UNIVERSITÉ DE GENÈVE

Archive ouverte UNIGE
https://archive-ouverte.unige.ch

Thèse 2025

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Traduction automatique du français vers les pictogrammes pour améliorer la communication entre médecins et patients avec une déficience intellectuelle en milieu hospitalier

Norré, Magali

How to cite

NORRÉ, Magali. Traduction automatique du français vers les pictogrammes pour améliorer la communication entre médecins et patients avec une déficience intellectuelle en milieu hospitalier. Doctoral Thesis, 2025. doi: 10.13097/archive-ouverte/unige:183266

This publication URL: https://archive-ouverte.unige.ch/unige:183266
Publication DOI: 10.13097/archive-ouverte/unige:183266

© The author(s). This work is licensed under a Creative Commons Attribution-NonCommercial (CC BY-NC 4.0) https://creativecommons.org/licenses/by-nc/4.0

UCLouvain

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Faculté de philosophie, arts et lettres
Centre de traitement automatique du langage

UNIVERSITÉ DE GENÈVE

UNIVERSITÉ DE GENÈVE

Faculté de traduction et d'interprétation
Département de traitement informatique multilingue

TRADUCTION AUTOMATIQUE DU FRANÇAIS VERS LES PICTOGRAMMES POUR AMÉLIORER LA COMMUNICATION ENTRE MÉDECINS ET PATIENTS AVEC UNE DÉFICIENCE INTELLECTUELLE EN MILIEU HOSPITALIER

Magali Norré

Thèse présentée en vue de l'obtention des titres de

Docteur en Langues, Lettres et Traductologie à l'Université catholique de Louvain

Docteur en Traitement Informatique Multilingue à l'Université de Genève

Prof. Dr. Thomas François, Université catholique de Louvain (promoteur)
Prof. Dr. Pierrette Bouillon, Université de Genève (promoteur)
Dr. Vincent Vandeghinste, Instituut voor de Nederlandse Taal & KU Leuven (promoteur)
Prof. Dr. Cédrick Fairon, Université catholique de Louvain (président du jury)
Prof. Dr. Didier Schwab, Université Grenoble Alpes (membre du jury)
Prof. Dr. Jean-Yves Antoine, Université de Tours (membre du jury)
Dr. Max De Wilde, Université de Genève & Université libre de Bruxelles (membre du jury)

2024–2025

To my mother
for her unconditional love and support.

Remerciements

Mener ces recherches a été une aventure passionnante qui n'aurait jamais été possible sans l'aide, les encouragements et le soutien indéfectible de nombreuses personnes connues depuis déjà fort longtemps ou rencontrées au fil de ces années.

Je souhaiterais remercier sincèrement mon promoteur Thomas François pour sa confiance, sa patience, son écoute en toutes circonstances, ses relectures et pour m'avoir sans cesse encouragée à me dépasser. J'ai appris énormément.

J'aimerais remercier mon promoteur (et professeur) Pierrette Bouillon de m'avoir offert de très nombreuses occasions de partager mes recherches, pour sa bienveillance et sa très grande disponibilité, incluant les heures de discussions sur divers sujets.

Je voudrais remercier mon promoteur Vincent Vandeghinste d'avoir tout de suite accepté de me guider dans le domaine du traitement automatique des langues, de m'avoir laissé tester librement de nouvelles choses et pour ses conseils. Thank you!

Merci à Leen Sevens pour le travail qu'elle m'a laissé, son livre que j'ai gardé précieusement (et emmené partout avec moi), ses présentations à Louvain-la-Neuve en 2017, ainsi qu'à Louvain en 2018. Sans cela, cette thèse aurait été très différente.

Je tiens à remercier chacun des membres de mon jury, que j'ai eu l'occasion de rencontrer à différents moments. Merci à Cédrick Fairon d'avoir immédiatement accepté d'en faire partie. Merci à Max De Wilde de m'avoir initié au TAL, d'avoir aussi été mon promoteur en master et de m'avoir fait rentrer en contact avec un expert en CAA.

Merci donc à Jean-Yves Antoine pour les passionnants échanges sur la CAA et son soutien depuis déjà plusieurs années. Merci à Didier Schwab pour toutes les discussions à Genève, Grenoble, Louvain-la-Neuve, mais également pour m'avoir proposé de participer au projet franco-suisse PROPICTO avec Benjamin Lecouteux.

J'en profite d'ailleurs pour remercier le Département de Traitement Informatique Multilingue (TIM) de Genève, ainsi que le Centre de Traitement Automatique du Langage (CENTAL) de Louvain-la-Neuve. [1] Je ne peux citer tout le monde.

Toutefois, j'aimerais remercier Trang, Eva, Rémi, Hubert, Johanna, Lucía, Joni ou encore Bastien pour leur aide, les divers échanges et l'accueil au sein de leur bureau. J'ai également passé de bons moments à discuter avec une certaine Anne le midi.

1. Merci à l'UNIGE et à l'UCLouvain pour le financement de cette thèse (les postes d'assistant de recherche, de boursier, ainsi que les mandats UCL-FSR N°13936.2020 et N°17005.2022).

Je remercie mes trois hôtes belges ou français à Genève, Elise, Isabelle et Gérard. Merci aux infirmières et médecins suisses d'avoir pris de leur temps pour me donner des informations sur la prise en charge des patients en situation de handicap.

J'adresse un grand merci à Anne C., logopède et fondatrice de l'association belge Comalso, qui m'a formée à la CAA pendant plusieurs années et qui m'a permis de faire de belles rencontres auprès de professionnels de la santé et d'utilisateurs de la CAA. Merci à Elisa D. de Pro Infirmis et à tous mes participants avec une DI pour leur aide.

Merci à Alain Guillet pour sa formation et ses conseils sur les modèles mixtes. Merci à Nicolas Ruytenbeek de m'avoir donné un avant-goût de la rigueur scientifique et des expériences en linguistique auprès de personnes en situation de handicap.

Merci à Gaëtan (alias Olé sempai) pour toutes les heures de jeu en ligne, les messages au sujet de tout et de rien, les colis de survie, les visites à Bruxelles et à Genève. Merci à Laura de m'avoir toujours considérée comme une deuxième petite sœur, la bonne humeur et tous les messages pour me demander des nouvelles.

Thanks my best friend forever Lisa (alias mon alter ego). Après notre première rencontre il y a déjà plus de vingt ans et toutes ces années d'amitié sincère, je te remercie infiniment pour ton soutien et ta confiance pour le meilleur et pour le pire.

Enfin et surtout, je voudrais dédier cette thèse de doctorat à ma famille, celle qui m'a élevée, ma mamy et ma maman. La seule raison pour laquelle je suis ici aujourd'hui, c'est vous. Je suis si fière d'avoir été adoptée par vous deux. Maman, tu m'as tout appris et tu seras toujours ma plus grande source d'inspiration où que nous soyons.

Résumé

L'objectif de cette thèse est d'explorer les façons dont la communication entre un médecin et un patient avec une Déficience Intellectuelle (DI) peut être améliorée à l'aide d'un système basé sur des pictogrammes. La communication entre médecins et patients est essentielle, car elle permet d'améliorer l'adhésion au traitement, l'état de santé des patients et leur qualité de vie. Ce type d'interactions peut toutefois entraîner diverses complications qui s'ancrent dans des problèmes de compréhension, en particulier, chez les personnes avec une DI. Nous étudions le recours à un système de communication reposant sur des pictogrammes dans le cadre de consultations médicales. Pour favoriser l'autodétermination des adultes avec une DI, nous évaluons leur compréhension de la langue médicale française lorsqu'elle prend la forme d'une traduction en pictogrammes. Nous proposons un cas d'utilisation original : nous adaptons au français Text-to-Picto, un système de traduction automatique de texte vers les pictogrammes et l'enrichissons de manière à intégrer les pictogrammes Arasaac. Nous analysons ensuite les erreurs du système sur un corpus de phrases médicales, explorons diverses stratégies de simplification et de désambiguïsation automatiques, afin de proposer des explications médicales efficacement traduites en pictogrammes. La simplification et la désambiguïsation lexicale pour la traduction automatique du français vers les pictogrammes sont étudiées au travers d'une approche faisant appel au Traitement Automatique des Langues (TAL) : avec des règles, des ressources linguistiques et des modèles de langue généraux et/ou médicaux (neuronaux, pré-entraînés, génératifs). Nos résultats sont comparables à l'état de l'art pour d'autres langues et montrent certaines améliorations par rapport à nos premières versions de Text-to-Picto. Dans le cadre de ces développements, nous avons créé plusieurs corpus d'évaluation, disponibles pour la recherche. Enfin, nous proposons deux études utilisateurs, à la fois quantitatives et qualitatives, qui évaluent la compréhension de pictogrammes isolés et de phrases en pictogrammes auprès d'adultes avec une DI. La première étude exploite des traductions générées par deux systèmes différents, Text-to-Picto et PictoDr, tandis que la deuxième se focalise sur la traduction d'une anamnèse dans le cadre du COVID-19. Les résultats montrent que plusieurs facteurs ont un impact significatif sur la compréhension. Ils mettent également en évidence le problème de la compositionnalité : les pictogrammes isolés sont mieux compris que les phrases en pictogrammes.

Abstract

The objective of this thesis is to investigate the ways in which communication between a doctor and a patient with an Intellectual Disability (ID) can be improved using a pictograph-based system. Communication between doctors and patients is essential because it improves adherence to treatment, patients' health status and quality of life. However, this type of interaction can lead to various complications that are rooted in problems of understanding, especially in people with ID. We study the use of a pictograph-based communication system in the context of medical consultations. In order to promote self-determination of adults with ID, we evaluate their understanding of the French medical language in the form of its translation into pictographs. We propose an original use case : we adapt Text-to-Picto to French, an automatic translation system from text to pictographs, and we extend it to the Arasaac pictographs. We then analyze the system's errors on a corpus of medical sentences, explore various automatic simplification and disambiguation strategies, in order to propose medical explanations efficiently translated in pictographs. Simplification and word sense disambiguation for automatic translation from French into pictographs are studied through an approach involving Natural Language Processing (NLP) : with rules, linguistic resources, and general and/or medical language models (neural, pre-trained, generative). Our results are similar to the state of the art for other languages and show some improvements over our first versions of Text-to-Picto. As part of these developments, we created several evaluation corpora, available for research. Finally, we propose two user studies, both quantitative and qualitative, which evaluate the understanding of isolated pictographs and the pictograph sentences among adults with ID. The first study exploits translations generated by two different systems, Text-to-Picto and PictoDr, while the second focuses on the translation of anamnesis in the context of COVID-19. The results show that several factors have a significant impact on the comprehension. They also highlight the problem of compositionality : isolated pictographs are better understood than pictograph sentences.

Table des matières

Résumé . . . 3
Liste des figures . . . 11
Liste des tables . . . 15
Liste des abréviations . . . 19
Préface . . . 21

Introduction **23**
1 Cadre de recherche . . . 25
2 Objectifs de la thèse . . . 27
3 Méthodologie générale . . . 28
4 Structure de la thèse . . . 30
5 Publications . . . 32

1 Contexte **33**
1 Communication alternative et améliorée . . . 34
1.1 Systèmes de communication alternative . . . 34
1.2 Différents utilisateurs . . . 35
1.2.1 Avec une déficience intellectuelle . . . 35
1.2.2 Autres personnes concernées . . . 37
2 Pictogrammes . . . 38
2.1 Définition et taxonomie . . . 38
2.2 Ensembles de pictogrammes . . . 39
2.2.1 Bliss . . . 40
2.2.2 Widgit . . . 40
2.2.3 Minspeak . . . 41
2.2.4 Picture Communication Symbols . . . 42
2.2.5 Sclera et Beta . . . 42
2.2.6 Arasaac . . . 43
2.3 Caractéristiques générales . . . 44
2.4 Évaluation des images . . . 47
2.5 Apprentissage des pictogrammes . . . 48
2.6 Effets des pictogrammes sur la communication . . . 50

3 Communication dans le domaine médical 51
3.1 Hôpitaux et accès aux soins 52
3.2 Expression et évaluation de la douleur 54
3.3 Applications médicales avec des images 58
4 Conclusion : des pictogrammes universels? 59

2 Traitement automatique : langues et pictogrammes 61
1 Traduction automatique 62
1.1 Typologie des systèmes de traduction 62
1.1.1 Traduction basée sur des règles 63
1.1.2 Traduction statistique 63
1.1.3 Traduction hybride 64
1.1.4 Traduction neuronale 64
1.1.5 Traduction avec des grands modèles de langue génératifs . 65
1.1.6 Traduction basée sur des phrases fixes 65
1.2 Systèmes de traduction vers les pictogrammes 66
1.2.1 Systèmes sans TAL 66
1.2.2 Systèmes avec TAL 67
1.3 Évaluation de la traduction vers les pictogrammes 73
1.3.1 Évaluation automatique 73
1.3.2 Évaluation manuelle 75
1.3.3 Étude utilisateur 77
1.4 Projet PROPICTO : Genève 78
1.4.1 Système BabelDr 78
1.4.2 Système PictoDr 80
1.5 Systèmes et évaluation de la traduction vers les pictogrammes . . . 82
2 Simplification de textes 84
2.1 Simplification manuelle 84
2.2 Simplification automatique 87
2.2.1 Simplification lexicale 87
2.2.2 Simplification syntaxique 87
2.2.3 Simplification neuronale 88
2.2.4 Simplification médicale 88
2.2.5 Ressources pour la simplification 89
2.2.6 Et les langues peu dotées? 89
2.2.7 Simplification pour les humains et les machines 90
3 Conclusion : peu de systèmes médicaux vers les pictogrammes 91

3 Système de base de Text-to-Picto **93**
1 Architecture du système de base . . . 94
2 WordNet : la route sémantique . . . 96
3 Corpus : des cas d'utilisation . . . 98
3.1 Corpus Email . . . 99
3.2 Corpus Livre . . . 99
3.3 Corpus Médical . . . 100
4 Paramétrage du système de base . . . 100
4.1 Quel WordNet français ? . . . 101
4.2 Quels paramètres ? . . . 101
5 Évaluation du système de base . . . 102
5.1 Évaluation automatique . . . 102
5.2 Évaluation manuelle . . . 105
6 Conclusion : des limites au système de base . . . 106

4 Spécialisation de Text-to-Picto **109**
1 Corpus de BabelDr et analyse syntaxique . . . 110
2 Analyse des stratégies de traduction en pictogrammes . . . 113
3 Analyse de la couverture médicale de Text-to-Picto . . . 115
3.1 Premières adaptations : version 2 de Text-to-Picto . . . 115
3.2 Paramétrage du système . . . 117
3.3 Évaluation automatique . . . 118
3.4 Évaluation manuelle . . . 119
3.4.1 Expérience 1 : le nombre de mots non traduits . . . 119
3.4.2 Expérience 2 : la couverture médicale . . . 120
4 Simplification générale et médicale . . . 123
4.1 Substitution lexicale : version 2.1 de Text-to-Picto . . . 123
4.1.1 Les synonymes de ReSyf . . . 123
4.1.2 Les paraphrases médicales . . . 128
4.1.3 Architecture du système . . . 129
4.2 Évaluation manuelle . . . 130
4.2.1 Expérience 1 : les synonymes de ReSyf . . . 130
4.2.2 Expérience 2 : les paraphrases médicales . . . 131
5 Simplification et grands modèles de langue génératifs . . . 133
5.1 État de l'art : grands modèles de langue génératifs . . . 133
5.2 Méthode de simplification . . . 134
5.3 Évaluation de la simplification . . . 137
6 Conclusion : des contributions . . . 143

5 Désambiguïsation lexicale **145**
1 État de l'art : désambiguïsation pour la traduction vers les pictogrammes . 146
2 Méthode de désambiguïsation 147
3 Évaluation de la désambiguïsation 151
3.1 Corpus d'évaluation 151
3.2 Évaluation automatique 152
4 Discussion : analyse des variables 156
4.1 Modèles de langue 156
4.2 Corpus d'entraînement 159
4.3 Synsets et pictogrammes 160
5 Et les grands modèles de langue génératifs ? 161
6 Conclusion : des perspectives pour la désambiguïsation 164

6 Études utilisateurs **165**
1 Évaluation de deux traductions en pictogrammes 166
1.1 Objectif et hypothèses de recherche 166
1.2 Participants 167
1.3 Design expérimental 168
1.4 Annotation et données collectées 173
1.5 Modèles linéaires généralisés à effets mixtes 176
1.5.1 Définition 176
1.5.2 Sélection 180
1.5.3 Prédiction 181
1.6 Résultats 183
1.6.1 Compréhension des pictogrammes et des phrases 183
1.6.2 Prédire automatiquement la compréhension 189
1.6.3 Communication, utilité des pictogrammes et amélioration 193
1.7 Discussion 194
2 Évaluation d'une anamnèse en pictogrammes 200
2.1 Objectif et hypothèses de recherche 200
2.2 Participants 201
2.3 Design expérimental 202
2.4 Annotation et données collectées 206
2.5 Modèles linéaires généralisés à effets mixtes 207
2.6 Résultats 208
2.6.1 Compréhension des questions 208
2.6.2 COVID-19, utilité des pictogrammes et amélioration 214
2.7 Discussion 217
3 Discussion générale : une méthodologie 224

4 Conclusion : vers de nouvelles recherches ? 228

Conclusion **229**

1 Contributions 229
2 Démo de Text-to-Picto 232
3 Limites et perspectives 233

A Corpus **237**

1 Corpus en pictogrammes Arasaac : Email, Livre, Médical 238
2 Corpus traduit en pictogrammes par une logopède 239
3 Corpus simplifié avec des grands modèles de langue génératifs 243
4 Corpus à désambiguïser 250

B Questionnaires de recherche **253**

1 Évaluation de deux traductions – Phrases 254
1.1 Toutes les phrases en pictogrammes 254
1.2 Liste de phrases et de pictogrammes : exemple 256
1.3 Phrase en pictogrammes : exemple 257
1.4 Liste de contrôle : exemple 258
2 Évaluation de deux traductions – Grille des participants 259
3 Évaluation de deux traductions – Guide du déroulement 260
4 Évaluation de deux traductions – Instructions 261
5 Évaluation de deux traductions – Consentement 262
6 Évaluation de deux traductions – Retranscriptions 263
7 Évaluation d'une anamnèse – Formulaire COVID-19 264
8 Évaluation d'une anamnèse – Phrases 266
8.1 Toutes les phrases en pictogrammes 266
8.2 Liste de phrases et de pictogrammes 267
8.3 Phrase en pictogrammes : exemple 268
9 Évaluation d'une anamnèse – Guide du déroulement 269
10 Évaluation d'une anamnèse – Instructions 270

C Code source pour l'analyse des données **271**

Bibliographie **273**

Liste des figures

1 Domaines de recherche de la thèse 25
2 Structure de la thèse 31

1.1 Représentations indirectes des pictogrammes Arasaac : exemple 39
1.2 Symboles Bliss : exemple 1 40
1.3 Pictogrammes Widgit : exemple 41
1.4 Pictogrammes Minspeak : exemple 41
1.5 Pictogrammes PCS : exemple 42
1.6 Pictogrammes Sclera, Beta et Arasaac : exemple 43
1.7 Pictogrammes pour la représentation d'un type de douleur 45
1.7a Arasaac : exemples 45
1.7b Sclera : exemples 45
1.8 Symboles Bliss : exemple 2 45
1.9 Pictogrammes Sclera et Arasaac : exemple 46
1.10 Images pour guider les patients dans des hôpitaux 53
1.10a Universal Symbols in Healthcare 53
1.10b Pictogrammes Arasaac en milieu médical 53
1.11 Supports papiers avec images utilisés dans des hôpitaux 57
1.11a Kit de communication 57
1.11b Medical Visual Language Translator 57
1.12 Pictogrammes Widgit Health : exemples 57
1.12a Chronologie d'événements 57
1.12b Porte-clés 57

2.1 Historique de la traduction automatique 62
2.2 Triangle de Vauquois 63
2.3 Système AraTraductor pour l'espagnol 68
2.3a Architecture 68
2.3b Exemple 68
2.4 Système text-to-picto pour l'espagnol 69
2.4a Architecture 69
2.4b Exemple 69

2.5 Système text-to-picto pour l'anglais 70
2.5a Architecture 70
2.5b Exemple 70
2.6 Système speech-to-picto pour le français : architecture et exemple 70
2.7 Système Voice2Picto pour le français 71
2.7a Architecture 71
2.7b Exemple 71
2.8 Système Glyph pour l'anglais : architecture et exemple 72
2.9 Système PicTalky pour l'anglais : architecture et exemple 73
2.10 Système BabelDr pour plusieurs langues : exemple 79
2.11 Système PictoDr UMLS-Arasaac pour le français : exemple 80
2.12 Exemples de traductions médicales de PictoDr en pictogrammes Arasaac 81
2.13 Images SantéBD : exemple 86

3.1 Text-to-Picto pour le français (système de base) : architecture 94
3.2 Text-to-Picto pour le français (système de base) : exemple 95
3.3 Lier les pictogrammes Arasaac aux WordNets français 98
3.4 Exemple de pictogrammes Sclera et Arasaac 99
3.5 Exemple de deux traitements du pluriel en pictogrammes Arasaac 100
3.6 Exemple de phrase du Corpus Médical en pictogrammes Arasaac 100

4.1 Méthodologie pour explorer les façons de traduire une phrase pour la communication entre médecins et patients avec Text-to-Picto : exemple . 110
4.2 Arbre syntaxique en constituants pour une phrase de BabelDr 111
4.3 Exemples de possibles traductions médicales en pictogrammes Arasaac . 115
4.4 Nombre de pictogrammes Arasaac créés par an entre 2008 et 2022 116
4.5 Pictogrammes Sclera, Beta et Arasaac : exemple des temps 117
4.6 Exemples de traductions médicales du Text-to-Picto pour le français (v2) en pictogrammes Arasaac 122
4.7 Ressource de synonymes ReSyf du Text-to-Picto pour le français (v2.1) : exemple 125
4.8 Exemples de synonymes avec ReSyf dans Text-to-Picto pour le français (v2.1) en pictogrammes Arasaac 127
4.9 Exemple de paraphrases médicales du Text-to-Picto pour le français (v2.1) en pictogrammes Arasaac 129
4.10 Text-to-Picto pour le français (v2.1) : architecture 129

5.1 Pictogrammes d'un mot à désambiguïser : exemple 1 145
5.2 Pictogrammes d'un mot à désambiguïser : exemple 2 147
5.3 Statistiques du corpus de 100 mots à désambiguïser 152

5.3a Nombre de lemmes par phrase 152
5.3b Nombre de synsets par mot 152
5.4 Résultats de la désambiguïsation lexicale (rang 1 et rang > 1) sur 100 mots polysémiques pour les ensembles de pictogrammes Sclera/Beta/Arasaac avec la métrique rappel : WOLF 153
5.5 Pictogrammes Arasaac : exemple de mots à désambiguïser 155
5.6 Résultats de la désambiguïsation lexicale (rang > 1) sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec les métriques rappel, précision et F-score : WOLF et WoNeF coverage, f-score, precision 156

6.1 Design expérimental de l'évaluation de deux traductions 170
6.2 Pictogrammes Arasaac : questions supplémentaires 171
6.2a Phrase 0 : exemple 171
6.2b Phrase 21 : exemple 171
6.3 Matrices de confusion avec les catégories « Correct » et « Incorrect » (A1), ainsi que les catégories « Correct » et « Incorrect / Ne sait pas » (A2) d'exactitude au niveau des pictogrammes et des phrases 174
6.4 Matrices de confusion avec les catégories « Correct » et « Incorrect » (A1), ainsi que les catégories « Correct » et « Incorrect » (A2) d'exactitude au niveau des pictogrammes et des phrases 175
6.5 Modèles linéaires généralisés à effets mixtes qui prédisent la compréhension des pictogrammes et la compréhension des phrases à partir des annotations 178
6.6 Compréhension selon un annotateur 184
6.6a Compréhension des pictogrammes par phrase et par type de traduction 184
6.6b Compréhension des phrases par phrase et par type de traduction . 184
6.7 Compréhension des pictogrammes et compréhension des phrases par type de traduction et par participant avec le niveau de langue selon un annotateur 187
6.8 Vérification des résidus des meilleurs modèles pour la compréhension des pictogrammes et la compréhension des phrases 191
6.9 Matrices de confusion pour la prédiction de la compréhension des pictogrammes et la compréhension des phrases, sans validation croisée . . . 191
6.10 Matrices de confusion pour la prédiction de la compréhension des pictogrammes et la compréhension des phrases, en validation croisée 192
6.11 Boîte à moustaches entre le pourcentage de pictogrammes corrects par phrase et la compréhension des phrases 193
6.12 Traductions de PictoDr : exemple 204

6.13 Matrice de confusion avec les catégories « Correct » et « Incorrect » (A1), ainsi que les catégories « Correct » et « Incorrect » (A2) d'exactitude au niveau des phrases 207
6.14 Compréhension des phrases de Text-to-Picto et PictoDr par participant selon un annotateur, résultats des deux études utilisateurs 209
6.15 Retranscription : exemple d'aide de la facilitatrice 210
6.16 Compréhension des phrases de PictoDr par phrase avec la section COVID-19 et par annotateur 211
6.17 Compréhension des phrases de PictoDr et temps en secondes par participant avec le niveau de langue selon un annotateur 213
6.18 Interface de Text-to-Picto 233
6.18a Entrée : phrase à traduire 233
6.18b Sortie : phrase traduite 233

B.1 Retranscriptions manuelles et automatiques : exemples 263
B.1a Retranscription manuelle : PictoDr 263
B.1b Retranscription manuelle : Text-to-Picto 263
B.1c Retranscription automatique : PictoDr 263
B.1d Retranscription automatique : Text-to-Picto 263

Liste des tables

2.1 Résultats de PictoDr pour la traduction de phrases vers une glose sémantique UMLS 81
2.2 Systèmes de traduction automatique vers les pictogrammes avec TAL . . . 83

3.1 Nombre de synsets non vides des WordNets français et pourcentage comparé au PWN 3.0 par POS 96
3.2 Résultats du Text-to-Picto français (système de base) pour WOLF et WoNeF pour les ensembles de pictogrammes Sclera/Beta/Arasaac avec la métrique BLEU 101
3.3 Valeurs des paramètres du Text-to-Picto français (système de base) sur le Corpus Email pour les ensembles de pictogrammes Sclera/Beta/Arasaac après réglage 102
3.4 Résultats du Text-to-Picto français et du Text-to-Picto néerlandais (systèmes de base) sur le Corpus Email pour les ensembles de pictogrammes Sclera/Beta/Arasaac avec les métriques BLEU, NIST, WER et PER 103
3.5 Résultats du Text-to-Picto français (système de base) sur le Corpus Livre pour l'ensemble de pictogrammes Arasaac avec les métriques BLEU, NIST, WER et PER 104
3.6 Résultats du Text-to-Picto français (système de base) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac avec les métriques BLEU, NIST, WER et PER 104
3.7 Résultats du Text-to-Picto français sur le Corpus Médical et du Text-to-Picto néerlandais/anglais/espagnol (systèmes de base) sur le Corpus Email pour les ensembles de pictogrammes Sclera/Beta/Arasaac avec les métriques précision, rappel et F-score 106

4.1 Structures syntaxiques présentes 30 fois ou plus dans le corpus de canoniques de BabelDr 112
4.2 Valeurs des paramètres du Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac après réglage 118

4.3 Résultats du Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac avec les métriques BLEU, NIST, WER et PER 118
4.4 Nombre de types (et nombre de tokens) non traduits avec Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac 119
4.5 Nombre de types non traduits avec Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac par POS les plus fréquents 120
4.6 Extrait de l'évaluation du Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac concernant la couverture par concept 121
4.7 Résultats du Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac concernant la couverture selon deux annotateurs 122
4.8 Résultats du Text-to-Picto français (v2.1), mode AZ, sur 100 phrases canoniques sans et avec le dictionnaire ReSyf pour l'ensemble de pictogrammes Arasaac avec les métriques précision, rappel et F-score avec amélioration relative 130
4.9 Résultats du Text-to-Picto français (v2.1), mode AZ, sur 100 termes techniques sans et avec le dictionnaire de paraphrases pour l'ensemble de pictogrammes Arasaac avec les métriques précision, rappel et F-score avec amélioration relative 131
4.10 Grands modèles de langue génératifs utilisés pour la simplification automatique de textes et la désambiguïsation lexicale 134
4.11 Résultats du Text-to-Picto français (v2.1), mode AZ, sur 100 canoniques, après simplification automatique avec des grands modèles de langue génératifs pour l'ensemble de pictogrammes Arasaac 137
4.12 Exemples de traductions médicales correctes en pictogrammes Arasaac du Text-to-Picto français (v2.1), après simplification automatique de l'entrée avec des grands modèles de langue génératifs 139
4.13 Exemples de traductions médicales incorrectes en pictogrammes Arasaac du Text-to-Picto français (v2.1), après simplification automatique de l'entrée avec des grands modèles de langue génératifs 141

5.1 Modèles de langue utilisés pour la désambiguïsation lexicale : Word2Vec, fastText, CamemBERT, FlauBERT, DrBERT et CamemBERT-bio 150

5.2 Résultats du Text-to-Picto français, mode AZ et ZA, sur 100 mots polysémiques sans désambiguïsation lexicale pour l'ensemble de pictogrammes Arasaac avec les métriques précision, rappel et F-score : WOLF 152
5.3 Résultats de la désambiguïsation lexicale (rang > 1) sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec les métriques précision et F-score : WOLF 154
5.4 Résultats de la désambiguïsation lexicale (rang > 1) de frWac2Vec et frWiki2Vec sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec les métriques rappel, précision et F-score : WOLF 155
5.5 Différents paramètres d'encodage pour les modèles BERT 157
5.6 Résultats de la désambiguïsation lexicale (rang > 1) des différentes configurations d'encodage pour BERT sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec la métrique précision : WOLF . . . 158
5.7 Résultats de la désambiguïsation lexicale avec des grands modèles de langue génératifs sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec les métriques rappel, précision et F-score : WOLF . 162

6.1 Description des participants de l'évaluation de deux traductions 168
6.2 Différences de traduction entre Text-to-Picto et PictoDr : variables 169
6.3 Différences de traduction entre Text-to-Picto et PictoDr : exemples 169
6.4 Annotation de la compréhension de deux pictogrammes et de la compréhension d'une phrase selon deux annotateurs : exemple 173
6.5 Nombre de phrases et nombre de pictogrammes Arasaac par questionnaire, évalués ou non par les deux annotateurs 174
6.6 Matrice de confusion pour la prédiction binaire : exemple 182
6.7 Compréhension des pictogrammes et compréhension des phrases par participant selon deux annotateurs 183
6.8 Corpus de test de l'évaluation de deux traductions 184
6.9 Compréhension de tous les pictogrammes, les moins compris selon un annotateur 185
6.10 Compréhension de tous les pictogrammes, les mieux compris selon un annotateur 186
6.11 Compréhension des pictogrammes et compréhension des phrases par niveau de langue selon un annotateur 187
6.12 Compréhension des pictogrammes et compréhension des phrases par type de traduction pour tous les participants selon un annotateur 187
6.13 Compréhension des pictogrammes et compréhension des phrases par type selon un annotateur 188
6.14 Compréhension des pictogrammes par POS selon un annotateur 188

6.15 Comparaison de l'AIC des modèles linéaires généralisés à effets mixtes pour la compréhension des pictogrammes et la compréhension des phrases189
6.16 Meilleurs modèles linéaires généralisés à effets mixtes pour prédire la compréhension des pictogrammes et la compréhension des phrases . . . 190
6.17 Niveaux de significativité des résultats de compréhension des pictogrammes et de compréhension des phrases par hypothèse de l'évaluation de deux traductions selon un annotateur 194
6.18 Description des participants de l'évaluation d'une anamnèse 202
6.19 Corpus de test de l'évaluation de l'anamnèse du COVID-19 203
6.20 Nombre de phrases et nombre de pictogrammes Arasaac du questionnaire lié au COVID-19, évalués ou non par les deux annotateurs 206
6.21 Compréhension des phrases de PictoDr par participant selon deux annotateurs 208
6.22 Compréhension des phrases de PictoDr par section COVID-19 pour tous les participants selon deux annotateurs 210
6.23 Compréhension des phrases de PictoDr par nombre de pictogrammes par phrase pour tous les participants selon deux annotateurs 212
6.24 Modèle linéaire généralisé à effets mixtes pour prédire la compréhension des phrases de PictoDr pour une anamnèse du COVID-19 212
6.25 Durée de l'évaluation de l'anamnèse par participant 213
6.26 Connaissances préalables de la maladie du COVID-19 215
6.27 Niveaux de significativité des résultats de compréhension des phrases par hypothèse de l'évaluation d'une anamnèse selon deux annotateurs . . 217

B.1 Questionnaire 1 : traductions de Text-to-Picto et PictoDr 254
B.2 Questionnaire 2 : traductions de Text-to-Picto et PictoDr 255
B.3 Questionnaire : traductions de PictoDr, anamnèse du COVID-19 266

Liste des abréviations

AAC Augmentative and Alternative Communication

AIC Akaike Information Criterion

API Application Programming Interface

BERT Bidirectional Encoder Representations from Transformers

BIC Bayesian Information Criterion

BLEU BiLingual Evaluation Understudy

CAA Communication Alternative et Améliorée

CBOW Continuous Bag Of Words

CECR Cadre Européen Commun de Référence pour les langues

DI Déficience Intellectuelle

FALC Facile A Lire et à Comprendre

GPT Generative Pre-trained Transformer

HUG Hôpitaux Universitaires de Genève

IC Intervalle de Confiance

ISAAC International Society for Augmentative and Alternative Communication

JSON JavaScript Object Notation

LLM Large Language Model

NIST National Institute of Standards and Technology

NLP Natural Language Processing

OR Odds Ratio

PER Position-independent word Error Rate

POS Part Of Speech

PROPICTO PRojection du langage Oral vers des unités PICTOgraphiques

PWN Princeton WordNet

QI Quotient Intellectuel

SVM Séparateur à Vaste Marge

TAL Traitement Automatique des Langues

TSA Trouble du Spectre de l'Autisme

UMLS Unified Medical Language System

URL Uniform Resource Locator

VIF Variance Inflation Factor

WER Word Error Rate

WOLF WOrdnet Libre du Français

WoNeF WordNet du Français

XML eXtensible Markup Language

Choix terminologiques

Communication Alternative et Améliorée Nous avons choisi d'employer le terme utilisé par l'association internationale pour la communication alternative et améliorée (ISAAC Francophone). Il est également utilisé par l'association belge Comalso. De plus, il s'agit du terme le plus fréquent dans Google Scholar, par rapport à « communication alternative et augmentée (ou augmentative) ».

Large Language Model Dans cette thèse, nous utiliserons des grands modèles de langue génératifs, c'est-à-dire des réseaux de neurones avec une architecture de type *tranformer* (où seule la partie consacrée au décodage (le décodeur) est néanmoins conservée) de plusieurs milliards de paramètres.

Préface

Au cours des dernières années, nous avons travaillé sur différentes applications en Traitement Automatique des Langues (TAL). Les projets (inter)nationaux auxquels nous avons eu le plaisir de participer sont tous liés plus ou moins étroitement avec la présente thèse de doctorat même si, pour certains, le lien peut sembler moins évident. C'est pourquoi, nous n'avons pas toujours inclus le contenu de ces publications, mais nous nous permettons de toutes les mentionner ici et de remercier nos co-auteurs.

Tout d'abord, nous avons commencé à nous intéresser au TAL, ainsi qu'à la Communication Alternative et Améliorée (CAA) avec notre mémoire de master (Norré et De Wilde, 2018; Norré, 2020a,b), qui portait sur le système de prédiction de mots Presage et l'interface utilisateur ACAT. Nous avons configuré et évalué ces systèmes automatiquement, puis auprès d'un adulte non verbal avec une infirmité motrice cérébrale.

L'annotation est une tâche souvent nécessaire. Nous y avons été confronté plusieurs fois durant nos recherches, y compris avec le projet du CENTAL AMesure (François *et al.*, 2020) ou quand nous avons annoté un corpus pour la simplification automatique (Cardon *et al.*, 2022, 2023). Nous avons pu faire un parallèle entre les opérations de simplification annotées et certaines stratégies de traduction vers les pictogrammes.

Le système sur lequel nous avons le plus travaillé et sur lequel porte cette thèse est Text-to-Picto, conçu à l'origine par Vandeghinste, notre co-promoteur d'INT et de la KU Leuven, pour les personnes avec une Déficience Intellectuelle (DI). Les études sur le développement et l'évaluation de la version française de ce système de traduction vers les pictogrammes ont aussi été publiées (Norré *et al.*, 2021c,d, 2022, 2023, 2024).

Un de nos objectifs dans cette thèse est d'utiliser Text-to-Picto pour traduire automatiquement les phrases du système de traduction médicale BabelDr destiné aux patients allophones. Nous avons débuté nos recherches en travaillant sur cet outil et en testant l'efficacité de son interface de réponses incluant les pictogrammes Arasaac et Sclera utilisés dans Text-to-Picto (Norré *et al.*, 2020, 2021a,b; Bouillon *et al.*, 2023).

Le projet BabelDr a été prolongé avec le projet franco-suisse PROPICTO (Ormaechea Grijalba *et al.*, 2023a,b). Au sein du Département de TIM de l'Université de Genève, nous avons participé à la constitution et l'évaluation de ressources en pictogrammes pour le système de traduction PictoDr (Mutal *et al.*, 2022; Gerlach *et al.*, 2023). Un de nos corpus a été intégré dans le projet Pantagruel (Segonne *et al.*, 2024a,b).

Introduction

There is no greater agony
than bearing an untold story inside you.
Maya Angelou

Accompagner un proche qui a un handicap cognitif chez le médecin ou à l'hôpital est une situation qui peut arriver à tout le monde au cours de la vie. Il peut s'agir d'une connaissance, d'un ami, d'un parent, etc. qui a plus de difficultés à s'exprimer, que ce soit de naissance, temporairement ou de façon permanente suite à un accident ou à une maladie parfois neurodégénérative. Ce cas, nous l'avons nous-même rencontré plusieurs fois, et à chaque fois, nous avons été confronté aux mêmes questions. En tant qu'aidant proche, est-il nécessaire d'aider la personne à exprimer sa douleur et ses antécédents médicaux ou faut-il ne pas intervenir et laisser la personne s'exprimer, quitte à ce que la consultation dure plus longtemps, que le message ne soit pas toujours compréhensible, cohérent ou complet, avec toujours un risque non négligeable de mauvais diagnostic dans les deux cas ?[2] Comment peut-on améliorer la communication entre le médecin et le patient si celui-ci a une Déficience Intellectuelle (DI) ?

C'est cette deuxième question qui a inspiré cette thèse. Nous nous sommes demandé s'il était possible de développer un système qui pourrait favoriser l'autodétermination du patient et qui permettrait au médecin de communiquer des informations médicales autrement que par le langage verbal. Notre propos ici n'est pas de proposer un système adapté à tous les patients avec une DI qui remplacerait complètement l'accompagnant, mais de contribuer à notre niveau aux recherches dans le domaine de l'aide au handicap et de la Communication Alternative et Améliorée (CAA), en nous basant sur des techniques de Traitement Automatique des Langues (TAL). C'est pourquoi, notre premier objectif principal est d'adapter et d'améliorer un outil existant en le spécialisant pour la communication entre médecins et patients avec une DI. Il s'agit de Text-to-Picto, un système conçu en Belgique (Vandeghinste *et al.*, 2017; Sevens, 2018).

2. Des problèmes de communication avec des personnes ayant une DI peuvent engendrer des erreurs ou des retards de diagnostic (Tuffrey-Wijne et McEnhill, 2008; Ibba *et al.*, 2020), ainsi que des (sous-)traitements inadaptés (de Knegt *et al.*, 2013), à cause d'une mauvaise évaluation de la douleur. Or, d'après la Convention des Nations Unies relative aux droits des personnes handicapées de 2006, les « professionnels de la santé » doivent dispenser « aux personnes handicapées des soins de la même qualité que ceux dispensés aux autres », ainsi qu'obtenir leur « consentement libre et éclairé ».

Text-to-Picto a été développé pour les personnes avec une DI. Il traduit automatiquement des emails néerlandais, anglais et espagnols vers des pictogrammes (Sevens *et al.*, 2015a). Les pictogrammes sont un moyen de CAA utilisé par les personnes avec une DI pour communiquer dans la vie de tous les jours (Beukelman et Mirenda, 2017). Le système Text-to-Picto est libre de droit, utilisable en ligne et exploite déjà différentes méthodes de TAL bien documentées dans la littérature scientifique, plus complètes et avancées que d'autres systèmes. [3] Bien que l'outil propose plusieurs paires de langues, le français n'est pas disponible. De plus, les ensembles de pictogrammes intégrés sont très limités en termes de représentations médicales. Nous proposons donc d'étendre le système en ajoutant le français comme langue source et le large ensemble de pictogrammes pour la CAA appelé Arasaac comme langue cible.

Nous évaluons Text-to-Picto et le spécialisons à notre cas d'utilisation en partant des phrases médicales françaises que nous voulons traduire automatiquement en pictogrammes : des instructions ou questions fréquemment posées lorsque l'on arrive aux urgences hospitalières et que le médecin pourrait entrer au clavier pour effectuer son anamnèse. Par exemple, « je vais palper le ventre » ou « depuis combien de temps avez-vous mal? ». Pour ce faire, nous recourons à différentes disciplines du TAL et nous évaluons leur impact sur la traduction automatique vers les pictogrammes. Une fois Text-to-Picto adapté, nous pouvons répondre à notre deuxième objectif principal, qui consiste en une évaluation de la compréhension de pictogrammes et de phrases en pictogrammes traduites avec ce système auprès d'adultes avec une DI. Elle vise à déterminer quels sont les facteurs qui ont un impact significatif sur la compréhensibilité. Comme il ne s'agit pas d'une thèse en informatique, ni en ergonomie, notre but n'est pas de développer un système complet directement utilisable auprès de potentiels utilisateurs. À la place, nous évaluons scientifiquement l'apport de diverses techniques et ressources dans un contexte spécifique : la traduction automatique de dialogues médicaux français en pictogrammes pour des adultes avec une DI. En ce qui concerne notre deuxième objectif, il s'agit bien d'évaluer l'utilité de la traduction en pictogrammes générée automatiquement auprès de ce public, plutôt que d'évaluer l'utilisation de ce système de traduction en situation réelle ou sa qualité en termes d'erreurs.

La première section montre où nous nous situons parmi les trois domaines de recherche dans lesquels s'inscrit notre travail (section 1). La deuxième section présente nos questions en lien avec la revue de la littérature, ainsi que nos deux objectifs divisés en sous-objectifs et hypothèses de recherche par chapitre (section 2). La troisième section donne un aperçu de la méthodologie générale que nous utilisons pour répondre à chacun de nos sous-objectifs (section 3). La quatrième section détaille la structure de ce manuscrit par chapitre (section 4). La cinquième et dernière section liste nos publications et communications incluses dans cette thèse (section 5).

3. https://picto.ccl.kuleuven.be

1 Cadre de recherche

Ce projet de recherche se veut interdisciplinaire, trois domaines sont concernés : le médical, la CAA et le TAL (Figure 1).

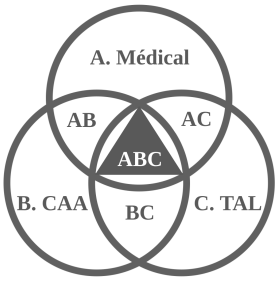


FIGURE 1 – Domaines de recherche de la thèse

- **A. Médical** – Nous avons décidé de nous concentrer sur la communication entre un médecin et un patient, en particulier, la situation de l'anamnèse où l'on utilise la langue médicale. Ce choix a un impact sur les deux autres domaines.
- **B. CAA** – Il existe beaucoup de systèmes et de méthodes différentes utilisées pour la Communication Alternative et Améliorée (CAA), qui permettent de remplacer ou d'améliorer la communication des personnes – enfants ou adultes avec divers handicaps –, qui ont tous un besoin communicationnel au quotidien (Cataix-Nègre, 2011; Beukelman et Mirenda, 2017). Nous nous intéressons ici aux technologies (text-to-picto), qui traduisent des langues naturelles en pictogrammes pour les adultes avec une DI, en étudiant leur compréhension des pictogrammes (versant réceptif de la CAA). Nous avons fait le choix de ne pas travailler sur les systèmes (picto-to-text), qui permettent aux utilisateurs avec une DI de saisir directement des pictogrammes via des grilles (versant expressif), afin de composer un message en langue naturelle qui sera ensuite affiché ou oralisé avec une synthèse vocale comme c'est le cas des tableaux de communication numériques.
- **C. TAL** – Le Traitement Automatique des Langues (TAL) regroupe de nombreuses disciplines, tâches et applications différentes, qui ont chacune leurs propres histoires et méthodes (Jurafsky et Martin, 2024). Ici, nous abordons la traduction automatique, la simplification automatique et la désambiguïsation lexicale, des larges champs de recherche concernant les données textuelles, mais relativement peu exploités pour les données pictographiques. Ces deux dernières disciplines sont notamment déjà utilisées pour améliorer la traduction automatique de langues avec moins de ressources (Štajner et Popović, 2016). Elles pourraient permettre ici de traduire correctement plus de mots en pictogrammes. Nous utilisons différents types de modèles de langue et des ressources lexico-sémantiques générales pour la langue française. Nous les comparons à des ressources développées pour la langue médicale quand celles-ci sont disponibles.

- **AB. Médical | CAA** – Les systèmes de CAA et les pictogrammes sont utilisés dans la vie de tous les jours, y compris dans des situations médicales : pour exprimer la douleur (de Knegt *et al.*, 2016a), faciliter les rendez-vous médicaux en préparant la consultation (Vaz, 2013), etc. Il existe cependant moins d'outils de CAA médicaux que de systèmes généraux. Des applications comme SantéBD ou MediPicto sont simples et représentent des phrases médicales sous la forme d'une seule image plutôt que via une séquence de pictogrammes comme nous le proposons.
- **AC. Médical | TAL** – Chaque discipline du TAL qui nous intéresse peut concerner la langue générale, mais aussi la langue médicale et ses spécificités. Pour spécialiser Text-to-Picto, nous nous basons sur des études en TAL qui permettent de traduire, désambiguïser et simplifier automatiquement des phrases médicales.
- **BC. CAA | TAL** – À notre connaissance, peu d'outils de CAA utilisent des méthodes de TAL et les derniers modèles de langue. La traduction automatique vers les pictogrammes est une sous-tâche du TAL relativement récente et méconnue. Seuls quelques systèmes de traduction de phrases vers les pictogrammes avec du TAL existent pour l'anglais (Imam *et al.*, 2019), l'espagnol (Bautista *et al.*, 2017), voire un pour le français (Vaschalde *et al.*, 2018), mais ils ne sont pas adaptés pour le domaine médical et/ou nos utilisateurs cibles.
- **ABC. Médical | CAA | TAL** – Il s'agit d'un vide dans la littérature scientifique que nous tentons de combler en adaptant Text-to-Picto pour la communication médicale. À notre connaissance, il n'existait pas de recherches similaires lorsque nous avons débuté cette thèse, même si quelques études ont depuis émergé parallèlement à nos propres travaux. Dans cette thèse, nous testons différentes méthodes et ressources utilisées en TAL, en évaluant leur impact sur un cas d'étude original : la traduction automatique du français vers les pictogrammes pour la CAA médicale. C'est la raison pour laquelle nous proposons aussi de nouveaux corpus pour développer et améliorer ce type de systèmes. Nous menons finalement les deux premières études utilisateurs pour évaluer des phrases médicales en pictogrammes générées par ce système auprès d'adultes avec une DI.

Ce projet doit donc faire face à trois problèmes majeurs liés à ces trois domaines : **(1)** l'absence de corpus en pictogrammes authentique pour la CAA médicale, c'est-à-dire de corpus d'évaluation et de corpus parallèles qui lient des phrases médicales aux pictogrammes correspondants et qui sont nécessaires pour entraîner des modèles de traduction en TAL ; **(2)** le manque de directives de professionnels de la santé pour traduire des phrases en pictogrammes ; **(3)** mais également la difficulté de représenter le sens des phrases médicales avec les ensembles de pictogrammes de CAA libres de droit pour des personnes avec une DI, qui n'ont pas toujours appris à communiquer avec des images et les pictogrammes que nous utilisons.

2 Objectifs de la thèse

Au-delà de ces obstacles, le but des chapitres 1 et 2 est de répondre aux questions ci-dessous. Les autres chapitres répondent à deux objectifs et plusieurs sous-objectifs. Notre thèse propose de tester l'impact de méthodes de TAL pour améliorer la traduction vers les pictogrammes et d'évaluer des traductions auprès d'adultes avec une DI.

- Qu'est-ce que la CAA et qui sont les utilisateurs ? (cf. chapitre 1)
- Comment les patients avec une DI communiquent-ils des informations médicales ? (cf. chapitre 1)
- Quelles sont les barrières auxquelles sont particulièrement confrontés les patients avec une DI ? (cf. chapitre 1)
- Existe-t-il des outils avec des images sur papier et/ou technologique afin de faciliter la communication, ainsi que le diagnostic des médecins dans les hôpitaux ? (cf. chapitre 1)
- Existe-t-il des systèmes de traduction automatique vers les pictogrammes ? Si oui, quelle est leur architecture et comment les évalue-t-on ? (cf. chapitre 2)

Objectif 1 : développer et améliorer un système de traduction automatique du français vers les pictogrammes pour la communication entre médecins et patients avec une DI

1. Adapter Text-to-Picto au français (cf. chapitre 3)
2. Étendre Text-to-Picto à l'ensemble de pictogrammes Arasaac (cf. chapitre 3)
3. Explorer des façons de traduire une phrase médicale (issue de l'échelle de tri des urgences hospitalières) en pictogrammes ; déterminer des stratégies de traduction pour le développement du système (cf. chapitre 4, sections 1 et 2)
4. Explorer des façons de spécialiser un système pour la traduction automatique de dialogues médicaux vers les pictogrammes et son amélioration (cf. chapitre 4, section 3)
5. Évaluer l'impact de la substitution lexicale avec des ressources générales ou médicales, ainsi que l'influence de la simplification automatique de textes générée avec des grands modèles de langue génératifs sur la traduction automatique du français vers les pictogrammes (cf. chapitre 4, sections 4 et 5)
6. Évaluer l'impact de la désambiguïsation lexicale sur la traduction automatique du français vers les pictogrammes, dont plusieurs variables (cf. chapitre 5)

- Évaluer les – types de – modèles de langue (incluant différents paramètres) : plongements statiques, contextuels | neuronaux, pré-entraînés, génératifs
- Évaluer les – types de – corpus d'entraînement : généraux et/ou médicaux
- Évaluer la ressource lexico-sémantique du système de traduction français : WOrdnet Libre du Français (WOLF), WordNet du Français (WoNeF)

Objectif 2 : évaluer les traductions médicales en pictogrammes chez des adultes avec une DI

Notre deuxième objectif vise à répondre à la question de recherche suivante : des systèmes de traduction automatique du français vers les pictogrammes produisent-ils une sortie utile et compréhensible au niveau des pictogrammes et/ou des phrases en pictogrammes pour des adultes avec une DI ? Elle est composée de 3 sous-objectifs qui peuvent comprendre plusieurs hypothèses de recherche.

1. Évaluer la compréhension de pictogrammes et la compréhension de phrases en pictogrammes générées par Text-to-Picto et PictoDr chez des adultes avec une DI ; déterminer des facteurs qui l'influencent de façon significative (cf. chapitre 6, sections 1 et 3)

2. Évaluer la compréhension de questions en pictogrammes en vue d'une anamnèse du COVID-19 chez des adultes avec une DI (cf. chapitre 6, sections 2 et 3)

3. Proposer une méthodologie quantitative et qualitative pour évaluer la compréhension des pictogrammes et la compréhension des phrases en pictogrammes, qui répond à nos sous-objectifs et hypothèses de recherche ci-dessus, tout en étant adaptée aux adultes verbaux – qui utilisent le langage oral ici – avec une DI qui ne connaissent pas les pictogrammes (cf. chapitre 6, sections 1, 2 et 3)

3 Méthodologie générale

Afin d'atteindre nos sous-objectifs 1, 2, 3, 4 et 5 (objectif 1), nous utilisons une approche de développement de système itérative. Dans le chapitre 3, nous proposons une première version, le système de base du Text-to-Picto français. Après des évaluations et l'analyse de ses limitations, nous l'adaptons, puis nous proposons des améliorations, etc. jusqu'au chapitre 4 (section 4). Le système évalué dans le dernier chapitre n'est pas le système complet. La simplification automatique via des grands modèles de langue génératifs (sous-objectif 5), ainsi que la désambiguïsation lexicale (sous-objectif 6) ne sont pas intégrées dans le système Text-to-Picto. Différentes investigations sont réalisées en parallèle. En ce qui concerne l'objectif 1, nous utilisons les

évaluations automatiques classiques en traduction automatique avec des métriques bien connues (sous-objectifs 1, 2 et 4), des évaluations manuelles avec un ou plusieurs annotateurs (sous-objectifs 1, 2, 4, 5 et 6). Nous adaptons et construisons plusieurs corpus pour le développement ou l'évaluation des différentes versions du système de traduction automatique du français vers les pictogrammes (le système de base sur plusieurs cas d'utilisation, la spécialisation au langage médical et la désambiguïsation lexicale). Nous faisons appel à une logopède (experte en CAA), afin d'avoir son avis sur la traduction de nos phrases médicales en pictogrammes pour adapter le système (sous-objectif 3). Pour l'objectif 2 (sous-objectifs 1, 2 et 3), nous recourons à des études utilisateurs, quantitatives et qualitatives, auprès de personnes cibles avec une DI et une facilitatrice sous la forme d'entretiens. Nous constituons ainsi un corpus français original, disponible en libre accès. Il contient les retranscriptions des 19 entretiens. Aucun corpus de ce type n'existait quand nous avons commencé ce travail. Nous utilisons ensuite des modèles linéaires généralisés à effets mixtes pour identifier les variables qui ont un impact sur la compréhension de pictogrammes et la compréhension de phrases en pictogrammes.

Impliquer des personnes cibles

Dans cette thèse, nous allons voir que les pictogrammes sont déjà utilisés par différents publics, par exemple, les personnes allophones ou en situation de handicap. En ce qui concerne la DI, elle regroupe également des profils très variés. Nous avons donc dû faire un choix en fonction de différentes contraintes, par exemple, temporelle (la durée de la thèse qui est limitée à quatre ans) et géographique (nous avions des contacts spécifiques en Belgique, d'autres en Suisse). Nous avons choisi de cibler des adultes avec une DI légère verbaux capables d'expliquer oralement le sens des phrases évaluées en pictogrammes. [4] Cela reste cohérent de les faire participer car cela permet de voir les adaptations et options qui pourraient être apportées pour un public (avec une DI) plus large. Impliquer des personnes avec une DI plus sévère non verbales aurait aussi été intéressant, mais cela n'a pas été possible ici pour plusieurs raisons. Nous aurions eu besoin d'encadrants paramédicaux qui connaissent bien les participants avec une DI. Or, les personnes contactées (logopèdes, infirmières spécialisées, etc.) étaient déjà débordées et peu disponibles. La pandémie de COVID-19 a en effet eu un impact direct sur notre thèse. Au début du projet, il n'était pas possible d'organiser des focus groups centrés sur une analyse des besoins avec les confinements successifs. Nous avons donc impliqué des personnes cibles à différents moments : une logopède pour la constitution d'un corpus qui a servi à proposer de premières adaptations au système de traduction et des adultes avec une DI les deux dernières années du projet.

4. La passation de questionnaires écrits ou en ligne ne nous semblait pas être envisageable car elle nécessite des connaissances que toutes les personnes avec une DI n'ont pas.

4 Structure de la thèse

Ce travail s'organise en six chapitres (Figure 2). Le premier décrit le cadre théorique dans lequel nous présentons le contexte général (**chapitre 1**). Il s'agit de donner un aperçu des systèmes de CAA et des ensembles de pictogrammes dans ce contexte, avant d'introduire la communication dans le domaine médical avec et sans images, leurs avantages, mais aussi les limites pour les patients avec une DI. Nous introduisons ensuite les disciplines du TAL, qui sont exploitées dans les autres chapitres, à savoir la traduction automatique et la simplification automatique de textes (**chapitre 2**). Nous présentons ensuite le système Text-to-Picto, que nous adaptons (**chapitre 3**). Nous détaillons la configuration d'un système de base pour le français vers trois ensembles de pictogrammes conçus pour les personnes avec une DI. Il fait l'objet d'une évaluation automatique et manuelle sur plusieurs corpus. Nous analysons des phrases que nous souhaitons faire traduire automatiquement en pictogrammes (**chapitre 4**). Il s'agit de questions et d'instructions de l'échelle de tri des urgences hospitalières, qui proviennent de BabelDr, un système de traduction de la parole pour le dialogue médical. Nous apportons des modifications au système Text-to-Picto pour les mots non traduits ou incorrects. Nous testons l'impact de différentes ressources pour la substitution lexicale, ainsi que l'influence de la simplification automatique avec des grands modèles de langue génératifs. Ensuite, nous cherchons à évaluer l'apport de la désambiguïsation lexicale à un système de traduction automatique vers les pictogrammes tel que Text-to-Picto (**chapitre 5**). Plusieurs types de modèles de langue sont évalués automatiquement : des modèles neuronaux et des modèles pré-entraînés généraux et/ou médicaux, mais également des grands modèles de langue génératifs. Enfin, nous proposons deux études utilisateurs avec une facilitatrice qui visent à tester la compréhension de pictogrammes et la compréhension de phrases en pictogrammes par des adultes avec une DI (**chapitre 6**). La première sur des traductions de phrases indépendantes générées automatiquement avec deux approches de traduction en pictogrammes : Text-to-Picto et PictoDr. La deuxième sur un cas plus réaliste, une anamnèse de la maladie du COVID-19 en pictogrammes.

Ces chapitres sont suivis par une conclusion et des annexes contenant des informations supplémentaires, incluant des corpus (annexe A), des documents des deux études utilisateurs (annexe B), ainsi que le code source utilisé pour l'analyse des données de la première étude (annexe C). Le code source de Text-to-Picto, les modèles de langue (Word2Vec et fastText), les corpus, les documents des études utilisateurs et autres données relatives à cette thèse sont disponibles en libre accès ou à la demande. [5]

5. https://github.com/VincentCCL/Picto

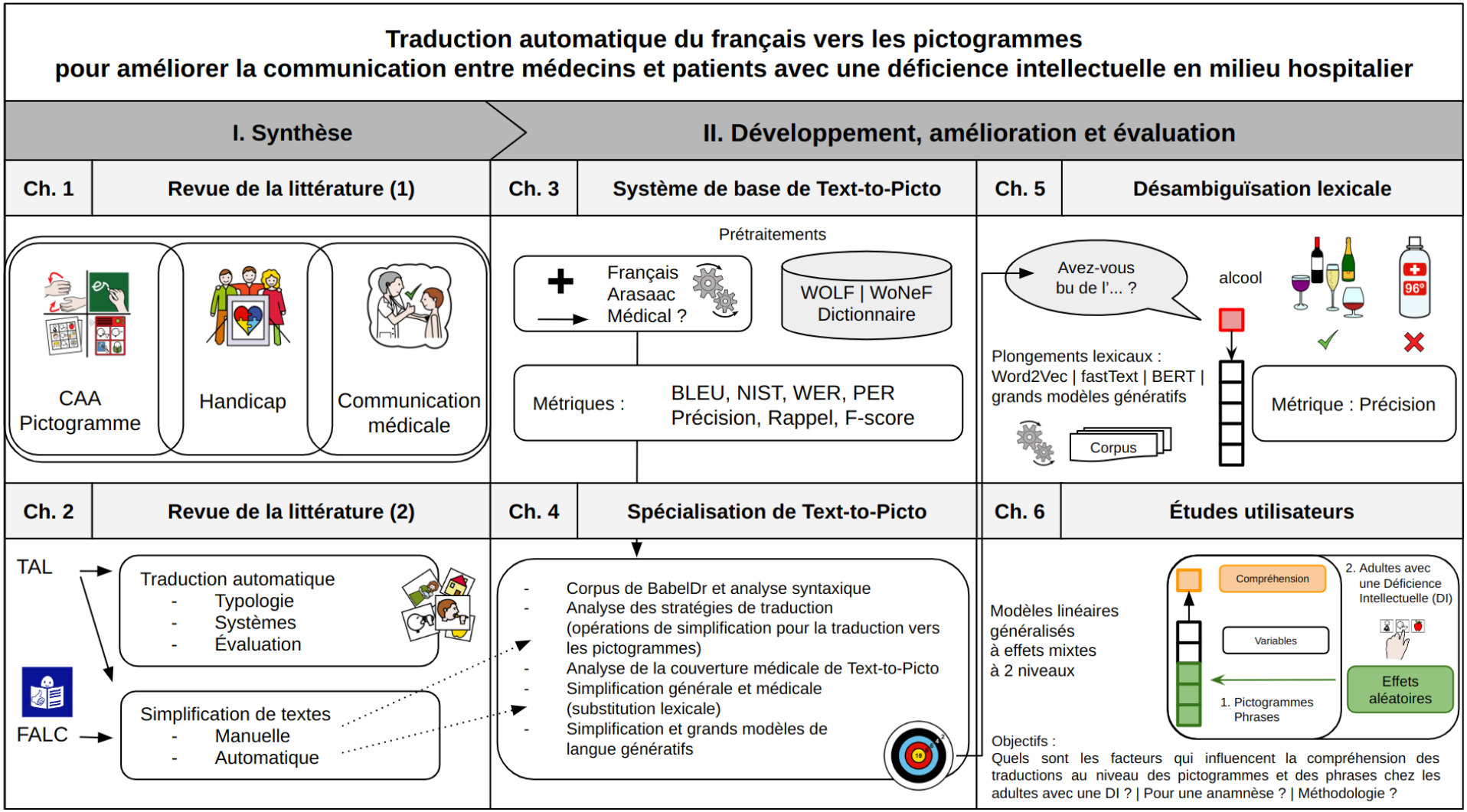


FIGURE 2 – Structure de la thèse

5 Publications

Comme nous l'avons déjà mentionné, les travaux présentés dans cette thèse sont basés sur des publications et des communications avec différents co-auteurs :

- le chapitre 3, le système de base de Text-to-Picto :

  Magali NORRÉ, Vincent VANDEGHINSTE, Pierrette BOUILLON et Thomas FRANÇOIS, Experiments for the Adaptation of Text2Picto to French. *In Meeting of Computational Linguistics in The Netherlands (CLIN31)*, 2021.

  Magali NORRÉ, Vincent VANDEGHINSTE, Pierrette BOUILLON et Thomas FRANÇOIS, Extending a Text-to-Pictograph System to French and to Arasaac. *In Proceedings of the 13th International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1050–1059, 2021. INCOMA Ltd.

- le chapitre 4, les stratégies de traduction en pictogrammes avec Text-to-Picto :

  Magali NORRÉ, Vincent VANDEGHINSTE, Thomas FRANÇOIS et Pierrette BOUILLON, Investigating the Medical Coverage of a Translation System into Pictographs for Patients with an Intellectual Disability. *In Proceedings of the Ninth Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2022)*, pages 44–49, Dublin, Ireland, 2022. Association for Computational Linguistics.

- le chapitre 5, la désambiguïsation lexicale pour Text-to-Picto :

  Magali NORRÉ, Rémi CARDON, Vincent VANDEGHINSTE et Thomas FRANÇOIS, Word Sense Disambiguation for Automatic Translation of Medical Dialogues into Pictographs. *In Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing (RANLP 2023)*, pages 803–812, Varna, Bulgaria, 2023. INCOMA Ltd.

- le chapitre 6, une évaluation des traductions de Text-to-Picto et PictoDr auprès d'adultes avec une DI :

  Magali NORRÉ, Trang Trần Hạnh PHẠM, Pierrette BOUILLON, Vincent VANDEGHINSTE et Thomas FRANÇOIS, Évaluation de la traduction en pictogrammes pour la communication médecin-patient par des adultes avec une déficience intellectuelle. *In Actes de la 13e conférence de l'IFRATH sur les technologies d'assistance. Handicap 2024 : Des solutions personnalisées pour des besoins spécifiques*, pages 181–186, Paris, France, 2024. IFRATH.

Chapitre 1

Contexte

Introduction

Dans ce chapitre, nous commençons par introduire le contexte théorique dans lequel s'intègre le système de traduction automatique de texte vers les pictogrammes que nous développerons dans cette thèse. Nous présentons d'abord la Communication Alternative et Améliorée (CAA) qui est notamment utilisée par les personnes avec une Déficience Intellectuelle (DI) pour communiquer dans la vie de tous les jours (section 1). En effet, les pictogrammes et leur utilisation dans des tableaux de communication ou des systèmes de traduction automatique sont des exemples de CAA pouvant être conçue pour répondre aux besoins de différents utilisateurs. Ces outils exploitent divers systèmes de représentation du langage, dont les pictogrammes (section 2). Leurs caractéristiques sont détaillées. L'évaluation d'images est abordée, en particulier dans le domaine médical pour différents publics. Afin de contextualiser plus spécifiquement l'utilisation de la communication avec des pictogrammes, nous décrivons certaines méthodes d'apprentissage des images chez les personnes avec une DI. Nous montrons quels sont les enjeux d'un système de CAA dans les hôpitaux en présentant les difficultés de communication entre les médecins et les patients avec une DI pour exprimer et évaluer leur douleur, ainsi que les outils utilisés pour améliorer leur prise en charge (section 3). Nous verrons que des dispositifs avec des images existent pour les personnes allophones, aphasiques ou avec une DI, afin de les aider à se faire comprendre par le personnel soignant. Enfin, nous concluons ce chapitre en nous positionnant sur la question de l'universalité des pictogrammes (section 4). À noter que les quelques applications médicales abordées ici ne traduisent pas de phrases complètes vers une une séquence de pictogrammes et n'utilisent pas de méthodes de Traitement Automatique des Langues (TAL). Les systèmes de traduction automatique de texte vers les pictogrammes sont présentés dans le prochain chapitre.

1 Communication alternative et améliorée

Selon l'association internationale ISAAC, [1] la Communication Alternative et Améliorée (CAA) désigne l'ensemble des « moyens humains et matériels permettant de communiquer autrement ou mieux qu'avec les modes habituels et naturels, si ces derniers sont altérés ou absents ». Des millions de personnes sont concernées par ce terme générique qui couvre à la fois un aspect substitutif et supplétif. Comme il s'agit de la traduction française d'*Augmentative and Alternative Communication (AAC)*, on retrouve différentes appellations pour le terme *augmentative* (améliorée, augmentative ou augmentée, assistée, etc.), nous nous limiterons à la première. Ces méthodes sont « alternatives » lorsqu'elles remplacent totalement les moyens d'expression orale et « améliorées » lorsqu'elles permettent de suppléer les capacités déjà présentes, pouvant même aider à l'émergence de l'oralisation ou de la graphie pour certaines personnes (Cataix-Nègre, 2011). Elles visent à permettre aux personnes avec une DI d'exprimer leurs besoins tout en favorisant leur inclusion au sein de la société. La CAA est non seulement utilisée pour communiquer dans la vie quotidienne, mais aussi dans des domaines plus spécifiques comme celui de la santé, qui fait l'objet de cette thèse. Nous commençons par montrer qu'il existe plusieurs types de systèmes de CAA (section 1.1). Nous décrivons ensuite les différents utilisateurs (section 1.2).

1.1 Systèmes de communication alternative

La communication étant multimodale, il est possible de communiquer autrement que par le langage verbal. En fonction du handicap, Cataix-Nègre (2011, pp. 29-31) explique que les moyens de communication sont « multiples » et « variés » citant les « gestes, signes, images, objets, dessins, écrits... » (Cataix-Nègre, 2011, p. 22). La communication passe par différentes formes de représentation : des signes manuels, graphiques ou tangibles (Von Tetzchner et Martinsen, 2000). Elle peut parfois s'appuyer sur un dispositif simple ou plus avancé. Les aides sont complémentaires et très variées, allant du support papier (classeurs ou tableaux de communication, cahiers de vie, etc.) aux systèmes technologiques (cf. chapitre 2, section 1.2), qui évoluent très rapidement. En plus, elles nécessitent parfois un matériel spécifique pour être utilisées – comme un contacteur ou une commande oculaire pour la sélection d'items (pictogrammes, lettres, mots, phrases) –, en fonction du handicap de l'utilisateur. Certaines méthodes de communication ne requièrent toutefois aucune aide technique comme c'est le cas des langues des signes ou des expressions faciales. Dans cette thèse, nous nous intéressons aux systèmes de communication avec un support technologique proposant des signes graphiques, en particulier, des pictogrammes.

1. *International Society for Augmentative and Alternative Communication (ISAAC)* : https://www.isaac-fr.org (cf. section 1.2.2).

1.2 Différents utilisateurs

Une communication alternative peut être utilisée par des personnes ayant divers besoins. Aucun système ne peut convenir à tous les handicaps, tout comme il existe des variabilités interindividuelles au sein d'une même pathologie.[2] C'est pourquoi, l'ergonomie et l'adaptation aux besoins spécifiques de chaque utilisateur jouent un rôle majeur. Nous introduisons ici les utilisateurs que nous ciblerons pour le développement du système Text-to-Picto, à savoir, les personnes avec une déficience intellectuelle (section 1.2.1), mais également d'autres intervenants en CAA (section 1.2.2).

1.2.1 Avec une déficience intellectuelle

Beaucoup de personnes rencontrent des difficultés à communiquer verbalement, que ce soit temporairement ou tout au long de leur vie. Cette situation peut être due à différents types de handicap plus ou moins sévères, qu'il soit moteur (comme la sclérose latérale amyotrophique ou le locked-in syndrome), langagier (l'aphasie, la dysarthrie), voire intellectuel. C'est sur cette dernière catégorie que nous nous concentrerons dans cette thèse. Elle inclut entre autres les personnes avec un Trouble du Spectre de l'Autisme (TSA), le syndrome de Down (la trisomie 21), le syndrome de Rett, etc.

Les utilisateurs de CAA ont souvent un polyhandicap, c'est-à-dire un handicap grave à expressions multiples, ce qui influence le choix de l'aide technique qui doit toujours être adaptée en fonction de l'éventuelle évolution de la pathologie et de la spécificité des besoins de l'utilisateur. Pour toutes personnes présentant des troubles sévères du langage oral, une première étape consiste à faire reconnaître officiellement son handicap et à procéder à une évaluation pour faire le point sur ses capacités de communication expressives et réceptives (Crunelle, 2018).

La Déficience Intellectuelle (DI) peut concerner, à des degrés divers, le fonctionnement intellectuel (la mémoire, l'attention, le langage, etc.) et les troubles du comportement. Les personnes avec une DI éprouvent plus de difficultés à raisonner, communiquer des concepts abstraits, comprendre et traiter de nouvelles informations (Tuffrey-Wijne et McEnhill, 2008). Elles ont souvent des besoins complexes de communication et présentent des troubles du développement ou d'apprentissage. Comme l'avait déjà noté Sevens (2018), la DI est à distinguer du handicap cognitif qui est un terme plus large. Il ne faut cependant pas non plus confondre la DI avec les maladies mentales qui apparaissent après la période de développement (> 18 ans).

La terminologie relative aux personnes avec une DI a évolué au cours du temps. On a retrouvé diverses appellations discriminantes dans la littérature : malade mental, aliéné, idiot, stupide, etc. Cependant, la prise en charge de ces patients a elle aussi

2. Par exemple, les personnes avec un Trouble du Spectre de l'Autisme ont des profils très divers (verbal ou non, avec différents niveaux de compréhension du langage, etc.). Tous les autistes n'ont donc pas les mêmes besoins, ni les mêmes préférences.

évoluée. Les personnes avec une DI ont progressivement reçu une « éducation sensorielle » (cf. la citation de Bernard Dan ci-dessous), [3] dont certains éléments sont encore enseignés aujourd'hui, notamment lorsque des professionnels leur apprennent à utiliser un outil de CAA pictographique, comme nous le verrons par la suite.

> “ Le vocabulaire médical utilisé aux XIXe et XXe siècles pour caractériser la [DI] comprenait des termes comme *idiotie*, *imbécilité*, *débilité mentale* ou encore *crétinisme*, rapidement devenus péjoratifs. Les personnes affectées étaient considérées comme *arriérées* ou *attardées* dans le développement de leur intelligence et leurs capacités d'apprentissage. Des efforts de plus en plus fructueux ont été déployés pour apprendre à ces enfants à devenir plus autonomes, en stimulant leurs sens, car on postulait un rapport entre les sens et le fonctionnement intellectuel. Des manuels d'éducation sensorielle visaient à renforcer l'audition, la vision, le goût, l'odorat et la coordination manuelle dans les activités quotidiennes au sein d'établissements spécialisés. Cet enseignement insistait sur l'imitation, la mémoire et la capacité à généraliser les apprentissages. Parmi les méthodes pédagogiques initiées alors, plusieurs restent d'actualité. ”

Ce changement de considération à l'égard des personnes avec une DI s'observe aussi dans la littérature scientifique. En comparant l'évolution des recherches qualitatives basées sur des interviews, Beail et Williams (2014) ont constaté que les chercheurs ont souvent hésité à impliquer les personnes avec une DI en tant que participants. Elles étaient encore très rarement interviewées dans les années 1980 alors que les neurotypiques étaient déjà considérés comme des « co-chercheurs » dans les études qualitatives de sociologie à partir des années 1970. Aujourd'hui, les chercheurs et associations insistent sur l'importance d'impliquer les personnes en situation de handicap (Schwab, 2021), y compris celles avec une DI quel que soit le domaine (Sevens, 2018; Beail et Williams, 2014; Biotteau *et al.*, 2013; Inclusion Europe, 2009).

Dans cette thèse, nous avons choisi de cibler les adultes verbaux – qui utilisent le langage oral – avec une DI légère (cf. chapitre 6). Impliquer des personnes avec une DI nécessite du temps et certaines adaptations que nous mentionnerons dans la partie concernée. Le nombre de participants avec une DI peut varier d'une étude à l'autre, en particulier quand on utilise une approche qualitative (Beail et Williams, 2014).

3. Citation reprise d'un professeur de l'Université libre de Bruxelles (ULB) provenant d'un panneau de l'exposition « *Art of Difference* » organisée par l'ULB au BOZAR de Bruxelles en août 2018.

1.2.2 Autres personnes concernées

Lorsque l'on aborde la CAA, il est essentiel de prendre en compte différents intervenants qui accompagnent les personnes avec une DI. Il y a tout d'abord les familles, les aidants et les amis, qui favorisent le bain langagier quotidien, nécessaire à la personne avec une DI dès son plus jeune âge pour simuler et développer ses compétences tant sur le versant réceptif qu'expressif (Cataix-Nègre, 2011). Nous verrons en effet que des aménagements peuvent être mis en place afin d'encourager l'utilisation et l'apprentissage des images ou pictogrammes. Il convient toutefois de noter que souvent les personnes avec une DI ne vivent pas à la maison toute la semaine, mais dans des centres spécialisés. Elles côtoient ainsi différents professionnels, professeurs et éducateurs qui doivent communiquer et comprendre leurs besoins. Ces intervenants jouent tous un rôle actif en tant que « partenaires de communication ».

La CAA est un domaine multidisciplinaire réunissant de nombreuses personnes qui participent à son développement. Beukelman et Mirenda (2017) citent entre autres les développeurs de logiciels, les éducateurs, les ingénieurs, les linguistes, les ergothérapeutes, les psychologues et les logopèdes. [4] Ce sont souvent des professionnels qui guident l'entourage de la personne avec une DI vers le système de CAA le plus adapté à ses besoins. En Belgique, il existe quelques associations ou centres qui donnent des conseils quant au choix de la CAA et qui proposent des formations à destination des différents intervenants qui encadrent la personne avec une DI. Nous pouvons citer l'association Comalso (COMmunication Alternative SOlution), le CRETH (Centre de Ressource et d'Evaluation des Technologies pour les personnes Handicapées) ou le SATIH (Service d'Accompagnement Technologie, Informatique et Handicap).

Des organisations internationales promeuvent l'usage de toute forme de CAA partout dans le monde. En 1983, l'*International Society for Augmentative and Alternative Communication (ISAAC)* a ainsi été créée pour améliorer la vie des personnes ayant des besoins complexes de communication. [5] Elle soutient la recherche autour de la CAA. Depuis 1985, elle publie le journal *Augmentative and Alternative Communication (AAC)*. Les membres de cette organisation sont des personnes qui utilisent la CAA, leur famille, des thérapeutes, des enseignants, des étudiants, des médecins, des chercheurs, des organisations et des entreprises qui fabriquent des aides à la communication. En 2024, une soixantaine de pays étaient représentées, réparties en 17 sous-groupes correspondant à un ou plusieurs pays, dont l'ISAAC Francophone, qui rassemble la France, la Belgique, le Luxembourg, la Suisse et le Québec. L'association a pour but de faire reconnaître la CAA comme une communication à part entière.

4. Au Canada, il existerait d'ailleurs un métier d'assistant de communication (Cataix-Nègre, 2011). L'assistant aide à interpréter ce que la personne en situation de handicap essaie de communiquer avec ou sans son système de CAA. L'association APF France handicap tente de faire reconnaître ce métier.

5. https://isaac-online.org

2 Pictogrammes

Dans cette section, nous proposons de commencer par une définition des pictogrammes et une taxonomie (section 2.1). Nous présentons différents ensembles de pictogrammes conçus pour les personnes avec une DI (section 2.2), mais également les caractéristiques des pictogrammes (section 2.3), qui peuvent être utilisés dans les systèmes de CAA, y compris ceux qui traduisent du texte en pictogrammes. Nous verrons que des études ont été réalisées pour évaluer des images médicales auprès de diverses populations (section 2.4). Enfin, nous chercherons à comprendre comment les personnes avec une DI apprennent à communiquer avec les pictogrammes (section 2.5), mais aussi quels sont leurs effets sur la communication (section 2.6).

2.1 Définition et taxonomie

Un moyen de communication souvent recommandé aux personnes avec une DI et des troubles de la parole ou du langage – ne sachant pas toujours parler, lire ou écrire – consiste à développer la communication non verbale et à recourir aux images.

En CAA, un pictogramme représente graphiquement un concept (le référent). Il peut s'agir d'une action, un objet, une personne, etc. (Pereira *et al.*, 2022). Dans la littérature scientifique, un pictogramme est parfois appelé un symbole ou une icône. Dans la vie de tous les jours, ils sont omniprésents sur les emballages, dans nos ordinateurs et les lieux publics. Les pictogrammes permettent de transmettre une information instantanément à un grand nombre de personnes ayant diverses origines. Ces représentations graphiques visent à soutenir ou remplacer des instructions textuelles écrites.

Il est possible d'identifier différentes stratégies de représentation. Nakamura et Zeng-Treitler (2012) ont par exemple décrit une classification sémantique des pictogrammes, mais hors contexte de la CAA. En se basant sur une analyse de 846 pictogrammes utilisés dans le domaine médical, ils ont proposé une taxonomie qui classe les relations entre les deux éléments constitutifs des pictogrammes, à savoir la représentation graphique et le référent. Leur taxonomie a notamment été reprise par Sevens (2018) pour présenter les pictogrammes Sclera (cf. section 2.2.5). Nakamura et Zeng-Treitler (2012, pp. 542-548) ont relevé trois types de représentation, qui peuvent être combinés au sein d'un seul pictogramme :

A) **Les représentations conventionnelles** où la relation entre le pictogramme et son référent est établi à travers un contrat social entre le lecteur et le décodeur. Il n'y a pas de réelle connexion entre la représentation et son référent. Cette convention arbitraire peut être abstraite, concrète ou transposée (Sevens, 2018, p. 20). Le symbole pour désigner la femme ♀ est un exemple de ces représentations.

B) **Les représentations non conventionnelles** qui reposent sur une similarité visuelle et qui sont utilisées pour désigner les noms ou les objets physiques avec une forme facilement identifiable. Il existe un lien entre la représentation et son référent qui transmet des caractéristiques visuelles au pictogramme. Le choix de l'horloge 🕒 pour symboliser le temps est un exemple.

C) **Les représentations indirectes ou médiées** basées sur l'association sémantique. Plusieurs sous-types sont observés, que nous illustrons (Figure 1.1) avec des pictogrammes Arasaac (cf. section 2.2.6) : (1) la comparaison ou le contraste (la maman par rapport au bébé), (2) l'exemplification (différentes sortes de médicaments), (3) le rétrécissement sémantique (examiner un patient), (4) la décomposition physique (des éprouvettes pour désigner le laboratoire), (5) la décomposition temporelle ou l'instantané pour les verbes (le mouvement de tomber), (6) le langage corporel (les expressions faciales pour un état comme fatigué), (7) la métaphore (la pièce de puzzle pour le TSA) et (8) la contiguïté comprenant (8.1) la contiguïté physique (une flèche indique la cicatrice sur un bras), (8.2) le conteneur (le tube et la boîte de la pommade), (8.3) la source (le sang de la main), (8.4) l'utilisation (la piqûre faite dans le bras), (8.5) l'outil (le radiologue est le médecin avec une radio), (8.6) la cause ou l'effet (la croix pour la constipation), ainsi que (8.7) l'objet (la cigarette – barrée – pour – l'interdiction de – fumer).

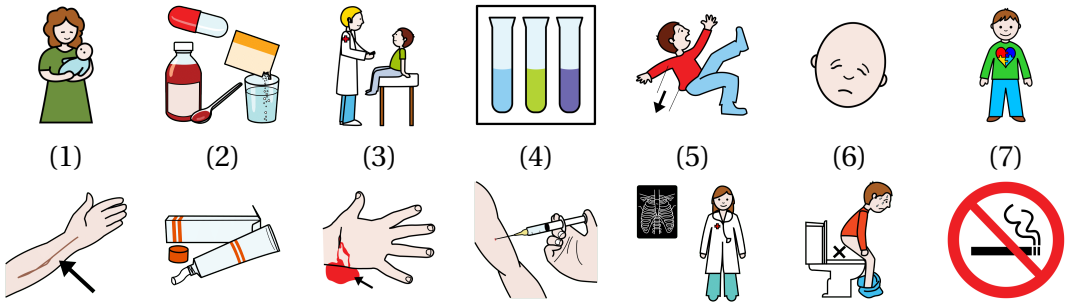


FIGURE 1.1 – Représentations indirectes des pictogrammes Arasaac : exemple des mots (1) « maman », (2) « médicaments », (3) « examiner », (4) « laboratoire », (5) « tomber », (6) « fatigué », (7) « autiste », (8.1) « cicatrice », (8.2) « pommade », (8.3) « sang », (8.4) « piqûre », (8.5) « radiologue », (8.6) « constipation » et (8.7) « interdiction de fumer »

2.2 Ensembles de pictogrammes

Les pictogrammes sont généralement regroupés dans un ensemble conçu par un même groupe de personnes selon certains principes de composition définis préalablement. Ces règles de conception permettent d'assurer de la consistance entre les images

et ainsi d'améliorer leur apprentissage. Nous pouvons citer plusieurs ensembles destinés spécifiquement aux personnes en situation de handicap – contrairement aux emojis, aux symboles de signalisation routière, aux pictogrammes sur les produits chimiques, etc. Certains sont gratuits, d'autres nécessitent une licence pour être utilisés par une ou plusieurs personne(s) sur différents supports. Les technologies de CAA et beaucoup de grands ensembles de pictogrammes ont été conçus durant les années 1980 (Cataix-Nègre, 2011).

2.2.1 Bliss

Les symboles Bliss ont été développés pendant la Seconde Guerre mondiale par Karl Kasiel Blitz avec l'objectif de créer un langage international, appelé *Semantography* (Bliss, 1965). En 1971, Shirley McNaughton a adapté le Bliss afin d'en faire le premier outil de CAA permettant aux personnes en situation de handicap moteur de communiquer en désignant des images (Cataix-Nègre, 2011, pp. 43-44). Il regroupe plus de 4 500 symboles relativement abstraits comme le montre la phrase (Figure 1.2) : « *"Help me!" he said. "The baby is sick! Can you come here?"* » ("Aide-moi!" dit-il. "Le bébé est malade! Peux-tu venir ici?). Ce langage idéographique nécessite un certain apprentissage [6] et s'inspire de la combinaison des caractères chinois (Tenny, 2016). Son développement est contrôlé par la *Blissymbolics Communication International (BCI)*.

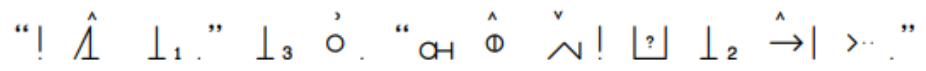


FIGURE 1.2 – Symboles Bliss© : exemple de la phrase « *"Help me!" he said. "The baby is sick! Can you come here?"* », reproduite de Blissymbolics Communication International (2004, p. 7)

2.2.2 Widgit

Les pictogrammes Widgit ont été utilisés aux États-Unis au début des années 1970. Ils avaient pour but de soutenir la communication en face à face de ceux qui parlent peu ou pas du tout. [7] Cet ensemble couvre un vocabulaire anglais de 55 000 mots et propose plus de 20 000 pictogrammes représentant un concept unique. Ils sont tous disponibles en noir et blanc, ainsi qu'en couleurs. La Figure 1.3 représente la phrase « *Symbols help communicate ideas and information* » (les pictogrammes aident à communiquer des idées et des informations). En 2021, chaque pictogramme a été

6. Beukelman et Mirenda (2017, p. 34) citent des études montrant que les symboles Bliss sont « moins transparents et plus difficiles à apprendre et à mémoriser que d'autres types de pictogrammes ». Le Bliss est néanmoins encore utilisé. Un système de traduction automatique du Bliss vers l'anglais a d'ailleurs été développé récemment (Sohail et Traum, 2019).

7. https://www.widgit.com (d'où provient la Figure 1.3). La traduction en pictogrammes de chaque mot du site est affichée en infobulle.

revu afin de les rendre plus représentatifs de la culture et de la technologie du monde dans lequel nous vivons. Il existe aussi une extension pour le domaine de la santé appelée Widgit Health. [8] Ces pictogrammes sont payants bien que quelques supports pour la communication médicale soient disponibles gratuitement sur le site.

FIGURE 1.3 – Pictogrammes Widgit : exemple de la phrase « *Symbols help communicate ideas and information* »

2.2.3 Minspeak

Utilisé dans les années 1980, le système de pictogrammes polysémiques Minspeak a été conçu par le linguiste Bruce Baker (1982). Selon l'association belge Comalso, il nécessite que l'utilisateur ait à la fois un certain niveau mental ainsi qu'une bonne mémoire visuelle. Le pictogramme lexical « pomme » peut avoir différents sens (« fruit », « rouge », « faim », etc.) [9] en fonction du contexte (*Semantic Compaction*), c'est-à-dire du pictogramme avec lequel il est combiné (Beukelman et Mirenda, 2017). Pour former le sens de « *eat* » ou « *ate* », l'utilisateur doit combiner le pictogramme représentant une « pomme » avec celui du « verbe » ou du « verbe +ed » (Tenny, 2016). Les icônes verbales et adjectivales peuvent représenter n'importe quel verbe ou adjectif comme le montrent les représentations des mots (1) « boire », (2) « dormir », (3) « soif » et (4) « fatigué » formés à partir des pictogrammes « jus » et « lit » (Figure 1.4).

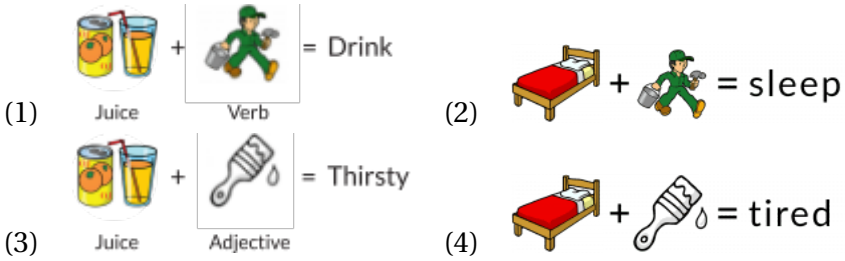


FIGURE 1.4 – Pictogrammes Minspeak® : exemple des mots (1) « *drink* », (2) « *sleep* », (3) « *thirsty* » et (4) « *tired* »

8. https://widgit-health.com
9. http://www.minspeak.com (d'où provient la Figure 1.4).

2.2.4 Picture Communication Symbols

Les pictogrammes PCS (*Picture Communication Symbols*) développés par Mayer-Johnson et commercialisés aujourd'hui par l'entreprise Tobii Dynavox [10] sont apparus aux États-Unis en 1980. Selon Beukelman et Mirenda (2017, p. 33), il s'agit de l'ensemble de pictogrammes le plus utilisé pour la communication partout dans le monde. Il contient plus de 45 000 pictogrammes payants, qui représentent « des mots, des phrases et des concepts sur un large éventail de sujets ». Comme le montre la Figure 1.5, il en existe plusieurs types (*Classic*, *High Contrast*, *ThinLine* et *InContext*) pour répondre aux besoins de différents utilisateurs (adolescents, adultes, avec ou sans une déficience visuelle, etc.).

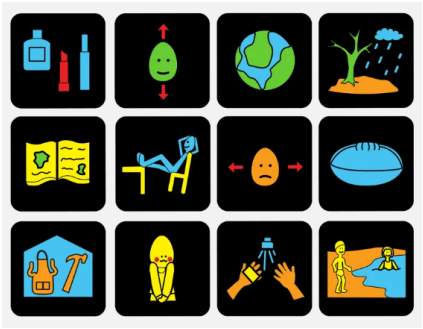
(1) (2) 

FIGURE 1.5 – Pictogrammes PCS® : exemple de symboles de type (1) « *Classic* » et (2) « *High Contrast* »

2.2.5 Sclera et Beta

Les pictogrammes Sclera [11] et Beta (*BEeldTAal*) [12] ont été conçus en Belgique (Flandre). Ils sont notamment utilisés dans le système Text-to-Picto (Sevens, 2018), que nous adapterons. Sclera comprend 13 000 pictogrammes à haut contraste, blancs sur fond noir (excepté pour la représentation de certains concepts comme les couleurs ou l'interdiction). Beta inclut 3 500 pictogrammes. Ils existent en couleurs, ainsi qu'en noir et blanc. Selon Sevens (2018, pp. 37-39), Sclera représente beaucoup de noms, de verbes, des concepts simples et complexes (par exemple, « sortir le chien » ou « pomme et poire ») ; tandis que Beta propose des concepts simples, ainsi que diverses catégories morphosyntaxiques (adverbes, prépositions, etc.). Les pictogrammes Sclera sont plus génériques que ceux en Beta (Figure 1.6). Les noms des pictogrammes de ces deux ensembles sont disponibles en néerlandais, français et anglais.

10. https://us.tobiidynavox.com/products/picture-communication-symbols-pcs (d'où provient la Figure 1.5).
11. http://www.sclera.be ; l'association Sclera vzw les publie sous licence Creative Commons 2.0.
12. http://www.betasymbols.com

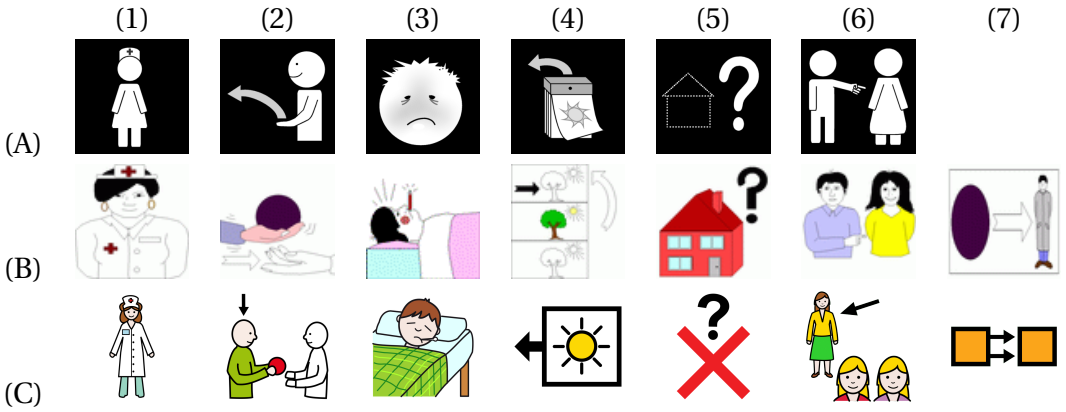


FIGURE 1.6 – Pictogrammes (A) Sclera, (B) Beta et (C) Arasaac : exemple des mots (1) « infirmière », (2) « donner », (3) « malade », (4) « hier », (5) « où », (6) « elle » et (7) « pour »

2.2.6 Arasaac

Les pictogrammes Arasaac (*ARAgoneSe portal of Augmentative and Alternative Communication*) sont créés par Sergio Palao depuis 2007, à l'initiative du Centre Aragonais de Technologies pour l'EDUcation (CATEDU). [13] Il s'agit d'un ensemble regroupant plus de 13 000 pictogrammes disponibles en couleurs (Figure 1.6) ou en noir et blanc, qui sont de plus en plus utilisés en CAA comme le montrent les différents supports partagés librement sur le site. De nouveaux pictogrammes sont encore ajoutés. Leurs noms sont traduits par une équipe de bénévoles (souvent des logopèdes) dans une trentaine de langues, dont le français, le néerlandais, le russe, le chinois, etc. Il est possible de les personnaliser directement sur le site (couleurs, légendes, indications pour la catégorie du concept, le pluriel, le temps pour les verbes, etc.), avant de les télécharger individuellement. Les développeurs d'Arasaac ont conçu une *Application Programming Interface (API)* permettant d'accéder aux métadonnées des images au format *JavaScript Object Notation (JSON)*. Elles sont régulièrement mises à jour.

En pratique, les personnes en situation de handicap utilisent souvent des pictogrammes de différents ensembles. Comme le notent Cataix-Nègre (2011), Von Tetzchner et Martinsen (2000), utiliser plusieurs banques de pictogrammes peut s'avérer utile dans certaines situations ; quand il n'existe pas de représentation d'un concept ou que celle-ci ne convient pas à l'utilisateur. Les applications récentes destinées au grand public intègrent des fonctionnalités permettant d'aller sur Internet ou de prendre des photos (par exemple, de personnes) pour ajouter des images, les renommer manuellement et ainsi enrichir le vocabulaire de base en le personnalisant en fonction des besoins et préférences des utilisateurs. Les logiciels utilisés pour construire des tableaux de communication au format papier (Boardmaker®, Picto-Selector, etc.) offrent également la possibilité de combiner des pictogrammes de plusieurs ensembles.

13. https://arasaac.org ; ces symboles pictographiques sont la propriété du Gouvernement d'Aragon et sont l'œuvre de Sergio Palao pour Arasaac qui les distribue sous licence Creative Commons.

2.3 Caractéristiques générales

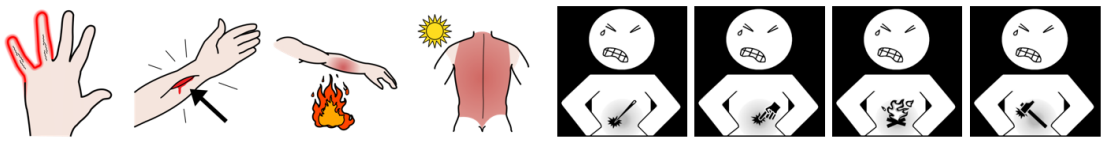
Dans les banques d'images, les pictogrammes ont généralement un nom (de fichier) qui désigne le sens qu'a voulu donner son (ou ses) concepteur(s). Il peut s'agir d'un lemme ou d'une expression multi-mots représenté.e par un pictogramme simple (en général une catégorie morphosyntaxique comme le nom « température ») ou complexe (plusieurs catégories morphosyntaxiques, les verbe(-article)-nom « prendre la température »). En Arasaac, chaque pictogramme peut avoir plusieurs noms (synonymes). Le pictogramme 15992 a dix noms {accompagnant éducatif et social, aidant, aide de vie scolaire, aide-médicopédagogique, aide-médicopsychologique, AMP, auxiliaire de vie, AVS, moniteur, tierce-personne}. Deux pictogrammes avec un nom de fichier identique – et un identifiant numérique différent pour Arasaac – peuvent être différents. C'est le cas des mots polysémiques ; une « souris » peut être représentée par l'animal ou l'accessoire de l'ordinateur, deux concepts différents (Vaschalde, 2018).

Un même concept peut également être représenté par différentes images (Nakamura et Zeng-Treitler, 2012). Le « non » peut être représenté par une croix ✕ rouge ou un visage pas content ☹ avec un mouvement horizontal ↔ (Norré *et al.*, 2021b, 2020). [14] Et inversement, il est possible qu'une représentation graphique exprime plusieurs sens – plus ou moins – différents (Nakamura et Zeng-Treitler, 2012). Dans la Figure 1.1 (3), le pictogramme « examiner » peut aussi désigner le verbe « ausculter ». Des représentations différentes d'un même concept peuvent être dues : a) à des différences culturelles (Mihalcea et Leong, 2008) ; b) au niveau de connaissance des utilisateurs (Christensen *et al.*, 2017). Ainsi, a) une « maison » ne sera pas représentée de la même façon en Asie et en Europe ; b) l'image d'un « défibrillateur » ne peut être reconnue – et donc utilisée – que si on connaît déjà le référent. Une certaine phase d'apprentissage est nécessaire et c'est la raison pour laquelle nous l'aborderons dans la section 2.5.

Nous avons aussi constaté que certains pictogrammes étaient polysémiques lorsqu'ils sont utilisés dans des contextes précis, par exemple, pour répondre à des questions médicales (Norré *et al.*, 2021b, 2020). En Arasaac (Figure 1.7a), ils représentent toujours un symptôme (picotement, coupure, brûlure) sur une partie du corps spécifique (main, bras, dos), ce qui pourrait limiter leur utilisation dans certains cas. En effet, les personnes questionnées dans ces études ont décrit ces images avec plus ou moins de précision lorsqu'on leur en demandait la signification. Autrement dit, il pourrait arriver que des patients ne choisissent pas le pictogramme « brûlure » s'ils ont une

14. Nous avions d'ailleurs remarqué que la représentation du « oui » et du « non » était une question souvent débattue par les logopèdes et les ergothérapeutes lorsqu'une forme de CAA est envisagée pour un patient (Norré *et al.*, 2020). Dans son livre, Cataix-Nègre (2011, p. 122) avait par exemple soulevé la question suivante : « comment en effet demander sans malaise à Kim, qui sort de sa consultation chirurgicale, alors qu'elle n'a pour répondre précisément que deux faciès – content et triste – pour exprimer le "oui" et le "non", si elle a bien compris qu'elle va devoir subir une opération du bassin, porter un plâtre, rester trois semaines sur un chariot plat qu'elle ne pourra pas diriger ? ».

brûlure à un endroit non représenté par l'image. La couverture médicale est quelque peu limitée car il n'existe pas de pictogrammes représentant uniquement une brûlure.



(a) Arasaac : exemples pour (1) « picotement », (2) « coupure » et (3/4) « brûlure »

(b) Sclera : exemples pour (1) « picotement douleur », (2) « douleur pression », (3) « sensation de brûlure » et (4) « douleur lancinante »

FIGURE 1.7 – Pictogrammes pour la représentation d'un type de douleur

En Sclera (cf. section 2.2.5), le type de douleur (picotement, pression, brûlure, lancinant) est représenté par un personnage unisexe grimaçant avec un symbole spécifique (aiguille, doigt, feu de camp, marteau) pour illustrer le symptôme qui est toujours situé au niveau du ventre (Figure 1.7b). En évaluant la compréhension de ces quatre pictogrammes auprès de personnes avec une DI, de Knegt *et al.* (2016b) avaient noté que cela pouvait être difficile de comprendre que ces pictogrammes puissent également représenter d'autres localisations douloureuses que les douleurs abdominales.

D'après le centre belge de la CAA, il existe des systèmes de communication polysémiques où chaque représentation graphique peut avoir plusieurs sens en fonction de son association avec une ou plusieurs autre(s) image(s). En Bliss (cf. section 2.2.1), « maison », « médical » et « livre » sont des caractères, des unités de sens indivisible qui peuvent être combinés pour former des mots avec un sens lexical spécifique comme « hôpital » et « bibliothèque » (Figure 1.8). Nous avions déjà vu un autre exemple de système polysémique lorsque nous avons présenté les pictogrammes Minspeak (2.2.3). Selon Comalso, la Belgique utiliserait principalement la monosémie en CAA.

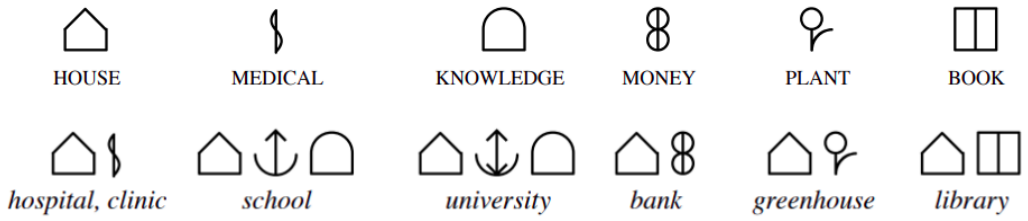


FIGURE 1.8 – Symboles Bliss© : exemple de caractères et de mots, reproduite de Blissymbolics Communication International (2004, pp. 2-3)

Certains concepts n'existent pas dans tous les ensembles ou avec des noms différents (Norré *et al.*, 2020). C'est, par exemple, le cas des types de douleur en Sclera, le pictogramme « douleur irradiant » n'a pas d'équivalent en Arasaac et « médicaments gouttes » est appelé « compte-gouttes » en Arasaac (Figure 1.9). Les ensembles de pic-

togrammes ont aussi une couverture grammaticale plus ou moins variable, en particulier concernant le vocabulaire des classes fermées. Les pictogrammes pour les mots de fonction (conjonctions, prépositions, etc.) sont abstraits car ils sont difficilement représentables. Que ce soit en Sclera ou Arasaac (Figure 1.9), ces mots sont souvent représentés avec des flèches et des formes géométriques (triangles, ronds, carrés, etc.). Nous avons mentionné la comparaison dans les représentations indirectes (Nakamura et Zeng-Treitler, 2012). Les pictogrammes se comprennent aussi les uns par rapport aux autres et la couleur a également une signification. Par exemple, l'intérieur des formes placées en avant-plan ou en haut est blanche ou rouge pour les mots « devant » ou « au-dessus » – et inversement pour « derrière » ou « sous » – (Figure 1.9).

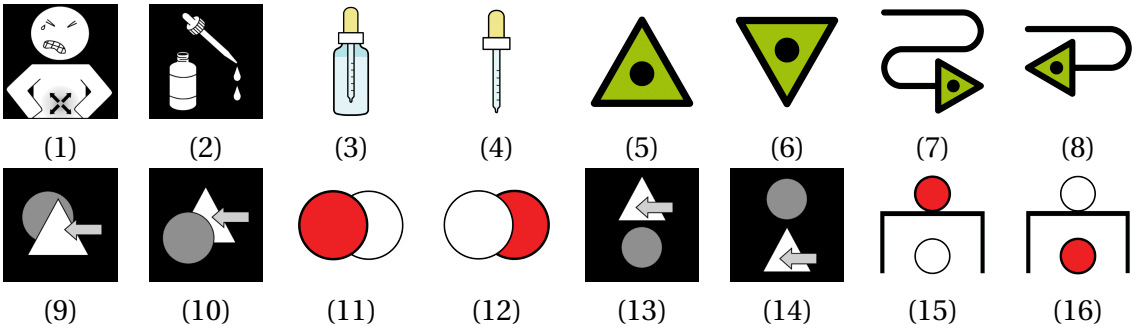


FIGURE 1.9 – Pictogrammes Sclera (1-2/9-10/13-14) et Arasaac (3-8/11-12/15-16) : exemple des mots (1) « douleur irradiant », (2) « médicaments gouttes », (3/4) « compte-gouttes », (5) « et », (6) « ou », (7) « bien que », (8) « mais », (9/11) « devant », (10/12) « derrière », (13/15) « au-dessus » et (14/16) « sous »

Certains ensembles de pictogrammes distinguent le nombre (singulier/pluriel). Il existe par exemple un pictogramme Arasaac pour les mots « œil » ou « médicament » et un autre pour les « yeux » ou les « médicaments ». Cette distinction n'est toutefois pas systématique – ce n'est pas le cas pour « oreille(s) », « bandage(s) », etc. Le concept au pluriel désigne parfois une catégorie d'éléments, c'est le cas des « animaux », « animaux volants », « animaux marins », « légumes », etc. Le pluriel peut aussi être marqué par un qualificatif sur le pictogramme en haut à droite (par exemple, un '+' en Arasaac et un '++' avec Widgit). Concernant la distinction du genre, Arasaac propose souvent un équivalent masculin, féminin et neutre (unisexe) pour les pictogrammes incluant des personnages humains, que ce soit pour les noms (« électroencéphalogramme »), les adjectifs (« timide »), les verbes (« boire »), etc.

Selon Fuller et Lloyd (1991, p. 215), plusieurs recherches ont identifié différentes variables pouvant influencer l'apprentissage des symboles utilisés en CAA, dont la complexité, le caractère concret/abstrait, la fréquence d'occurrence, l'iconicité, l'imagerie et les catégories morphosyntaxiques. La terminologie autour de l'iconicité est floue. Les auteurs ont relevé de nombreux termes utilisés dans la littérature scientifique pour décrire la relation visuelle entre les symboles et leurs référents (« *"arbitrary", "enactive",*

"metonymic", "opaque", "representativeness", "symbolic", "translucent", and "transparent" »). La compréhension des pictogrammes dépend de l'iconicité, ainsi que d'autres facteurs tels que le contexte, l'apprentissage des images et de leurs référents.

2.4 Évaluation des images

Des travaux se sont concentrés sur la compréhension des images. Autrement dit, ils ont vérifié si différentes populations pouvaient facilement deviner leur sens en l'absence de leur référent. Il s'agit donc d'évaluer la notion de transparence des symboles graphiques telle que définie par Fuller et Lloyd (1991, p. 217) :

> “ A transparent symbol is one in which the shape, motion, or function of the referent is depicted to such an extent that the meaning of the symbol can be readily guessed in the absence of the referent. Transparency is operationally defined as *guessability*. ”

Les recherches sur la transparence des ensembles de pictogrammes de CAA ne concernent pas notre public cible, ni le domaine qui nous intéresse. En effet, il n'y a pas, à notre connaissance, d'études menées auprès d'adultes avec une DI sur la compréhension de pictogrammes médicaux de CAA, isolés et évalués indépendamment d'un outil. C'est pourquoi, nous avons décidé de mentionner ici quelques recherches sur l'évaluation de pictogrammes médicaux auprès de personnes sans handicap. Nous donnerons des exemples d'évaluation de séquences de pictogrammes générées avec divers systèmes de traduction vers les pictogrammes dans le chapitre suivant.

Des études ont expérimenté des protocoles participatifs pour la conception et l'évaluation de pictogrammes par des personnes d'origines diverses. Kim *et al.* (2009) ont testé la reconnaissance du sens de 40 pictogrammes médicaux auprès de 37 patients. Ils ont ainsi pu obtenir 719 réponses. Le crowdsourcing a été utilisé par Christensen *et al.* (2017), afin de concevoir et d'évaluer un grand nombre d'images médicales grâce à un jeu appelé Doodle Health. Les participants recevaient un mot (par exemple « ambulance ») ou une phrase médicale qu'ils devaient dessiner. D'autres joueurs devaient ensuite deviner le sens. Les dessins produits étaient destinés à réduire les disparités dans la communication médicale. Dans les journaux médicaux (Barros *et al.*, 2014), beaucoup d'études ont évalué la compréhension des pictogrammes pharmaceutiques utilisés pour améliorer la littératie en santé, notamment les pictogrammes USP (cf. section 3). Ces travaux ont concerné différents publics : les populations peu alphabétisées en Afrique du Sud (Dowse et Ehlers, 2001), les enfants (Hämeen-Anttila *et al.*, 2004), les personnes âgées (Faustino *et al.*, 2020), etc. Beaucoup de ces méthodes d'évaluation ne sont pas adaptées aux personnes avec une DI. C'est aussi le cas de la norme ISO 9186-1 sur l'évaluation de la compréhension des symboles graphiques.

2.5 Apprentissage des pictogrammes

Entre 2 et 5 millions de personnes vivant au sein de l'Union Européenne avec un trouble de la parole, un handicap langagier ou cognitif pourraient en fait bénéficier de l'usage de pictogrammes ou de symboles associés à du texte (Keskinen *et al.*, 2012). Tout comme pour le langage, la lecture ou l'écriture, l'utilisation régulière des pictogrammes nécessite en réalité un long apprentissage qui sera abordé ici.

Le vocabulaire d'un système de CAA doit être adapté aux capacités individuelles des utilisateurs (Cataix-Nègre, 2011), c'est-dire ni trop simple, ni trop compliqué afin de favoriser un usage régulier. La modélisation (cf. infra) est une technique d'apprentissage qui permet d'adapter le vocabulaire. D'après l'association Comalso, certaines personnes n'utiliseront jamais de mots liens comme les articles. Le choix des pictogrammes doit aussi être adapté au type de handicap et à sa sévérité. Pour une personne ayant aussi des troubles perceptifs, il faut tenir compte des couleurs et des contrastes ; les pictogrammes blancs et noirs de l'ensemble Sclera pourront être plus adaptés.

Dans son livre, Cataix-Nègre (2011, pp. 90-91) présente divers exercices simples sous la forme de jeux qui visent à évaluer certaines compétences des personnes en situation de handicap avant de commencer à proposer des pictogrammes pour apprendre à s'exprimer. Différentes capacités sont ainsi évaluées au cours de ces exercices : les connaissances lexicales et la dénomination d'images, l'appariement, la catégorisation, l'évocation, l'association d'idées ou encore la symbolisation. Il existe plusieurs techniques ou dispositifs qui permettent de choisir et favoriser la mise en place de l'outil de CAA – avec des images – le plus adapté à l'utilisateur.

Le CHESSEP (Communication Handicap complexe : Evaluer, Situer, S'adapter, Elaborer un projet Personnalisé) est une démarche qui est fondée sur une évaluation des besoins, forces et faiblesses des personnes en situation de handicap complexe, ainsi que sur une adaptation à leur environnement afin d'élaborer un projet individualisé (Crunelle, 2018). Les aidants et les professionnels de la santé remplissent deux grilles d'observation composées chacune de 10 items correspondant à un niveau de communication réceptive ou expressive. La « communication totale » est d'abord évaluée. Les niveaux 4 à 7 correspondent respectivement au « début de symbolisme », au « langage signifiant », au « langage verbal » et au « début de syntaxe », le dernier niveau étant celui de l' « accès à l'écrit ». Au niveau 4, sont notamment évaluées les capacités de pointage qui permettront de faciliter un projet de CAA et l'utilisation des images.

Selon le programme CHESSEP, le patient pourrait débuter son apprentissage des symboles lorsqu'il commence à pointer ce qu'il veut, et ce, quel que soit son âge. Cette étape est importante car cela signifie que la personne tente d'exprimer un choix (Crunelle, 2018), une intention (Cataix-Nègre, 2011). Il faut ensuite identifier les cibles, privilégier celles que la personne apprécie et qui sont alors des objets réels, avant de pro-

gressivement les remplacer par des objets miniatures avec leur photo, leur image ou pictogramme. Il peut s'agir de représentations graphiques sur différents supports intégrés au quotidien. Par exemple, un set de table avec des photos d'aliment (céréale, sel, etc.) que la personne peut demander quand elle en a envie. Pour encourager l'utilisation des pictogrammes, ces supports doivent rester disponibles tout le temps et être adaptés en fonction de la situation de communication. Il est recommandé aux utilisateurs de CAA d'emporter leur outil de communication dans tous leurs déplacements.

Une fois l'objet désigné, on demande à la personne si c'est bien ce qu'elle veut et on montre le substitut (photo ou image), avant de donner l'objet réel voire les substituts. Deux ou plusieurs objets sont ensuite présentés afin de renforcer sa capacité à faire des choix avec des questions fermées (« Tu préfères une pomme ou un biscuit ? Montre-moi »), puis ouvertes (« Qu'est-ce que tu veux manger ? ») en travaillant sur l'imitation et la relation cause/effet (Crunelle, 2018). Selon la logopède, les images doivent toujours être nommées de la même façon et bien représentatives pour permettre la généralisation des concepts. La personne apprend à associer de nouveaux mots (noms, verbes/actions, émotions, etc.) aux pictogrammes et à les catégoriser afin de pouvoir les désigner spontanément dans son cahier de communication personnel.

D'après l'ergothérapeute Cataix-Nègre (2011), il ne faut pas hésiter à créer des conditions pour que le patient utilise le plus possible la CAA au cours de la journée. Ainsi, elle propose des aménagements pour différentes situations de communication (demander un objet ou une action, faire un choix, demander de l'aide, refuser, etc.) tout en mettant à disposition « quelques supports de référence au signifié attendu » comme des pictogrammes (Cataix-Nègre, 2011, p. 131). Il est possible de faire en sorte que le patient demande qu'on lui serve encore un verre si celui-ci n'était pas assez rempli ou de l'obliger à faire un choix entre plusieurs choses pour obtenir ce qu'il veut vraiment.

Une méthode très connue pour enseigner la CAA s'appelle la modélisation (Sennott *et al.*, 2016). Modéliser en CAA signifie que l'on parle en pointant directement les mots-clés sous la forme de pictogrammes dans l'outil de communication. L'interlocuteur, que ce soit un professionnel ou un aidant, s'adresse directement à la personne avec une DI en utilisant son système de CAA. Le patient apprend à l'utiliser en voyant comment font les autres en situation réelle. C'est pourquoi, on dit que lorsque l'on modélise, on sert de modèle à une autre personne. En 2011, la logopède Jane Korsten a dit avoir comparé le temps d'exposition au langage chez des enfants ordinaires et chez des utilisateurs de CAA (cf. la citation ci-dessous). Étant donné qu'un enfant d'un an et demi a été très peu exposé au langage s'il utilise la CAA et est suivi par un logopède deux fois par semaine, il est important de modéliser le plus souvent possible.

> “ The average 18 month old child has been exposed to 4,380 hours of oral language at a rate of 8 hours/day from birth. A child who

> has a communication system (AAC) and receives speech/language therapy two times per week for 20-30 minutes sessions will reach this same amount of language exposure in 84 years. ”

En plus de communiquer en oralisant et en montrant les pictogrammes, on peut communiquer en les échangeant. Le PECS (*Picture Exchange Communication System*) est un système de communication par échange d'images développé aux États-Unis en 1985 par un psychologue comportementaliste et une logopède (Bondy et Frost, 1994). À l'origine, il visait à aider les personnes avec un TSA à mieux communiquer dans un but fonctionnel (requête, commentaire, etc.). Cette méthode a ensuite été appliquée plus largement aux personnes avec une DI. Elle est divisée en six phases, dont certaines se retrouvent dans le CHESSEP. Selon l'organisme Pyramid fondé par Bondy et Frost, [15] les patients apprennent à échanger des images contre les objets (ou les activités) qu'ils veulent (faire) en donnant le pictogramme dans la main de leur interlocuteur, puis à discriminer les images et à faire des phrases de plus en plus complexes avec des images représentant différentes catégories morphosyntaxiques (adjectifs, prépositions, etc.).

D'autres compétences peuvent être améliorées afin de favoriser l'utilisation des images. Il peut s'agir des différentes modalités pour communiquer, comme l'entraînement du regard qui permet de pointer et sélectionner les pictogrammes. Des chercheurs ont ainsi développé GazePlay (Schwab *et al.*, 2018), [16] une plateforme de jeux utilisables par oculométrie (*eye-tracking*). En plus d'apprendre aux enfants l'utilisation de certains pictogrammes, ils offrent la possibilité d'acquérir diverses compétences : « action-réaction, sélection, littératie, mémorisation, etc. » (Schwab, 2021, p. 43).

2.6 Effets des pictogrammes sur la communication

Bien qu'à notre connaissance, il existe peu d'études menées avec un nombre élevé de personnes ayant une DI sur les effets des pictogrammes et plus généralement de la CAA, l'état de l'art de Romski et Sevcik (2005) est souvent cité. Il est repris dans les formations – destinées aux professionnels et familles de personnes en situation de handicap –, que nous avons suivies en Belgique. Il démontre 6 mythes concernant l'utilisation de la CAA, pour lesquels nous proposons la traduction française ci-dessous.

1. la CAA est le "dernier recours" dans la prise en charge de la parole et du langage ;

2. la CAA freine ou empêche le développement de la parole ;

15. Le PECS repose sur l'approche ABA, c'est-à-dire l'*Applied Behavior Analysis* (https://pecs-france.fr/picture-exchange-communication-system-pecs).

16. https://gazeplay.github.io/GazePlay

3. les enfants doivent posséder certaines compétences pour pouvoir bénéficier de la CAA ;

4. les systèmes de CAA avec synthèse vocale sont destinés uniquement aux enfants avec des capacités cognitives intactes ;

5. les enfants doivent avoir un certain âge pour bénéficier de la CAA ;

6. il existe une hiérarchie représentationnelle des symboles qui va des objets aux mots écrits (orthographe traditionnelle).

Ces croyances sont largement répandues, mais fausses. Elles représentent souvent des freins à l'introduction d'un système de CAA dès le plus jeune âge. Tout comme Schwab (2021, p. 16), nous pensons que la CAA ne va pas pour autant « brusquement permettre aux enfants en question de se développer normalement ». [17]

3 Communication dans le domaine médical

Les outils de CAA sont souvent conçus pour la communication quotidienne afin d'exprimer des besoins basiques. Ils comprennent toujours un vocabulaire de base (*Core Vocabulary*), qui est constitué des mots importants, les plus couramment utilisés quel que soit le handicap (Beukelman et Mirenda, 2017, p. 20). Dans ce vocabulaire, on trouve notamment les pronoms, les déterminants, les auxiliaires, le « oui » et le « non », quelques adjectifs et adverbes, ainsi que des verbes courants (Cataix-Nègre, 2011, p. 265). Il s'agit surtout des mots de classe fermée et donc des pictogrammes abstraits.

Ces systèmes peuvent être limités à des domaines généraux tels que les routines, l'hygiène, la nourriture, l'école, etc. ou à quelques pictogrammes dans des domaines spécifiques. Ils sont toutefois extensibles avec du vocabulaire thématique et du vocabulaire spécifique (*Fringe Vocabulary*), c'est-à-dire les mots de contenu représentant les intérêts et expériences personnelles de l'utilisateur de CAA (van Tilborg et Deckers, 2016). La personnalisation de ces outils nécessite l'implication de la personne avec une DI, de l'entourage et des professionnels (logopèdes, ergothérapeutes, etc.). Nous constatons que dans les grands ensembles de pictogrammes pour CAA, de plus en plus d'images concernent des domaines plus spécialisés, dont celui de la santé. C'est le cas de l'extension Widgit Health (cf. section 2.2.2) ou de l'ensemble Arasaac (2.2.6).

L'usage d'images dans le domaine de la santé n'est toutefois pas récent, ni limité aux personnes en situation de handicap (Houts *et al.*, 2006). On les retrouve aussi sur les boîtes des médicaments et les notices pour indiquer la posologie (Katz *et al.*, 2006). Les 81 pictogrammes USP sont par exemple des images normalisées qui aident

17. C'est également l'avis des formatrices et des logopèdes que nous avons pu rencontrer.

à transmettre les instructions, les précautions et les avertissements concernant les médicaments aux patients peu alphabétisés ou ceux dont l'anglais n'est pas la langue maternelle. [18] Ils ont été évalués auprès de différentes populations (cf. section 2.4).

La crise sanitaire liée au coronavirus a eu un certain impact sur le développement de pictogrammes médicaux comme en témoignent les images diffusées par les médias, sur les sites web d'information ou dans les lieux publics. De nombreux supports graphiques ont été conçus partout dans le monde afin de sensibiliser la population aux gestes barrières et expliquer la pandémie aux personnes avec des troubles de la communication (Othman et Al-Sinani, 2021). Le besoin de systèmes de CAA, dont les tableaux de communication, a aussi été accru dans les hôpitaux pour les patients avec des symptômes du COVID-19 incluant de sévères problèmes respiratoires entravant la communication (Altschuler *et al.*, 2021). [19]

Dans cette section, nous nous intéresserons plus spécifiquement à la communication médicale, à l'accès aux soins et donc aux questions suivantes. Comment les patients avec une DI communiquent-ils des informations médicales ? Quelles sont les barrières auxquelles sont particulièrement confrontés les patients avec une DI ? Existe-t-il des outils avec des images sur papier et/ou technologique afin de faciliter la communication, ainsi que le diagnostic des médecins dans les hôpitaux ?

3.1 Hôpitaux et accès aux soins

En milieu hospitalier, la communication est indispensable dans de nombreuses situations, et ce, avec différents intervenants, que ce soit pour réaliser une anamnèse lors d'une consultation (dialogue médecin-patient), une hospitalisation au cours de laquelle les infirmières échangent avec le patient et l'entourage, au service des urgences, à l'accueil avec les assistants administratifs, avant ou après la prise en charge (directives de sortie), etc. Les hôpitaux sont un lieu qui représente souvent une source de stress importante susceptible d'avoir un impact négatif sur la communication pour tous les patients (Alvarez, 2014), y compris ceux avec une DI.

Or, les patients avec une DI nécessiteraient encore plus de soins que les autres en raison notamment de leur santé souvent plus fragile (Ibba *et al.*, 2020; de Knegt *et al.*, 2013). Biotteau *et al.* (2013, p. 5) ont relevé que ce besoin serait deux fois et demi plus plus important par rapport à la population générale. Des problèmes de communication avec des personnes ayant une DI peuvent engendrer des erreurs ou des retards de diagnostic (Ibba *et al.*, 2020; Tuffrey-Wijne et McEnhill, 2008), ainsi que des (sous-)traitements inadaptés (Ibba *et al.*, 2020; de Knegt *et al.*, 2013), à cause d'une mauvaise évaluation de la douleur (cf. section 3.2). La qualité de la prise en charge est alors im-

18. https://www.usp.org/health-quality-safety/usp-pictograms

19. Comme le mentionnent les auteurs, beaucoup de ressources sont gratuites et facilement disponibles (https://www.patientprovidercommunication.org).

pactée, ce qui peut entraîner des conséquences sur le long terme. Différents supports et systèmes de CAA sont envisagés pour communiquer et informer les patients.

Tout comme dans les aéroports, les villes ou d'autres endroits publics avec beaucoup de circulation, les pictogrammes sont utilisés dans les hôpitaux pour guider les patients et leurs proches (Figure 1.10). Nandy (2019) avait ainsi relevé l'initiative de la *Society for Experiential Graphic Design (SEGD)* et d'*Hablamos Juntos*, qui ont proposé plusieurs ensembles de symboles conçus dans ce but. On peut les retrouver sous le nom d'*Universal Symbols in Healthcare* (Figure 1.10a). [20] L'ensemble de pictogrammes Arasaac est déjà exploité dans des hôpitaux (Paolieri et Marful, 2018). Ils guident par exemple les patients avec un TSA à l'*Hospital Municipal Josefina* en Amérique du Sud ou à l'*Hospital Universitario del Sureste* en Espagne (Figure 1.10b). [21]

(a) *Universal Symbols in Healthcare* du SEGD et d'*Hablamos Juntos*, reproduite de Nandy (2019, p. 21)

(b) Pictogrammes Arasaac en milieu médical (Madrid)

FIGURE 1.10 – Images pour guider les patients dans des hôpitaux

Concevoir des pictogrammes pour les patients avec une DI est une des façons pouvant contribuer à un accès plus équitable aux soins, à l'information et à la communication. Cette prérogative s'inscrit dans les articles 9, 21 et 25 de la Convention des Nations Unies relative aux droits des personnes handicapées de 2006, qui oblige les États signataires à prendre des mesures pour faire respecter les droits et devoirs de ces personnes. [22] Elle promeut l'inclusion et l'engagement des personnes en situation de handicap dans tous les domaines de la société, y compris celui de la santé. Il s'agit de favoriser leur autonomie, ainsi que leur autodétermination. En 2022, cette convention avait déjà été ratifiée par 185 pays, dont la Belgique en 2009, la France en 2010 ou encore la Suisse en 2014. En mars 2021, la Constitution belge a finalement intégré l'article 22 ter, qui établit que « chaque personne en situation de handicap a le droit à une pleine inclusion dans la société, y compris le droit à des aménagements raisonnables ».

En France, l'Unapei (Biotteau *et al.*, 2013) constatait déjà dans son livre blanc

20. https://segd.org/resources/hablamos-juntos
21. https://aulaabierta.arasaac.org/mapa-de-arasaac-visual (d'où provient la Figure 1.10b).
22. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-persons-disabilities

que les personnes avec un retard mental rencontraient plus de difficultés pour accéder aux soins de santé. Cette fédération française qui rassemble 550 associations proposait plusieurs actions comme le développement d'outils de communication. Elle préconisait l'élaboration de « documents médicaux en "[Facile A Lire et à Comprendre (FALC)]" largement illustrés » et l'utilisation « des recueils des pictogrammes adaptés aux situations de santé » (Biotteau *et al.*, 2013, pp. 25-26). Selon l'institution belge Unia qui lutte contre la discrimination et défend l'égalité des chances en Belgique, la crise du coronavirus a renforcé les problèmes existants en matière d'accessibilité aux soins de santé et de communication chez les personnes en situation de handicap et leur entourage. Parmi les mesures attendues des autorités par les proches, figure la demande d'une « information accessible à tous les types de handicap [dont l'information] sur l'impact du COVID-19 [les concernant] », mais aussi le souhait d'avoir une « communication adaptée aux personnes handicapées sur les mesures à suivre » (Unia, 2020, p. 42).

L'enquête d'Unia a notamment questionné 502 personnes avec différents handicaps, dont 216 ont déclaré avoir éprouvé d'importantes difficultés dans l'accès aux soins de santé durant cette période. Or, pour les proches, il est nécessaire d'impliquer directement les personnes en situation de handicap dans les questions de santé qui les concernent. Nous verrons qu'il n'est pas toujours facile pour les personnes avec une DI d'exprimer leurs problèmes, ce qui rend la communication difficile avec les médecins.

3.2 Expression et évaluation de la douleur

En plus d'être un concept relativement abstrait, la douleur est aussi subjective. Une même douleur peut être ressentie de différentes manières. Il faut savoir que les personnes avec une DI ne l'expriment pas toujours, et ce, pour diverses raisons. Elles ne savent pas toujours la communiquer ; de Knegt *et al.* (2016a) l'expliquent aussi par la peur de la réaction des autres et la peur de les déranger. Certains patients avec une DI expriment la douleur verbalement en indiquant la localisation, d'autres différemment, de façon non verbale (Tuffrey-Wijne et McEnhill, 2008). En se basant sur 27 études qualitatives menées auprès de patients avec une DI, de Knegt *et al.* (2013) ont recensé au moins 14 catégories de comportements différents pour exprimer la douleur :

1. l'absence ou la présence d'une activité motrice comme des mouvements involontaires, une certaine position du corps ;
2. une activité faciale, par exemple, avec les yeux ou la bouche ;
3. des indicateurs socio-émotionnels tels que des signes de peur, d'anxiété, de frustration, mais aussi le fait de vouloir rester seul ou de rechercher du confort ;
4. des expressions vocales non verbales, en particulier, les pleurs (dans 60-90 % des cas), les gémissements (entre 60-80 %), les hurlements ;

5. des indicateurs physiologiques, dont le rougissement ou le pâlissement du visage dans certaines situations ;
6. des comportements d'auto-mutilation qui seraient une conduite typique des personnes avec une DI ;
7. des expressions verbales, mais qui ne sont pas utilisées par tous les patients avec une DI et qui sont plus ou moins précises (avec ou sans la description de la localisation de la douleur : « j'ai mal à l'estomac » vs « je ne me sens pas bien ») ;
8. une forme d'agression comme des crises de colère ;
9. des troubles de l'alimentation ou du sommeil ;
10. de l'agitation voire parler sans s'arrêter ;
11. des changements dans la vie quotidienne qui empêchent le maintien d'une vie indépendante ;
12. une exagération des symptômes habituels du handicap ;
13. une réponse aux analgésiques ;
14. des mouvements stéréotypés tels que les battements des mains qui seraient un comportement spécifique de certaines personnes avec une DI.

Les quatre premières catégories sont les plus fréquentes, c'est-à-dire mentionnées dans plus d'une dizaine d'études, tandis que les quatre dernières catégories ne sont abordées que dans une ou deux études. En ce qui concerne la douleur, les patients avec une DI ont difficile à identifier leurs caractéristiques (effet, type, intensité et localisation). Afin d'améliorer cette identification, de Knegt *et al.* (2016b) ont testé des pictogrammes Sclera auprès de 39 patients adultes avec un syndrome de Down pour évaluer la douleur (effet et type). Les participants ont reconnu la bonne représentation de la douleur de deux façons : a) à l'aide des traditionnelles échelles [23] – dont la *Facial Affective Scale* ou FAS conçue pour les enfants –; b) avec des pictogrammes (cf. Figure 1.7b). Les auteurs ont montré qu'utiliser des pictogrammes peut optimiser la communication de la douleur chez certains patients avec une DI. Il est intéressant de noter que 74 % des participants avaient déjà des pictogrammes dans leur environnement, principalement pour communiquer des activités ou des objets ; seulement 5 personnes les utilisaient pour exprimer des émotions dans leur vie quotidienne.

En se basant sur ces résultats, de Knegt *et al.* (2016a) ont développé STOP-ID (*Self-Reporting Tool On Pain in People with Intellectual Disabilities*), une application

23. Pour faciliter l'auto-évaluation de la douleur, en particulier chez les enfants, il existe plusieurs types d'échelles avec des images : échelle visuelle analogique, numérique, avec des visages, etc.

avec des pictogrammes sur ordinateur et tablette pour l'auto-évaluation de la douleur, qu'ils ont fait tester par 40 patients adultes avec un syndrome de Down. Ils ont montré que plus de participants avaient reconnu les représentations graphiques de la localisation et de l'effet de la douleur que les représentations du type et de l'intensité de la douleur. Les pictogrammes Sclera ont encore été évalués, en plus d'un dessin représentant le corps humain de face et de dos. Ce type de représentation est souvent utilisé pour permettre au patient de situer la douleur, par exemple, avec le *Medical Visual Language Translator* (cf. infra) ou l'application MediPicto AP-HP (section 3.3).

En cas de communication verbale entre un médecin et un patient avec une DI : le médecin doit s'adapter au patient (Tuffrey-Wijne et McEnhill, 2008). Les auteurs recommandent d'éviter les questions fermées car les patients avec une DI auraient tendance à répondre « oui » quelle que soit la question, même si les patients répondraient souvent le dernier choix après des questions ouvertes. Le médecin doit donc s'assurer que le patient a bien compris en lui demandant de reformuler avec ses propres mots. Iacono et Johnson (2004) rappellent que les patients avec une DI rencontrent des problèmes de communication, tant au niveau de la compréhension (de ce qui est dit) que de l'expression (qui est difficile à comprendre et dont le vocabulaire, ainsi que la formulation des phrases sont limités). Les auteures recommandent que le patient utilise son propre système de CAA et soit accompagné d'un membre de la famille, d'une personne de soutien rémunérée ou d'un autre professionnel durant les consultations.

Elles conseillent aussi au médecin d'utiliser des images et de prévoir un double rendez-vous pour avoir plus de temps pour la communication. Mais lorsque le médecin n'a pas le temps de se former aux outils de CAA et que l'auto-évaluation par le patient n'est pas possible, on peut également recourir à une hétéro-évaluation par des tierces personnes. Il existe des outils spécifiques comme la Grille d'Évaluation de la Douleur - Déficience Intellectuelle ou GED-DI (Zabalia *et al.*, 2011) – l'adaptation francophone de la NCCPC-PV – et l'échelle de la Douleur Enfant San Salvadour ou DESS destinée à la prise en charge des enfants avec une DI sévère ou en situation de polyhandicap qui ont des difficultés à s'exprimer (Collignon et Giusiano, 2001). La DESS est utilisée dans les hôpitaux, par exemple, les Hôpitaux Universitaires de Genève (HUG). Un questionnaire doit être rempli avec la famille ou la personne qui connaît le mieux le patient. Il constitue le dossier de base décrivant le comportement habituel du patient. La grille d'observations de 10 items est complétée par l'équipe médicale à l'hôpital.

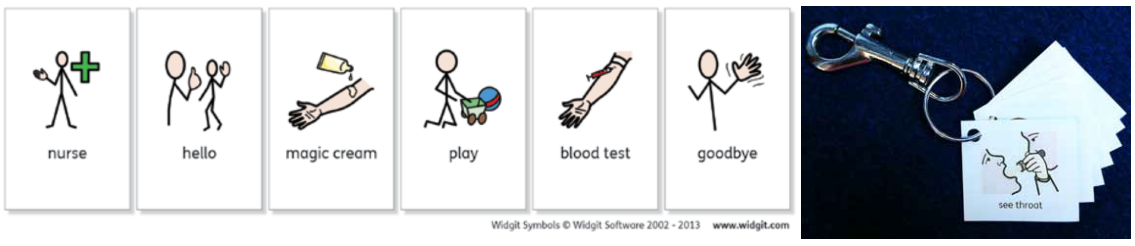
En ce qui concerne la CAA, elle peut aussi être envisagée dans les hôpitaux pour les enfants (Mesko *et al.*, 2011) et les patients sans DI qui ne peuvent pas communiquer avec le langage oral à cause de leur état (traumatisme crânien, accident vasculaire cérébral, etc.) ou de certaines opérations comme l'intubation ou la trachéotomie (Beukelman et Mirenda, 2017; Hurtig et Downey, 2008). Différents outils sont mis à leur disposition, dont des tableaux de communication pour exprimer des besoins de base,

des sentiments, la douleur, etc. Contrairement aux personnes avec une DI, ces patients ont rarement employé des outils de CAA. Ils doivent apprendre rapidement comment les utiliser à l'hôpital pour pouvoir communiquer avec les soignants et leurs proches.

(a) Kit de communication (Paris) (b) *Medical Visual Language Translator* (Genève)

FIGURE 1.11 – Supports papiers avec images utilisés dans des hôpitaux

Certains hôpitaux créent eux-mêmes leurs propres images pour faciliter la communication avec leurs patients (Figure 1.11). C'est le cas de l'Assistance Publique - Hôpitaux de Paris (AP-HP), qui avait développé un Kit de communication au format papier (Figure 1.11a),[24] avant de concevoir l'application MediPicto AP-HP.[25] L'unité des urgences ambulatoires des HUG de Genève utilise notamment le *Medical Visual Language Translator* au format papier (Figure 1.11b).[26] Créé pour la communication instantanée entre le personnel soignant et les patients, ce support propose un vaste choix d'images dans différentes thématiques : les procédures de diagnostic et de traitement, les requêtes, les allergies, etc. Le système BabelDr a ensuite été développé pour la traduction de la parole aux HUG et une interface de réponses avec des pictogrammes a été implémentée comme nous le mentionnerons dans le prochain chapitre.



(a) Chronologie d'événements : prise de sang (b) Porte-clés

FIGURE 1.12 – Pictogrammes Widgit Health,© reproduite de Vaz (2013, p. 157)

Le recours aux images en milieu médical est aussi envisagé pour les personnes avec une DI. Vaz (2013) a évalué les pictogrammes Widgit Health présentés différem-

24. http://unikom.blogs.apf.asso.fr/archive/2010/02/index.html (d'où provient la Figure 1.11a).
25. https://www.aphp.fr/medipicto
26. https://www.kwikpoint.com/products-page/medical-translators-prodcat-travel/medical-visual-language-translator (d'où provient la Figure 1.11b).

ment (Figure 1.12) : via une chronologie d'événements (Figure 1.12a) ou sur un porte-clés (Figure 1.12b) représentant des examens médicaux, des procédures cliniques et des traitements. Son étude a montré qu'utiliser des pictogrammes dans le domaine de la santé et les envoyer à la maison avant les rendez-vous médicaux améliore la compréhension et la coopération des enfants autistes ou avec des troubles de l'apprentissage.

3.3 Applications médicales avec des images

Il existe aussi plusieurs applications sur ordinateur, tablette ou smartphone pour améliorer la communication médicale avec des images. Elles sont souvent simples et développées afin de répondre aux besoins d'un grand nombre de personnes (*design for all*), n'ayant pas forcement un handicap, comme les patients allophones, les migrants, les enfants, etc. Bien que les performances des systèmes de traduction automatique comme Google Traduction continuent de s'améliorer, leur utilisation dans le domaine médical reste problématique (Bouillon *et al.*, 2017; Patil et Davies, 2014). En effet, des erreurs dans ce type de technologies ne sont pas acceptables (Wołk *et al.*, 2017). C'est pourquoi, des recherches sont menées pour concevoir des systèmes de communication médicale qui tentent de réduire un maximum les risques de mauvais diagnostic.

Pour la communication médicale entre médecins et patients, Wołk *et al.* (2017) ont présenté un prototype de montre connectée avec des pictogrammes, aussi utilisable sur smartphone et tablette. Elle se veut simple d'utilisation en comparaison avec les systèmes exploitant le TAL, afin d'éviter le risque d'erreur des systèmes de traduction automatique. Cet outil est destiné aux patients étrangers et aux personnes avec des troubles de la parole, de l'audition ou une DI. Au Canada, Alvarez (2011) a commencé par évaluer différentes représentations biomédicales pour concevoir une interface graphique pluriculturelle de l'échelle de triage. Elle a ensuite lancé l'application Mon traducteur de symptômes (Alvarez, 2014).

Comme nous l'avons déjà mentionné, l'Assistance publique - Hôpitaux de Paris (AP-HP) a conçu un support papier (cf. Figure 1.11a), dont l'utilisation est proche des systèmes de CAA. Elle a ensuite développé l'application MediPicto pour améliorer la communication médicale entre les soignants et les patients. Elle propose des images liées à des mots (« piqûre », « boire »), des assertions ou instructions (« j'ai la diarrhée », « ne respirez plus ») et des questions souvent fermées (« avez-vous mal ? », « est-ce que ça brûle ? »). L'utilisateur peut choisir son profil – soignant ou patient –, ainsi que la langue associée à ces deux rôles à chaque lancement du système. Pour accéder plus rapidement au pictogramme souhaité, il peut faire une recherche par mot-clé. Divers outils sont intégrés, dont une échelle pour évaluer la douleur et les émotions, le dessin d'un corps humain pour situer différentes douleurs, des calendriers, des pictogrammes représentant des formes pharmaceutiques pour indiquer la posologie, etc.

4 Conclusion : des pictogrammes universels ?

Dans ce chapitre, nous avons donc introduit le contexte général de notre projet. Il existe un vaste domaine de recherches qui concerne la CAA pour les personnes avec une DI. Les images peuvent être utilisées pour communiquer de différentes façons dans la vie quotidienne, mais aussi dans des domaines plus spécialisés comme celui de la santé pour évaluer la douleur et/ou pour des consultations médicales. Des ensembles de pictogrammes ont été conçus pour les personnes en situation de handicap. Ils ont chacun leurs caractéristiques. Contrairement à Sevens (2018), nous avons choisi d'introduire les pictogrammes du point de vue de la CAA et moins de la sémantique avec les théories du signe de Peirce et de Saussure. Il faut toutefois mentionner qu'il existe des travaux en sémiologie ainsi que dans d'autres domaines comme en psychologie cognitive et en ergonomie avec le concept d'affordance (Norman, 2013), qui peut être utilisé pour les images. Nous avons vu que les pictogrammes médicaux ont également fait l'objet d'évaluations, mais celles-ci concernent surtout différents publics sans DI. Des outils avec des images sont aussi développés pour faciliter la communication des personnes qui n'ont pas forcément appris à se servir d'images. Les pictogrammes ne sont pas universels. Nous avons vu que leur utilisation en CAA est susceptible de varier en fonction de plusieurs facteurs, dont leur apprentissage, la connaissance de leur référent, la catégorie morphosyntaxique du concept représenté, le contexte, les préférences des utilisateurs, la présence d'un trouble perceptif ou encore des différences culturelles – les pictogrammes Sclera et Beta sont plus utilisés en Belgique que dans d'autres pays, [27] où d'autres ensembles sont plus répandus. [28] Les pictogrammes peuvent nécessiter aux personnes avec une DI plusieurs années d'apprentissage auprès de professionnels de la santé (logopèdes, ergothérapeutes, etc.). Dans le prochain chapitre, nous présentons les systèmes qui permettent de traduire automatiquement des phrases vers une séquence de pictogrammes qui proviennent d'un ou plusieurs ensembles utilisés en CAA que nous avons mentionnés ici.

27. C'était le constat de Vaschalde (2018) pour Sclera et Beta en France. C'est pourquoi, elle avait utilisé d'autres ensembles pour son système français speech-to-picto.

28. Tawasol est un ensemble de pictogrammes pour la CAA arabe (Othman et Al-Sinani, 2021), qui serait plus utilisé dans ces pays en raison de ses caractéristiques culturelles visuelles, par exemple, pour la représentation des personnages (https://tawasol.mada.org.qa).

Chapitre 2

Traitement automatique : langues et pictogrammes

Introduction

Dans ce chapitre, nous introduisons les deux principales tâches de Traitement Automatique des Langues (TAL) qui font l'objet de cette thèse : la traduction automatique (section 1) et la simplification de textes (section 2).

La première section commence par poser le cadre plus large de la traduction automatique en présentant la typologie des systèmes de traduction automatique. Nous y abordons les systèmes basés sur des règles, les outils statistiques, hybrides, neuronaux, avec des grands modèles de langue génératifs, ainsi que les traducteurs à phrases fixes. Ensuite, nous nous focalisons sur les systèmes de traduction automatique de phrases en langue naturelle vers les pictogrammes, lesquels peuvent inclure ou non des techniques de TAL. Nous verrons pourquoi, dans cette thèse, nous n'utilisons pas l'approche de traduction automatique classique aujourd'hui, à savoir la traduction neuronale. De plus, nous décrivons les méthodologies d'évaluation de ces systèmes de traduction vers les pictogrammes. Nous présentons ensuite le projet PROPICTO, qui vise plus spécifiquement à développer et évaluer des systèmes de traduction automatique du français vers les pictogrammes, dont un pour les patients en milieu hospitalier. C'est pourquoi, nous évaluerons également ce système avec nos utilisateurs cibles dans le dernier chapitre (pour répondre à l'objectif 2 et aux sous-objectifs 1, 2 et 3). Étant donné que nous voulons évaluer l'impact de la simplification automatique de phrases entrées dans un système de traduction automatique vers les pictogrammes sur la sortie (objectif 1, sous-objectif 5), la deuxième section présente le domaine de la simplification de textes. Nous y donnons des exemples d'utilisation de la simplification manuelle réalisée par des humains, avant de présenter la simplification automatique effectuée par des machines pour différents publics et/ou pour des systèmes de TAL.

1 Traduction automatique

Comme l'avait défini Vandeghinste (2008, p. 9), la traduction automatique est l'automatisation du processus de traduction, c'est-à-dire la conversion d'une phrase d'une langue source vers une langue cible, en conservant un sens identique ou similaire. La traduction automatique est une technologie qui est aujourd'hui utilisée par de plus en plus de monde. Elle peut être utile dans beaucoup de situations différentes. Dans cette thèse, nous nous intéressons plus particulièrement aux systèmes de traduction automatique d'une langue source naturelle vers une langue cible pictographique. Ce type de systèmes peut s'intégrer dans une application médicale basée sur la Communication Alternative et Améliorée (CAA), destinée à améliorer la communication des personnes en situation de handicap, dont celles avec une Déficience Intellectuelle (DI). Nous verrons qu'il en existe peu répondant à tous ces critères.

Nous donnons d'abord un aperçu des différentes approches en traduction automatique (section 1.1), ainsi que des systèmes de traduction automatique de phrases vers les pictogrammes (section 1.2). Nous décrivons aussi les méthodes qui ont été utilisées pour évaluer ces systèmes (section 1.3). Nous présentons ensuite le système de traduction médicale vers les pictogrammes PictoDr (section 1.4). Enfin, nous résumons tous les systèmes de traduction automatique vers les pictogrammes avec TAL que nous avons cités avec leurs caractéristiques et méthodes d'évaluation (section 1.5).

1.1 Typologie des systèmes de traduction

La traduction automatique est une tâche de TAL qui a fait l'objet de nombreuses recherches que nous ne détaillerons pas. [1] Elle peut se présenter sous différentes architectures (Figure 2.1) : à base de connaissances linguistiques (section 1.1.1), de connaissances statistiques (section 1.1.2), de connaissances linguistiques et statistiques (section 1.1.3), avec des réseaux de neurones (section 1.1.4), avec des grands modèles de langue génératifs (section 1.1.5), voire basée sur des phrases fixes (section 1.1.6).

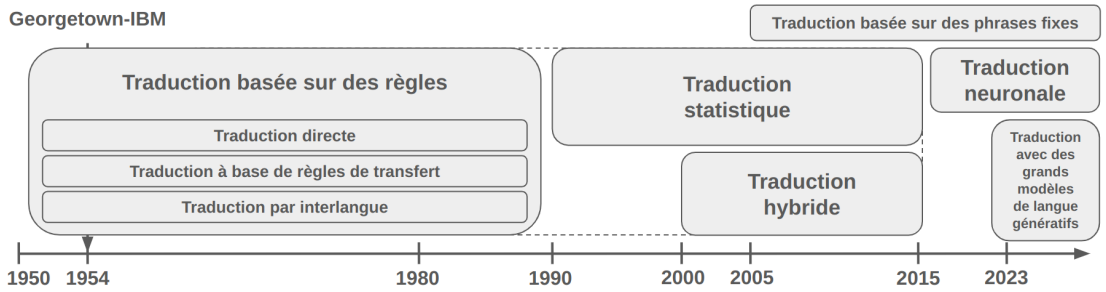


FIGURE 2.1 – Historique de la traduction automatique

1. Pour un historique détaillé et des précisions sur les méthodes, le lecteur peut se référer à Koehn (2009), Hutchins (2023), Jurafsky et Martin (2024).

1.1.1 Traduction basée sur des règles

Les premiers systèmes sont apparus au début des années 1950. En 1954, la première démonstration publique de traduction automatique a eu lieu (du russe vers l'anglais), plus connue sous le nom d'expérience Georgetown-IBM. Jusqu'à la fin des années 1980, la traduction automatique reposait uniquement sur des connaissances linguistiques qui sont facilement interprétables pour les humains. Produire des règles à la main est une tâche qui prend beaucoup de temps aux experts linguistes. Ces règles visent à lever des ambiguïtés à différents niveaux de la langue (morphologique et syntaxique, ainsi qu'à extraire la représentation symbolique). Un dictionnaire et des analyseurs morphosyntaxiques, voire sémantiques sont nécessaires. Cette approche symbolique est de moins en moins utilisée aujourd'hui. Elle l'est encore, notamment pour les langues peu dotées, qui disposent de peu de ressources informatisées.

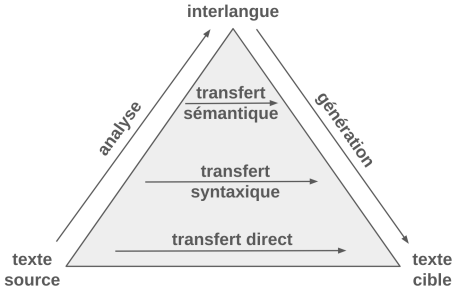


FIGURE 2.2 – Triangle de Vauquois

Il existe trois sous-approches pour passer d'un texte en langue source à un texte en langue cible : (1) directe, (2) à base de règles de transfert, (3) par interlangue. Elles sont traditionnellement représentées sous la forme d'un triangle (Figure 2.2), le triangle de Vauquois (1968), qui illustre les trois étapes de traduction. Le processus de traduction débute par une phase d'analyse du texte en langue source. Cette étape peut aboutir à une représentation abstraite du sens du texte, correspondant au sommet du triangle. Les systèmes produisent une telle représentation en utilisant une interlangue, un énoncé structuré dans un langage artificiel. Enfin, le système génère une version du texte qu'il considère comme la meilleure traduction dans la langue cible.

1.1.2 Traduction statistique

La traduction automatique statistique s'est largement développée dans les années 1990. Ces systèmes sont basés sur des connaissances statistiques liées aux théories des probabilités. Ils fonctionnent essentiellement avec deux types de modèles statistiques : un modèle de langue appris automatiquement à partir d'un corpus monolingue dans la langue cible ; un modèle de traduction appris à partir d'un corpus parallèle bilingue avec des textes en langue source et en langue cible (des traductions

réalisées par des humains). Ces systèmes nécessitent donc des corpus de plusieurs milliers, voire des millions de phrases, afin d'obtenir les tables d'alignement contenant les probabilités de chaque expression en langue source et en langue cible. L'outil Moses implémente cette approche statistique (Koehn *et al.*, 2007). Google Translate a été le premier système de traduction automatique statistique sur le web, mais il est passé au neuronal en 2016. Des études ont montré que cette architecture n'est pas optimale pour la traduction de phrases dans le domaine médical (Patil et Davies, 2014; Bouillon *et al.*, 2017), les erreurs étant trop fréquentes. Selon Patil et Davies (2014), la précision des phrases médicales générées avec Google Translate atteignait seulement 57,7 %.

1.1.3 Traduction hybride

La traduction automatique hybride a commencé à apparaître aux environs des années 2000. Elle combine des connaissances linguistiques et statistiques. Par exemple, Vandeghinste (2008) a proposé le système hybride LoRe-MT pour la traduction automatique du néerlandais vers l'anglais. [2] Il y a vingt ans, les corpus parallèles étaient plus rares. La traduction hybride avait pour objectif de combiner les avantages des deux types d'approches précédentes, afin d'optimiser les performances du système.

1.1.4 Traduction neuronale

Il s'agit du type de traduction automatique qui est le plus connu aujourd'hui. Il utilise des modèles de langue neuronaux basés sur le fonctionnement des neurones biologiques. Les systèmes de traduction automatique neuronale se basent sur une représentation du mot (*word embedding*), qui permet d'encoder le sens en contexte à tous les niveaux linguistiques, contrairement à la traduction automatique statistique. Ils ont après été basés sur une architecture encodeur-décodeur (séquence-à-séquence), de type *transformer* avec un mécanisme d'auto-attention (Vaswani *et al.*, 2017). Ils encodent la phrase source dans le réseau d'encodeurs et décodent la phrase cible conditionnée aux sorties précédentes dans le réseau de décodeurs. L'attention permet de se concentrer sur des parties de la séquence d'entrée pendant que le modèle prédit la séquence de sortie. La traduction neuronale nécessite toutefois un grand volume de données pour l'entraînement des modèles. Parmi les systèmes qui utilisent cette approche d'apprentissage profond, nous pouvons citer OpenNMT (Klein *et al.*, 2017), Marian-NMT (Junczys-Dowmunt *et al.*, 2018) ou encore les systèmes commerciaux DeepL et Google Translate. Lorsque nous avons commencé cette thèse, il n'existait pas de données pour la langue pictographique médicale en français pour affiner des modèles pré-entraînés. Nous n'avons donc pas utilisé cette approche pour le système Text-to-Picto.

2. Dans sa thèse, Vandeghinste (2008) a cité plusieurs systèmes hybrides, dont Matador.

1.1.5 Traduction avec des grands modèles de langue génératifs

Aujourd'hui, il est difficile de ne pas aborder les grands modèles de langue génératifs de type décodeur seulement. Leur taille dépasse les dizaines, voire les centaines de milliards de paramètres (Zhao *et al.*, 2023). Les grands modèles de langue génératifs sont des boîtes noires, leur fonctionnement interne est opaque pour les utilisateurs. Ce sont des modèles multilingues et majoritairement pré-entraînés sur des textes provenant d'Internet (Wikipedia, Common Crawl, WebText, BookCorpus, etc.). Ils sont utilisés par des (non-)experts pour résoudre diverses tâches réelles, qui ne sont pas forcément liées au TAL. Tout le monde peut y accéder et les interroger directement via une interface de chat ou une *Application Programming Interface (API)*. Ils peuvent également être utilisés pour la traduction automatique. De nombreux grands modèles de langue génératifs existent, tels que *Generative Pre-trained Transformer (GPT)*, ChatGPT, Vicuna, LLaMA, Gemini, etc. Des premières études ont montré que par rapport à des systèmes de traduction commerciaux (par exemple, Google Translate, DeepL et Tencent TranSmart), ChatGPT est compétitif sur des langues bien dotées (Jiao *et al.*, 2023; Hendy *et al.*, 2023). Leurs résultats sont toutefois moins bons pour les langues avec peu de ressources ou plus distantes entre elles. Selon Jiao *et al.* (2023), les systèmes commerciaux sont meilleurs que ChatGPT pour le domaine biomédical, mais celui-ci est potentiellement un bon traducteur pour la langue parlée. Peng *et al.* (2023) ont aussi évalué des grands modèles de langue génératifs et montré qu'il était possible d'améliorer leurs performances pour la traduction automatique en utilisant certains paramètres ou en spécifiant les requêtes en langue naturelle adressées à ces modèles. Contrairement à l'approche précédente, celle-ci nécessite peu, voire pas du tout de données (exemples). Au moment où nous réalisons cette thèse, les grands modèles de langue génératifs ne traduisent pas de phrases vers des séquences de pictogrammes de CAA.

1.1.6 Traduction basée sur des phrases fixes

Comme l'ont mentionné Bouillon *et al.* (2023, p. 129), les traducteurs à phrases fixes (Seligman et Dillinger, 2013), aussi appelés phraselator, sont souvent utilisés dans le domaine médical pour des questions de sécurité et de précision. Nous pouvons par exemple citer l'application mobile CALD Assist, Universal Doctor Speaker, MediBabble ou encore BabelDr, qui sera présenté dans la section 1.4.1. Ils sont basés sur une liste limitée de phrases pré-traduites, qui peuvent ensuite être présentées au patient au format écrit ou oral, en utilisant soit la synthèse vocale, soit des enregistrements audio humains (Bouillon *et al.*, 2023, p. 130). Encore récemment, des études ont mis en évidence que tant les patients que les médecins manifestent une préférence pour l'utilisation d'un traducteur à phrases fixes comme BabelDr par rapport à un système de traduction automatique générique tel que Google Translate (Bouillon *et al.*, 2017).

1.2 Systèmes de traduction vers les pictogrammes

Il existe beaucoup de systèmes de CAA commerciaux, qui permettent aux utilisateurs en situation de handicap de sélectionner une suite de pictogrammes souvent accompagnés d'un nom et disposés sous la forme de grilles. Ils génèrent automatiquement un message dans une fenêtre de phrases (traduction picto-to-text), qui peut ensuite être synthétisé oralement (traduction picto-to-speech). Nous pouvons citer le logiciel de communication Mind Express et les applications sur tablette ou smartphone Proloquo2Go, Snap Core First, LetMe Talk ou Avaz qui ont été traduites en français. La sélection manuelle des images se fait alors au moyen de menus et sous-menus représentant les catégories (personnes, lieux, nourriture, etc.) de chaque pictogramme.

Dans cette thèse, nous ne détaillerons pas ces outils et nous présenterons les systèmes permettant de traduire automatiquement des phrases en langue naturelle vers les pictogrammes (traduction text-to-picto). Ils peuvent être utilisés pour améliorer la communication et l'accès à l'information en la rendant plus simple (compréhensible) pour les personnes en situation de handicap, dont celles avec une DI.

1.2.1 Systèmes sans TAL

Les outils text-to-picto commercialisés par des sociétés de CAA sont moins fréquents que ceux qui permettent de sélectionner directement des pictogrammes, même s'il existe des logiciels payants ou gratuits pour concevoir du matériel de CAA imprimable au format papier pour différents usages quotidiens (par exemple, Boardmaker® ou Picto-Selector). Ils se limitent souvent à une simple recherche textuelle avec des mots-clés dans une ou plusieurs banque(s) de pictogrammes différentes. Ces logiciels sont destinés aux aidants, aux enseignants, ainsi qu'aux professionnels (logopèdes, ergothérapeutes, etc.). Plusieurs traducteurs de textes vers les pictogrammes Arasaac sont accessibles en ligne, mais ils sont limités et proposent souvent des traductions littérales. C'est le cas des applications AraWord ou DictaPicto, qui permettent d'écrire ou de dicter simultanément des phrases espagnoles en pictogrammes Arasaac. [3]

Dans le service web espagnol Pictotraductor, [4] lorsque l'utilisateur saisit un mot qui est lié à plusieurs images Arasaac, il doit sélectionner lui-même le pictogramme qu'il veut parmi ceux proposés par l'outil. Par exemple, s'il écrit : « *cae en coma* » (il tombe dans le coma), l'utilisateur peut choisir parmi plusieurs pictogrammes pour le verbe (une personne qui tombe, faire tomber un verre ou une dent qui tombe), ainsi que pour le nom (une virgule ou deux représentations différentes du verbe manger « *comer* », car le pictogramme médical n'existe pas). Le pictogramme « *en* » (dans) est aussi affiché. Bien que cet outil ne détecte pas les expressions multi-mots, il semble

3. https://aulaabierta.arasaac.org/araword-1-introduccion ; https://www.fundacionorange.es/aplicaciones/dictapicto-tea

4. https://www.pictotraductor.com

toutefois lemmatiser les entrées – étant donné que le verbe manger est proposé –, ce qui n'est pas non plus toujours le cas dans toutes les applications.

Le site web Pictofacile permet de convertir un texte dans 27 langues naturelles (dont le français, l'arabe, le polonais ou le chinois) en pictogrammes Arasaac. [5] L'utilisateur doit ici aussi sélectionner un pictogramme pour chaque mot parmi les propositions du système (c'est à l'utilisateur de gérer le problème d'ambiguïté). Il peut supprimer un pictogramme, le renommer ou ajouter une image personnelle. Bien que l'interface soit pratique et permette de générer des documents en pictogrammes, l'application est simple et sans TAL. Les propositions sont basées sur les mots entrés, non tokenisés et non lemmatisés, ce qui génère des mauvaises propositions de pictogrammes. Il ne traduit pas les mots qui ne se trouvent pas dans la base de données Arasaac.

1.2.2 Systèmes avec TAL

Cet état de l'art ne se veut pas exhaustif. Nous avons fait le choix de présenter uniquement les technologies text-to-picto basées sur des techniques de TAL à partir de la période du Text-to-Picto de Sevens (2018), [6] à l'exception des études de Mihalcea et Leong (2008), Bautista *et al.* (2017).

Mihalcea et Leong (2008) ont utilisé la ressource sémantique WordNet (Miller, 1995), mais sans exploiter les relations lexicales. De plus, leur système visait seulement à traduire les noms et les verbes. Ce n'est pas le cas du système néerlandais Text-to-Picto pour les emails des personnes avec une DI (Sevens, 2018; Vandeghinste *et al.*, 2017), que nous adapterons. Comme nous le verrons dans le prochain chapitre, leur système de base exploite plusieurs types de relations issus de WordNet. Il comprend un dictionnaire pour traduire tous les mots, un algorithme de recherche A* (Hart *et al.*, 1968), ainsi qu'une étape de prétraitements du texte (analyse linguistique superficielle). Sevens (2018) a aussi ajouté trois modules optionnels : la correction orthographique automatique pour des utilisateurs d'Internet avec une DI (Sevens *et al.*, 2016b), la simplification automatique avec des règles pour 12 phénomènes syntaxiques (Sevens *et al.*, 2017), [7] ainsi que la désambiguïsation lexicale (Sevens *et al.*, 2016a).

Text-to-Picto a été conçu pour traduire automatiquement le néerlandais en pictogrammes Sclera et Beta, puis étendu à l'anglais et l'espagnol (Sevens *et al.*, 2015a). Kultsova *et al.* (2017) ont intégré cet outil dans une application mobile destinée aux personnes avec une DI pour les voyages et la communication en russe (*Travel and*

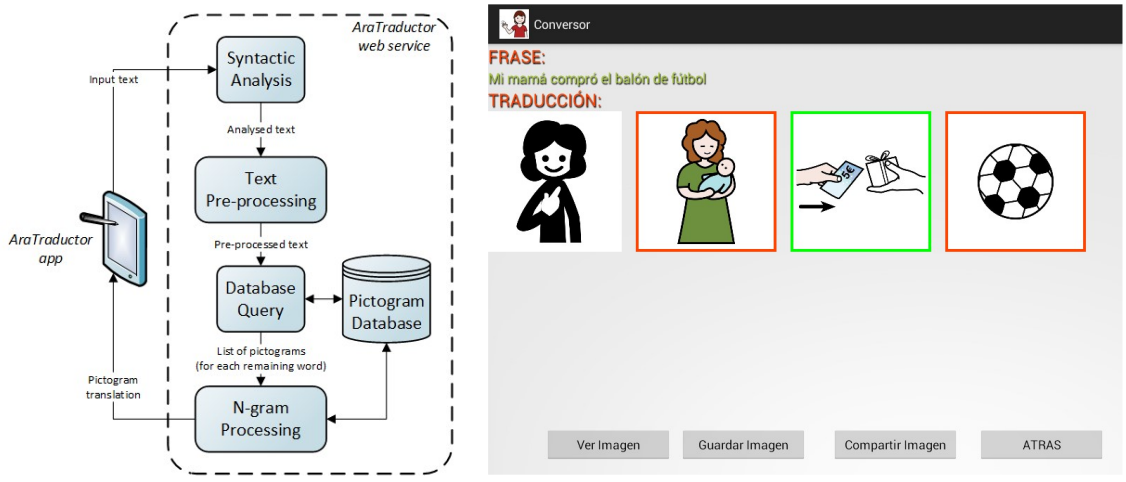
5. https://www.pictofacile.com

6. Le lecteur intéressé par des systèmes plus anciens pourra se référer à la thèse de Sevens (2018).

7. Les 12 phénomènes sont : phrase coordonnée, proposition subordonnée, ordre des propositions, groupe participial, (om) te 'pour' + proposition infinitive, proposition relative, proposition appositive incorporée, groupe adverbial ou groupe prépositionnel en position de thème, sujets ou objets cachés dans la proposition indépendante nouvellement créée, ordre non sujet + verbe + objet, voix passive, compression de phrases (Sevens, 2018, p. 116).

Communication Assistant). Nandy (2019) a aussi proposé d'adapter ce système pour les langues indiennes. La version néerlandaise de Text-to-Picto a continué d'être améliorée, puisque les auteurs l'ont adapté, afin de faciliter la communication et l'intégration des migrants, en utilisant l'ensemble de pictogrammes belges Rand conçu pour ce public (Bulté *et al.*, 2021).[8] Un module pour traiter certaines "entités nommées" a été ajouté. Pour ces cas, il permet d'afficher des images – sans restrictions – de Wikipédia (par exemple, pour le jeu de Go chinois ou les villes belges d'*Oostende* et d'*Antwerpen*).

Des outils ont été développés pour traduire automatiquement du texte espagnol en pictogrammes Arasaac, puisqu'il s'agit de la plus grande base d'images libres de droit créée en Espagne. AraTraductor (Figure 2.3) est une application web qui est aussi disponible sur Android (Bautista *et al.*, 2017). Elle utilise des techniques de TAL (analyse syntaxique, prétraitement du texte, gestion des expressions multi-mots via des modèles n-grammes) pour améliorer la traduction en pictogrammes (Figure 2.3a). Dans AraTraductor, chaque pictogramme généré est entouré d'un cadre dont la couleur représente la catégorie du mot (les noms propres en jaune, les noms communs en orange, les actions en vert, les adjectifs en bleu, etc.). Ainsi, dans l'exemple (Figure 2.3b) : « *mi mamá compró el balón de fútbol* » (ma maman a acheté le ballon de foot), les noms « *mamá* » et « *balón de fútbol* » – une expression multi-mot traduite par un seul pictogramme – sont en orange, tandis que le verbe « *compró* » est en vert. L'article « *el* » a également été omis durant l'étape de prétraitement du texte.



(a) Architecture

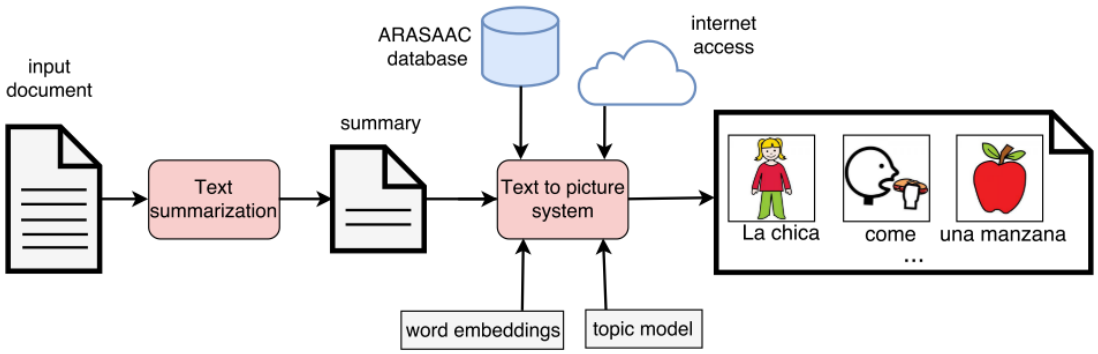
(b) Exemple de la phrase « *mi mamá compró el balón de fútbol* »

FIGURE 2.3 – Système AraTraductor, reproduite de Bautista *et al.* (2017, pp. 4-5)

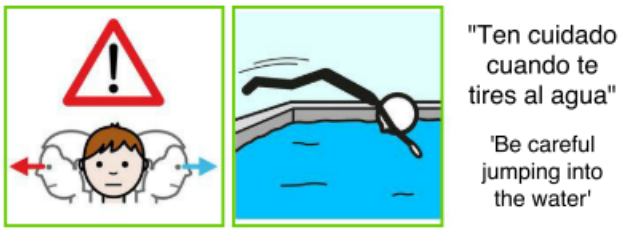
L'objectif du système text-to-picto espagnol de Cabello *et al.* (2018) est d'aider les personnes avec des difficultés langagières à mieux comprendre les textes complexes (Figure 2.4). Les auteurs ont testé des techniques de résumé automatique de texte, des

8. https://www.pictogrammendatabank.be

algorithmes d'analyse sémantique latente. Ce système génère de courtes séquences de pictogrammes : les expressions multi-mots sont traduites par un seul pictogramme et les mots de fonction (articles, prépositions, etc.) ne sont pas pris en compte, c'est le cas de : « *la chica come una manzana* » (la fille mange une pomme) et « *ten cuidado cuando te tires al agua* » (fais attention quand tu sautes dans l'eau), qui sont respectivement traduits par trois et deux pictogrammes (Figures 2.4a et 2.4b).



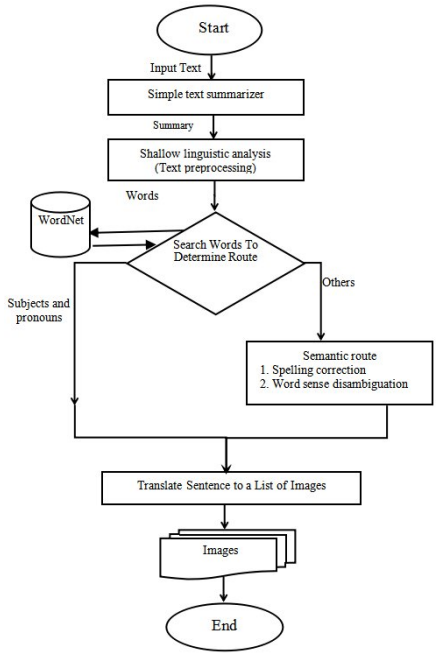
(a) Architecture



(b) Exemple de la phrase « *ten cuidado cuando te tires al agua* »

FIGURE 2.4 – Système text-to-picto, reproduite de Cabello *et al.* (2018, pp. 16-17)

Imam *et al.* (2019) ont comparé plusieurs méthodes de désambiguïsation lexicale – par exemple, l'algorithme de Lesk et un Séparateur à Vaste Marge (SVM) – dans leur système de traduction de l'anglais vers les pictogrammes pour les personnes ayant des troubles du langage. L'outil proposé fonctionne avec WordNet et ImageNet (Deng *et al.*, 2009). Il possède une architecture très proche de celle du système Text-to-Picto. En effet, le système d'Imam *et al.* (2019) comprend aussi une phase d'analyse linguistique superficielle – dont une tokenisation, un étiquetage morphosyntaxique et une lemmatisation –, une route directe pour traiter certains types de mots comme les pronoms, ainsi qu'une route sémantique qui gère les ensembles de synonymes de WordNet (Figure 2.5a). Imam *et al.* (2019) ont donné des exemples de phrases traduites avec leur système (Figure 2.5b) : « *I read a book* » (je lis un livre) et « *a cat eats a fish* » (un chat mange un poisson). La première a été bien traduite, tandis que le mot « *cat* » (chat) de la deuxième phrase a été mal traduit, c'est-à-dire par « *tree leaf* » (feuille d'arbre), car il s'agit d'un des sens de « *cat* » dans WordNet (Imam *et al.*, 2019, p. 30). À noter qu'elles ont utilisé une horloge 🕒 pour représenter le temps du verbe.



(a) Architecture

Example	Input	Output by kids	Precision	Recall	FScore
1	I read a book	I read a book	1	1	1
2	A cat eats a fish	A tree leaf eats a fish	.66	.8	0.727
3	I love my home	I love baseball	.66	.5	0.571
4	I love dogs	I love car wheels	.5	.66	0.571

(b) Exemple des phrases « *I read a book* » et « *a cat eats a fish* »

FIGURE 2.5 – Système text-to-picto anglais, reproduite d'Imam *et al.* (2019, pp. 29-30)

En ce qui concerne la traduction à partir du français, Vaschalde (2018), Vaschalde *et al.* (2018) ont implémenté à Grenoble un outil speech-to-picto avec le WordNet anglais et un module de reconnaissance vocale (Figure 2.6). Ce système permet de traduire automatiquement la parole vers les pictogrammes Arasaac et Makaton.[9] Basé sur le travail de Vandeghinste *et al.* (2017), Vaschalde (2018) et Vaschalde *et al.* (2018) ont évalué leur prototype en testant notamment la désambiguïsation lexicale et la simplification automatique des structures passives (« a été mangée par » → « mange »).

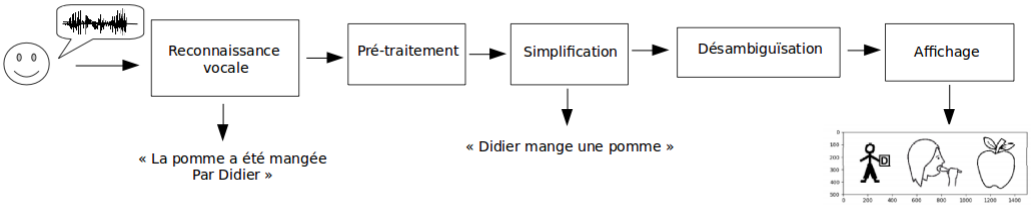
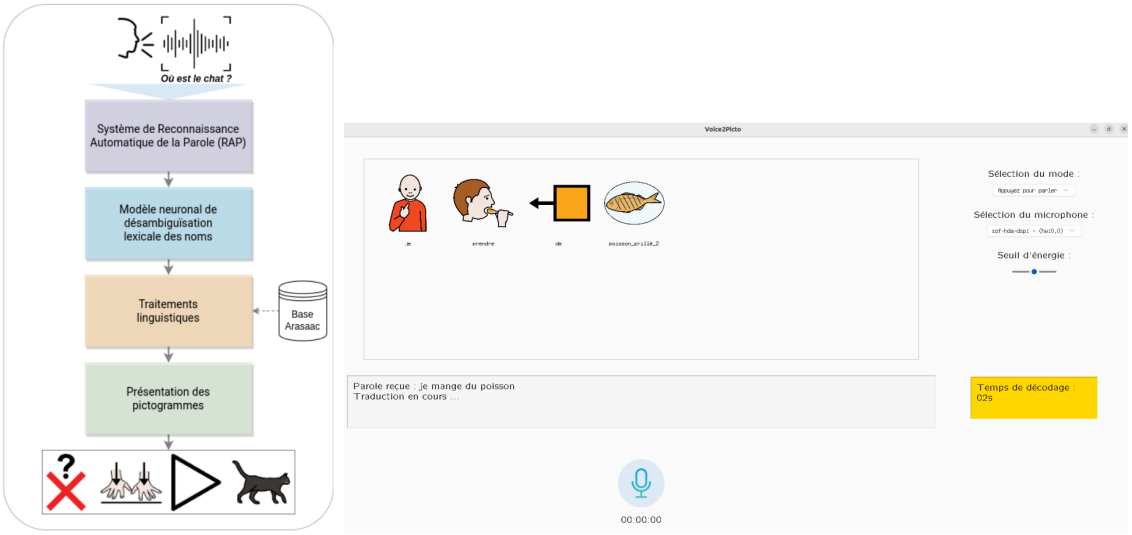


FIGURE 2.6 – Système speech-to-picto pour le français : architecture et exemple de la phrase « la pomme a été mangée par Didier », reproduite de Vaschalde (2018, p. 32)

9. Le Makaton combine la parole, les signes et les pictogrammes (https://www.makaton.fr).

À Grenoble, Macaire *et al.* (2022, 2023b) ont poursuivi ces recherches dans le cadre du projet PROPICTO (cf. section 1.4). Ils ont conçu Voice2Picto (Figure 2.7). [10] Celui-ci comprend toujours un système de reconnaissance automatique de la parole au début (Figure 2.7a), mais la désambiguïsation lexicale est située avant des traitements linguistiques (tokenisation, étiquetage morphosyntaxique et lemmatisation). Ils ont aussi proposé une interface ergonomique (Figure 2.7b). Un an plus tard, Macaire *et al.* (2024b) n'utilisent plus WordNet, mais des modèles de langue séquence-à-séquence de traduction neuronale (NMT, mBART, T5, NLLB), qu'ils règlent sur un corpus général parole/texte/pictogrammes qu'ils ont créé (Macaire *et al.*, 2024a,c).



(a) Architecture (b) Exemple de la phrase « je mange du poisson »

FIGURE 2.7 – Système Voice2Picto, reproduite de Macaire *et al.* (2023b, pp. 11-12)

Dans le cadre du Lab ImageCLEF2024, il y a eu la première édition de la tâche ToPicto à Grenoble, [11] en septembre 2024 (Ionescu *et al.*, 2024). L'objectif était de réunir la communauté scientifique (linguistes, informaticiens, traducteurs, etc.), afin de développer de nouvelles méthodes de traduction permettant soit de convertir un texte en entrée, soit un discours oral en une séquence de pictogrammes Arasaac. Deux tâches distinctes ont été proposées : text-to-picto et speech-to-picto. Des données – transcriptions associées à des pictogrammes – au format JSON ont été fournies aux participants. Elles ont été construites avec le corpus du français parlé TCOF, [12] qui contient des enregistrements d'interactions. Des approches utilisant différents modèles de langue sans WordNet ont donc été présentées. Koushik *et al.* (2024) ont évalué le modèle encodeur-décodeur T5. Elliah *et al.* (2024) ont affiné les modèles pré-entraînés GPT-2 et Helsinki-BERT en utilisant la simplification lexicale. Anand *et al.* (2024) ont testé des plongements basés sur CamemBERT avec une technique d'apprentissage contrastif.

10. https://github.com/macairececile/Voice2Picto
11. ToPicto est organisé dans le cadre de PROPICTO (https://www.imageclef.org/2023/topicto).
12. https://hdl.handle.net/11403/tcof

L'application Glyph (Figure 2.8) vise à traduire automatiquement les instructions de sortie des patients en pictogrammes en utilisant des techniques de TAL (Hill *et al.*, 2016; Zeng-Treitler *et al.*, 2014; Bui *et al.*, 2012). Elle propose une étape de prétraitement de texte, incluant la division de phrases, la normalisation de mots et de synonymes, etc. Ainsi, l'ordre de la phrase « *avoid driving if you are light-headed, drowsy or dizzy* » (« évite de conduire si tu es étourdi, somnolent ou pris de vertiges ») a été modifié, les deux pictogrammes représentant la condition précédent la recommandation (Figure 2.8). Le système exploite une terminologie – dont 15 groupes sémantiques d'UMLS (Bodenreider, 2004) – et des bases de données médicales (incluant 600 pictogrammes) avec des techniques graphiques pour générer des images. Il ne s'agit pas d'un outil de CAA, conçu et testé spécifiquement pour les personnes en situation de handicap, mais bien d'une première application développée dans le domaine de la santé qui améliore automatiquement du texte avec du TAL et des images.

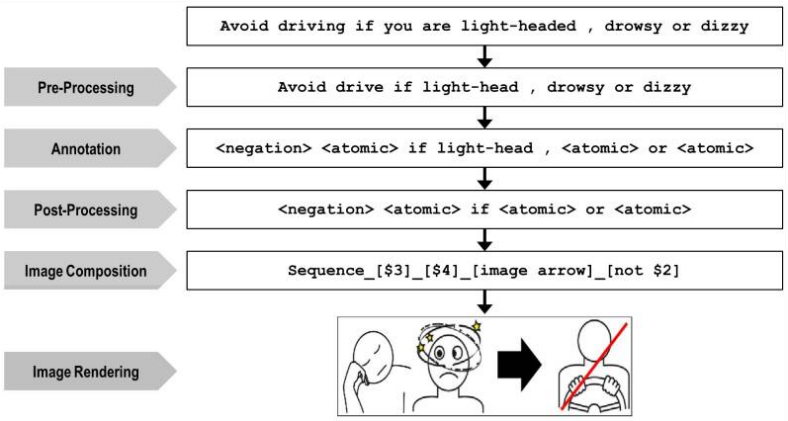


FIGURE 2.8 – Système Glyph pour l'anglais : architecture et exemple de la phrase « *avoid driving if you are light-headed, drowsy or dizzy* », reproduite de Bui *et al.* (2012, p. 1162)

Park *et al.* (2022) ont proposé PicTalky (Figure 2.9), un système de CAA pour l'anglais, qui est destiné aux enfants avec un handicap de développement langagier, leur famille et les soignants. Il exploite les réseaux de neurones pour la correction des erreurs de grammaire. En plus de faire de la reconnaissance vocale (speech-to-picto), il comprend un module text-to-picto par n-gramme, ainsi qu'un module de compréhension du langage naturel qui se charge du texte ne faisant pas partie du vocabulaire des pictogrammes Arasaac, incluant la résolution des coréférences et la similarité sémantique. Celle-ci est mesurée avec WordNet et Word2Vec (Mikolov *et al.*, 2013). Ce système fait aussi appel à la robotique, puisqu'il est connectable au robot NAO.

Bien qu'il ne s'agisse pas d'un système de traduction automatique, nous pouvons mentionner une approche récente, qui améliorerait la composition de messages dans des systèmes de CAA pour les personnes avec une DI. Il s'agit d'un modèle appelé PictoBERT (Pereira *et al.*, 2022), qui permet de faire de la prédiction de pictogrammes. [13]

13. Les systèmes de CAA utilisent souvent diverses techniques de TAL pour prédire et suggérer des

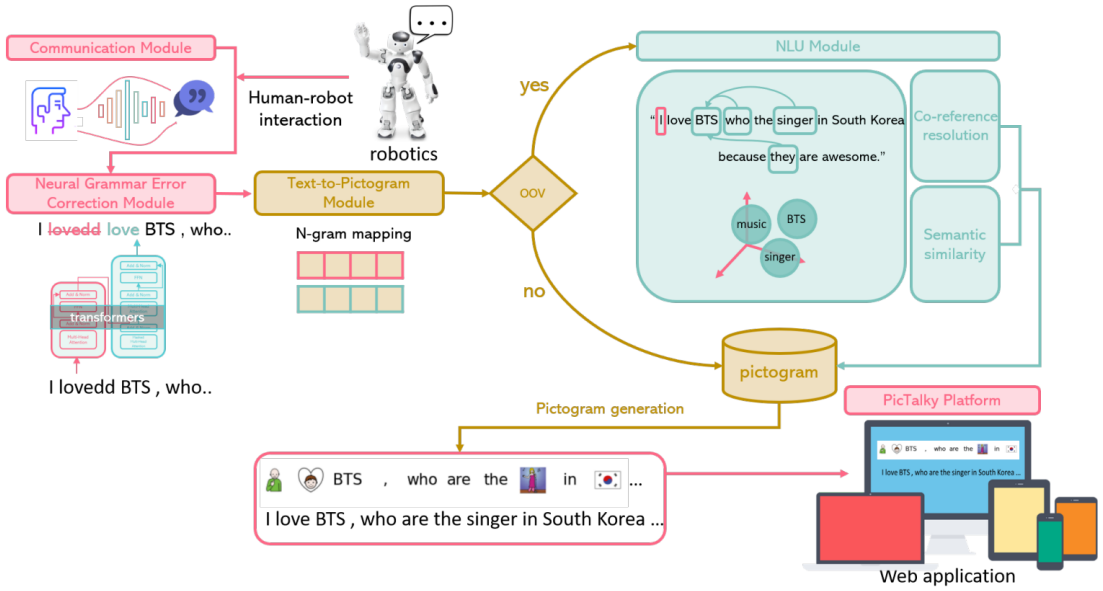


FIGURE 2.9 – Système PicTalky pour l'anglais : architecture et exemple de la phrase « *I lovedd BTS, who are the singer in South Korea* », reproduite de Park *et al.* (2022, p. 19)

1.3 Évaluation de la traduction vers les pictogrammes

Il faut distinguer l'évaluation du système de traduction automatique complet et l'évaluation d'un composant de ce système. Nous nous intéressons en particulier aux méthodes pour évaluer tout le système. Nous commençons par décrire les évaluations automatiques (section 1.3.1), manuelles (section 1.3.2), mais aussi certains tests utilisateurs réalisés dans le cadre de systèmes de traduction automatique de phrases vers les pictogrammes (section 1.3.3).

1.3.1 Évaluation automatique

En traduction, la métrique la plus utilisée pour les évaluations automatiques est la *BiLingual Evaluation Understudy (BLEU)*, développée par IBM (Papineni *et al.*, 2002). *National Institute of Standards and Technology (NIST)* de Doddington (2002) permet aussi d'évaluer la qualité du texte obtenu via la traduction automatique. Quant au *Word Error Rate (WER)* de Levenshtein (1966) et au *Position-independent word Error Rate (PER)*, ils sont plutôt utilisés dans le domaine de la reconnaissance vocale. Toutes ces métriques comparent la sortie du système – phrase candidate – à une ou plusieurs traduction(s) de référence. Une difficulté en traduction est le fait qu'il n'existe pas qu'une seule référence ; plusieurs phrases sont possibles et sont plus ou moins correctes – alors qu'en reconnaissance vocale, il y a qu'une seule phrase possible. La production de ces références par l'humain est relativement coûteuse en termes de temps

caractères, des mots ou des pictogrammes à l'utilisateur, afin d'améliorer la saisie de messages (Antoine et Maurel, 2007; Higginbotham *et al.*, 2012; Sevens *et al.*, 2015b; Norré et De Wilde, 2018; Norré, 2020a,b).

et celles-ci peuvent avoir un impact sur le score de ces métriques.

BLEU compare des n-grammes de poids identique. [14] NIST est une adaptation de BLEU, mais qui donne plus d'importance aux n-grammes les moins fréquents. WER et PER sont des métriques basées sur la distance d'édition. Une traduction est considérée comme meilleure si les scores BLEU et NIST sont élevés, contrairement au WER et au PER, qui indiquent que la reconnaissance est meilleure si ces taux sont faibles.

En ce qui concerne la traduction vers les pictogrammes, il s'agit d'une tâche de traduction monolingue, qui est basée sur les noms des images. Mihalcea et Leong (2008) avaient par exemple utilisé les métriques automatiques NIST et GTM. Sevens (2018, pp. 91, 235) avait réalisé une traduction manuelle de phrases en néerlandais, anglais et espagnol vers les pictogrammes Sclera et Beta, du mieux de ses capacités (« *to the best of our ability* »), afin de calculer des scores BLEU, NIST, WER et PER. Elle disposait d'un corpus d'emails écrits en pictogrammes Sclera et Beta. Elle a aussi traduit manuellement des tweets anglais et espagnols. Pour évaluer la qualité de son système, Vaschalde (2018, pp. 47-48) avait utilisé BLEU et réalisé deux traductions de référence : une version où tous les mots étaient traduits en pictogrammes et une version "simple". Son corpus d'évaluation était constitué d'histoires pour enfants en français. Le score BLEU a aussi été employé par Macaire *et al.* (2024b). Il s'agit de la métrique qui est la plus utilisée pour évaluer la traduction automatique de texte vers les pictogrammes.

Dans la récente tâche ToPicto, qui consistait à proposer des systèmes de traduction automatique du français vers les pictogrammes Arasaac – sur un corpus plus général que médical –, trois métriques automatiques ont été utilisées par les participants (Koushik *et al.*, 2024; Elliah *et al.*, 2024; Anand *et al.*, 2024) : BLEU, METEOR et *Picto-term Error Rate* ou PictoER (Woodard et Nelson, 1982). Pour une phrase source donnée (« tu peux pas savoir »), ces métriques comparent toutes l'hypothèse du système ou du modèle proposé – sous la forme d'une séquence de noms de pictogrammes (par exemple, « toi savoir non ») – à la référence (« toi pouvoir savoir non »). Contrairement à BLEU, METEOR (Banerjee et Lavie, 2005) exploite la racinisation et la synonymie. Cette métrique a aussi été utilisée par Sevens (2018, pp. 205, 212-213) pour évaluer son système de traduction Picto-to-Text, mais pas pour Text-to-Picto.

Park *et al.* (2022) ont proposé une métrique pour évaluer spécifiquement l'efficacité d'un module de traduction vers les pictogrammes : la *Text-to-Pictogram Accuracy* ou TPA. Elle a seulement été utilisée pour évaluer le système PicTalky. Les auteurs ont testé différents cas d'étude en fonction des catégories morphosyntaxiques (POS) à supprimer (déterminants, prépositions, conjonctions), mais aussi des mots vides. Ils ont implémenté un poids sur les entités nommées.

14. Un poids peut être ajouté pour la brièveté de la phrase candidate par rapport à la référence.

1.3.2 Évaluation manuelle

Les métriques automatiques sont indépendantes de la langue, mais ne donnent pas d'information sur le contenu de la traduction. C'est pourquoi, il est nécessaire de faire appel à des évaluateurs qui ont une certaine expertise de la langue cible. Il s'agit donc d'un jugement humain sur la qualité de la traduction. Pour établir un score global, ceux-ci utilisent souvent des échelles de Likert basées, par exemple, sur la fidélité et la fluidité (Callison-Burch *et al.*, 2006). Ce type d'évaluation est subjectif et prend plus de temps que les évaluations automatiques. En ce qui concerne l'évaluation manuelle de systèmes de traduction automatique de phrases vers les pictogrammes, Mihalcea et Leong (2008) avaient utilisé des taux de fidélité. Ils ont évalué si le concept était correctement représenté sur la base d'une échelle.

L'application Glyph a fait l'objet de plusieurs évaluations manuelles. L'exactitude de 49 instructions destinées aux patients (160 unités sémantiques), converties en images par Glyph a d'abord été évaluée par trois personnes (un infirmier, un designer graphique et un ingénieur logiciel). Elles ont utilisé une échelle subjective à trois points correspondant au niveau de représentativité des pictogrammes ; l'accord interannotateurs global était de 73,1 % (Bui *et al.*, 2012). Ces instructions de sortie ont ensuite été évaluées par 84 participants (Zeng-Treitler *et al.*, 2014), sans DI. Ils ont mesuré le rappel de ces informations après un certain temps. Leurs résultats ont montré une augmentation significative du rappel permettant d'affirmer que les personnes qui bénéficiaient le plus de ces illustrations étaient les moins de 40 ans qui n'avaient pas le niveau collège. Hill *et al.* (2016) ont poursuivi l'évaluation de ces instructions améliorées automatiquement avec des pictogrammes médicaux. Ils les ont comparées avec d'autres et ont montré que les instructions de Glyph permettaient d'améliorer le rappel et la satisfaction des patients interrogés.

Pour évaluer AraTraductor, Bautista *et al.* (2017) ont utilisé un corpus traduit en pictogrammes Arasaac ($n = 282$) : la Déclaration des droits des personnes handicapées. Ils ont calculé la précision, le rappel et le F-score sur le système baseline Pictotraductor et leur système avec des méthodes de TAL, qui a généré moins de pictogrammes et obtenu des F-scores plus élevés que celui qui effectue la traduction mot à mot. Pour évaluer la traduction vers les pictogrammes, Cabello *et al.* (2018) ont également utilisé cette Déclaration, ainsi que des contes pour enfants qui ont été illustrés manuellement par Arasaac ($n = 53$ phrases avec 245 lemmes différents et 10,5 mots en moyenne par phrase). Ils ont obtenu une précision moyenne de 74 %.

En comparant différentes techniques de désambiguïsation lexicale intégrées au sein de leur système, Imam *et al.* (2019) avaient obtenu, par exemple, un F-score de 89 % en utilisant la similarité maximale et 94 % avec un SVM. Quant à Sevens (2018, pp. 69-70), elle a évalué manuellement la précision, le rappel et le F-score de phrases

en pictogrammes Sclera et Beta pour les versions néerlandaise, anglaise et espagnole du Text-to-Picto de base comme nous le verrons dans le prochain chapitre. Afin de comparer le système Text-to-Picto de base avec le système complet (néerlandais vers Arasaac), Sevens (2018, pp. 167-174) a aussi réalisé des évaluations humaines objectives de l'exactitude avec 8 chercheurs divisés en deux groupes (les utilisateurs experts et les utilisateurs compétents). Ils avaient pour tâche de fournir des transcriptions de 15 phrases en pictogrammes (uniquement pour les mots de contenu). [15] Elle a effectué après une évaluation subjective de la qualité sur la base d'une échelle à cinq points mesurant la préservation du sens. La différence d'exactitude (globale) entre les transcriptions des traductions générées par le système de base et celles des générées par le système complet était significative en utilisant un test de Student pairé.

Bulté *et al.* (2021) ont également évalué la compréhension et la couverture lexicale de phrases générées automatiquement en pictogrammes Sclera, Beta et Rand avec Text-to-Picto pour le néerlandais sur trois corpus différents (cours de langue, articles de presse, sites Internet des gouvernements locaux). Afin de juger de l'utilité des traductions pour des migrants, le premier auteur et un évaluateur externe ont utilisé deux échelles à quatre points (« *No pictographs* », « *Not helpful* », « *Partially helpful* » et « *Potentially fully helpful* »). En plus d'avoir proposé une métrique automatique pour l'évaluation de leur système de CAA, Park *et al.* (2022, pp. 26-27) ont réalisé une enquête pour évaluer la satisfaction. Ils ont demandé à 53 personnes – 43 experts dans le domaine des troubles du langage et 10 enfants non verbaux – de répondre à cinq questions avec une échelle de Likert à trois points (« *Satisfied* », « *Neither agree nor disagree* », « *Dissatisfied* »). Ils expliquent qu'il est en effet difficile de réaliser des enquêtes à large échelle avec des enfants non verbaux. Leurs résultats ont notamment montré qu'entre 80 et 90 % des réponses étaient « Satisfait » de leur module de traduction de texte vers les pictogrammes.

Pour évaluer la qualité des traductions en pictogrammes Arasaac générées à partir de la parole française, Macaire *et al.* (2024a,b) ont effectué des évaluations humaines. Ils ont adapté un cadre analytique appelé *Multidimensional Quality Metrics* ou MQM (Burchardt, 2013). [16] Il s'agit d'une évaluation qui consiste à analyser les erreurs entre le texte source et le texte cible au niveau de la précision, de la fluidité, de la vérité et du design. Pour ce faire, deux experts de leur projet ont annoté manuellement le niveau de gravité des erreurs associé à un poids (neutre : 0, mineur : 1, majeur : 5, critique : 25). Les auteurs ont ainsi obtenu un score de qualité globale et ont pu comparer les traductions générées par leurs meilleurs modèles sur trois corpus (Macaire *et al.*, 2024b).

15. https://goo.gl/forms/Cafc1vyTUker8BOn1 ; https://goo.gl/forms/kmYok3WaBpQzy9dA3
16. https://themqm.org

1.3.3 Étude utilisateur

Cette sous-section présente des évaluations d'applications – qui intègrent un outil – de traduction automatique de phrases vers les pictogrammes, ayant fait l'objet de tests auprès de potentiels utilisateurs, les personnes avec une DI ou leur entourage.

Afin d'analyser les défauts du système de base de la version néerlandaise de Text-to-Picto, Sevens (2018, pp. 70-72) avait réalisé des études utilisateurs, dont un focus group et observé des utilisateurs cibles dans un centre de jour en Flandre : 5 adultes avec une DI et 2 entraîneurs. Durant les sessions pratiques, les participants avec une DI ont donc utilisé Text-to-Picto pour traduire du texte sur leur fil d'actualité et ont ensuite répondu à des questions orales qu'elle avait préparées : s'ils aimaient utiliser le système Text-to-Picto, s'ils aimeraient continuer à utiliser les réseaux sociaux, si les solutions text-to-speech sont plus pratiques que les pictogrammes et qu'est-ce qui rendait Text-to-Picto difficile à utiliser. C'est cette expérience qui a montré qu'il était nécessaire d'ajouter d'autres modules (le correcteur orthographique adapté aux capacités des personnes avec une DI, la simplification syntaxique pour réduire la complexité de certaines structures et la désambiguïsation lexicale).

Le module de simplification syntaxique de Text-to-Picto a fait l'objet d'études utilisateurs auprès de personnes faisant partie de l'environnement proche de ses utilisateurs cibles avec une DI (Sevens, 2018, pp. 136-148). Il s'agissait de familles, d'enseignants, de logopèdes ou de soignants qui devaient répondre en pensant à la personne avec une DI, car son questionnaire était trop long pour elle (47 questions). Sur la base de 12 phrases en pictogrammes (Sclera ou Beta) simplifiées ou non avec Text-to-Picto, ses 28 participants devaient noter le niveau de difficulté sur une échelle de 1 à 10 pour l'utilisateur cible. Les répondants devaient aussi traduire la phrase en pictogrammes, même si le sens de chaque pictogramme était donné. Elle a ainsi obtenu des résultats quantitatifs. Sur les 6 phénomènes syntaxiques évalués, 3 comportaient une différence significative entre la version de base et la version simplifiée : la simplification des propositions relatives, des subordonnées et des subordonnées avec la voix passive.

L'application mobile *Travel and Communication Assistant* (Kultsova *et al.*, 2017) a été évaluée par de jeunes enfants avec un Trouble du Spectre de l'Autisme (TSA) dans un centre de réadaptation. D'après leurs résultats, les spécialistes du centre ont observé que les enfants pouvaient communiquer leurs besoins et leur état de santé par messages sous forme de pictogrammes générés par le système. Ils ont reporté une diminution significative de certains comportements indésirables. Ils ont également noté ceci : « *children were improving their attention and memory, expanding the scope of perception, activating the impressive and expressive speech, began composing and pronouncing simple phrases from single words* » (Kultsova *et al.*, 2017, p. 297). Cependant, il n'y a pas plus de détails sur leurs tests dans cet article. Les résultats sont qualitatifs.

1.4 Projet PROPICTO : Genève

Une partie de cette thèse s'inscrit dans le projet franco-suisse (2021–2024), appelé PRojection du langage Oral vers des unités PICTOgraphiques (PROPICTO), entre l'Université Grenoble Alpes (GETALP) et l'Université de Genève (TIM). Il est dirigé respectivement par Benjamin Lecouteux et Pierrette Bouillon (Ormaechea Grijalba *et al.*, 2023a,b). [17] Nous avons décidé de présenter plus en détail la partie suisse du projet, qui vise à développer un système de traduction automatique de phrases en pictogrammes pour le dialogue médical. Nous revenons sur les origines de celui-ci, en décrivant le système BabelDr utilisé à Genève (section 1.4.1). Nous décrivons ensuite PictoDr (section 1.4.2). Nous comparerons les traductions médicales en pictogrammes de ce système avec celles générées par Text-to-Picto dans le chapitre 6.

1.4.1 Système BabelDr

BabelDr [18] est un système de traduction de la parole pour la communication entre les médecins et les patients allophones (Bouillon *et al.*, 2021; Spechbach *et al.*, 2019; Bouillon *et al.*, 2017), y compris les patients sourds. Le médecin pose oralement ses questions de façon libre. Elles sont ensuite traduites automatiquement dans une des langues cibles (albanais, arabe, dari, espagnol, farsi, russe, tigrigna, ukrainien, anglais simple ou langue des signes de Suisse romande). Le système répond à trois critères principaux : sécurité des données, fiabilité de la traduction et portabilité à de nouvelles langues (peu dotées), phrases ou domaines (Norré *et al.*, 2021a). Le projet est mené par le Département de Traitement Informatique Multilingue de l'Université de Genève, ainsi que les Hôpitaux Universitaires de Genève (HUG) où le système est déjà utilisé au service ambulatoire depuis 2018. Il a également été testé récemment dans des pharmacies en Suisse (Mali, 2023).

Le système BabelDr est un exemple de traducteur à phrases fixes. Il fonctionne avec un corpus de 10 000 phrases types (instructions et questions médicales), qui est dérivé de l'échelle de tri des urgences hospitalières suisses. [19] Les auteurs ont demandé à des médecins de poser les questions pour les points de l'échelle de tri. Elles ont ensuite été reformulées pour le système de BabelDr et traduites dans plusieurs langues, incluant les pictogrammes Arasaac. [20] Ce système a la particularité de lier le résultat de la reconnaissance vocale à la phrase pré-traduite la plus proche avec des techniques

17. Le projet PROPICTO (https://propicto.unige.ch) est financé par l'Agence Nationale de la Recherche (ANR-20-CE93-0005) et le Fonds National Suisse (N°197864).

18. https://babeldr.unige.ch ; https://babeldr.app.unige.ch

19. L'échelle de tri (ou triage) est un outil qui permet de prédire la complexité et la sévérité d'un patient venu consulter aux urgences. Il s'agit d'une liste de points à vérifier par domaine et par niveau d'urgence. Elle peut être différente par pays. Dans cette thèse, nous ferons toujours référence à celle utilisée en Suisse (l'Echelle Suisse de Tri ou EST®).

20. https://propicto.demos.unige.ch/pictoDrClient/corpora

neuronales. Pour ce faire, il utilise des données synthétiques, dérivées avec une grammaire (Bouillon *et al.*, 2017), qui lie des millions de variantes (par exemple, « avez-vous eu de la température ? », « est-ce que vous avez de la fièvre ? », etc.) aux phrases types (« avez-vous de la fièvre ? »). Les variantes sont des données synthétiques en raison du peu de données disponibles concernant les dialogues médicaux.

Nous avons eu l'occasion de travailler sur ce système dans le cadre de premiers travaux (Norré *et al.*, 2020, 2021a,b; Bouillon *et al.*, 2023). Nous avons réalisé des études préliminaires visant à évaluer des images avec des questionnaires en ligne et des tests utilisateurs par visioconférence. Nous avons testé la compréhension de pictogrammes Arasaac et Sclera associés aux phrases de BabelDr avec divers utilisateurs sans DI, afin de concevoir l'interface de réponses bidirectionnelle de BabelDr (Figure 2.10).

FIGURE 2.10 – Système BabelDr pour plusieurs langues : exemple de la phrase « pouvez-vous me montrer ce qui vous amène ? » traduite en arabe avec des réponses possibles en pictogrammes Arasaac, reproduite de Norré *et al.* (2021b, p. 56)

Pour ce faire, nous avons lié manuellement un ensemble limité de pictogrammes (entre 3-61) aux phrases types du système. Le patient peut répondre à la question posée par le médecin en sélectionnant un ou plusieurs pictogramme(s) qui constitue(nt) une possible réponse – le nom des pictogrammes est affiché en français pour le médecin. Bien qu'ils soient disposés sous la forme d'une grille comme dans les tableaux de communication en CAA, il n'est pas possible de former une phrase complète en pictogrammes, les mots de fonction ne sont pas proposés. Nous avons mis en correspondance les phrases types avec les pictogrammes constituant les réponses possibles.

1.4.2 Système PictoDr

Dans un premier temps, trois systèmes de traduction neuronale ont été entraînés par Mutal *et al.* (2022) avec OpenNMT toolkit – sur la base des données de BabelDr[21] –, afin de traduire automatiquement les phrases de BabelDr (phrases types et variantes) vers : **a)** une glose sémantique constituée d'une séquence de concepts du méta-thésaurus *Unified Medical Language System (UMLS)* de Bodenreider (2004) ; **b)** ces concepts sont ensuite traduits en pictogrammes Arasaac[22] (Figure 2.11).

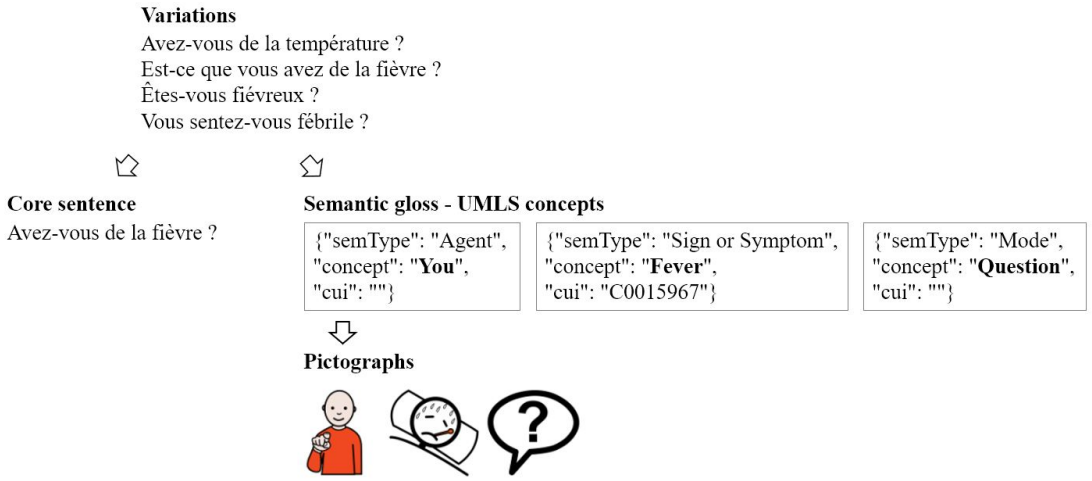


FIGURE 2.11 – Système PictoDr UMLS-Arasaac pour le français : exemple de la phrase type de BabelDr « avez-vous de la fièvre ? », reproduite de Mutal *et al.* (2022, p. 254)

(a) La première étape a fait l'objet d'évaluations manuelles (Mutal *et al.*, 2022). Le meilleur modèle a obtenu un F_β (avec $\beta = 0{,}5$) de 0,882 sur les concepts UMLS des phrases du domaine. Deux juges (dont l'auteur) ont chacun annoté les résultats de la traduction neuronale – 882 représentations de phrases médicales de BabelDr en séquences de concepts UMLS – selon les catégories suivantes : « Même sens », « Hyperonyme/hyponyme », « Partie » (si deux idées), « Contresens » (au moins un concept est faux), « Absence d'(au moins) un élément » (concept) important, « Ajout d'un élément » (concept) pas nécessaire, « Je ne sais pas ». Elles ont été fusionnées en : « *Same Meaning* » (Même sens), « *Different Meaning* » (Contresens), « *Related Meaning* » (Hyperonyme/hyponyme, Partie, Absence ou Ajout d'un élément), « *I don't know* » (Je ne sais pas).[23] L'étude a montré qu'environ 80 % (moyenne pour les deux juges) des concepts générés par ces modèles sont lisibles par les humains et peuvent potentiellement caractériser la langue pictographique (Table 2.1).

21. Il s'agit d'une ressource synthétique en raison du manque de données (dialogues médicaux). Elle est présentée par Mutal *et al.* (2022), Gerlach *et al.* (2023), puis plus en détail par Gerlach *et al.* (2024). Ce corpus contient des liens entre les variantes, les phrases types et les concepts UMLS.

22. Nous avons participé à la création de la première version de la ressource JSON (Mutal *et al.*, 2022; Gerlach *et al.*, 2023), qui lie les concepts UMLS aux pictogrammes Arasaac (https://propicto.demos.unige.ch/pictoDrClient/overview).

23. Cette évaluation est inspirée des taxonomies d'analyse d'erreurs en traduction automatique.

	Same Meaning	Different Meaning	Related Meaning	I don't know	Total
In Domain	67,20 % | 70,53 %	16,77 % | 15,29 %	13,56 % | 14,18 %	2,47 % | 0 %	811
Out Of Domain	08,45 % | 14,08 %	67,61 % | 64,79 %	21,13 % | 21,13 %	2,82 % | 0 %	71
In Coverage	82,43 % | 89,86 %	06,08 % | 06,08 %	10,81 % | 04,05 %	0,68 % | 0 %	148
Out of Coverage	58,45 % | 61,17 %	23,84 % | 21,93 %	14,85 % | 16,89 %	2,86 % | 0 %	734
Moyenne	**62,47 % | 65,99 %**	**20,86 % | 19,27 %**	**14,17 % | 14,74 %**	**2,49 % | 0 %**	**882**

TABLE 2.1 – Résultats de PictoDr pour la traduction de phrases vers une glose sémantique UMLS selon deux annotateurs (A1-A2), adaptée de Mutal *et al.* (2022, p. 259)

(b) Nous donnons des exemples de traductions en pictogrammes de PictoDr (version de 2023), avec le meilleur modèle (Figure 2.12). Gerlach *et al.* (2024) ont ensuite conçu de nouveaux pictogrammes et adapté des pictogrammes existants. En plus de proposer une nouvelle visualisation des pictogrammes sous la forme d'une bande dessinée (version de 2024), ils ont évalué d'autres modèles de langue pré-entraînés pour PictoDr : CamemBERT (Martin *et al.*, 2020), FlauBERT (Le *et al.*, 2020), DrBERT (Labrak *et al.*, 2023a,b) et XLM-R (Conneau *et al.*, 2020).[24] Selon leur évaluation (F-score sur deux corpus),[25] le meilleur modèle est CamemBERT (Gerlach *et al.*, 2024).

FIGURE 2.12 – Exemples de traductions médicales de PictoDr en pictogrammes Arasaac, reproduite de Mutal *et al.* (2022, p. 259)

Dans le chapitre 6, nous évaluerons la deuxième étape, c'est-à-dire la traduction de phrases médicales en pictogrammes Arasaac générées par PictoDr (une version du système de 2023 et une de 2024). Nous comparerons la compréhension de ces traductions avec celle de phrases générées par Text-to-Picto chez des adultes avec une DI.

24. Ces systèmes neuronaux traduisant des phrases médicales vers les pictogrammes sont disponibles en démo avec la reconnaissance vocale (https://propicto.demos.unige.ch/pictodrclient).

25. https://propicto.demos.unige.ch/lrec2024

1.5 Systèmes et évaluation de la traduction vers les pictogrammes

La Table 2.2 reprend tous les systèmes de traduction automatique vers les pictogrammes avec des techniques de TAL et leurs méthodes d'évaluation.

Bien qu'elle ne soit pas exhaustive, elle montre qu'à l'exception des systèmes développés dans le cadre du projet franco-suisse PROPICTO, aucun n'utilisait la traduction automatique neuronale – lorsque nous avons commencé cette thèse –, un seul concernait le domaine médical (Bui *et al.*, 2012; Zeng-Treitler *et al.*, 2014; Hill *et al.*, 2016) et il n'existait qu'un outil pour le français (Vaschalde, 2018; Vaschalde *et al.*, 2018). La langue cible Arasaac est souvent choisie, étant donné qu'il s'agit d'un grand ensemble de pictogrammes libres de droit en CAA. Les exemples de traduction en pictogrammes présentés dans les études sont souvent des phrases simples et déclaratives (pas de questions), de type sujet + verbe + objet (sans proposition subordonnée).

En raison du peu de données disponibles sous la forme pictographique, les systèmes text-to-picto exploitent différentes techniques de TAL, souvent basées sur des règles, en vue d'améliorer leurs traductions en pictogrammes. Beaucoup de ces systèmes emploient la ressource lexico-sémantique WordNet. En plus de la base de données d'images (dictionnaire), ils sont composés de diverses sous-tâches du TAL : un prétraitement de texte ou une analyse linguistique superficielle – tokenisation, étiquetage morphosyntaxique, lemmatisation – au début, une désambiguïsation lexicale, mais aussi une simplification, dont nous allons présenter un aperçu de la discipline.

Comme nous l'avons déjà mentionné, la métrique automatique la plus utilisée pour la traduction vers les pictogrammes est le score BLEU, même si d'autres ont été employées (METEOR, NIST, etc.). La taille des corpus d'évaluation en pictogrammes est limitée à quelques centaines de phrases, voire même parfois à quelques centaines de pictogrammes. Des évaluations manuelles avec des humains sont réalisées sur la base d'échelles ou de la précision, du rappel et du F-score. Les études utilisateurs sur des phrases traduites en pictogrammes par un système avec le public cible, en particulier si celui-ci a une DI sont toutefois très rares et peu documentées.

Système et référence(s)	**Techniques et caractéristiques (non exhaustives)**	**Langue(s) source(s)**	**Langue(s) cible(s)**	**Médical?**	**Évaluation automatique (métrique(s)), manuelle et/ou étude utilisateur**
(Mihalcea et Leong, 2008)	WordNet anglais (noms et verbes), étiquetage morphosyntaxique, lemmatisation	Anglais	Images, photos	Non	Auto (NIST, GTM), manuelle
Glyph (Bui *et al.*, 2012; Zeng-Treitler *et al.*, 2014; Hill *et al.*, 2016)	UMLS, prétraitement, annotation, règles	Anglais	Images	Oui	Manuelle
AraTraductor (Bautista *et al.*, 2017)	Analyse syntaxique, étiquetage morphosyntaxique (avec cadre coloré), lemmatisation, prétraitement, gestion des expressions multi-mots (modèle n-gramme)	Espagnol	Arasaac	Non	Manuelle
(Cabello *et al.*, 2018)	Résumé automatique de texte, analyse sémantique latente	Espagnol	Arasaac	Non	Manuelle
(Imam *et al.*, 2019)	WordNet anglais, tokenisation, étiquetage morphosyntaxique, lemmatisation, désambiguïsation lexicale	Anglais	ImageNet	Non	Manuelle
PicTalky (Park *et al.*, 2022)	WordNet anglais, correction orthographique neuronale, résolution des coréférences, similarité sémantique, speech-to-picto, robotique	Anglais	Arasaac	Non	Auto (TPA)
(Kultsova *et al.*, 2017)	WordNets russe et anglais, lemmatisation	Russe	Sclera, Beta, PECS	Non	Étude utilisateur
Text-to-Picto (Sevens *et al.*, 2015a; Sevens, 2018)	WordNets anglais et espagnol (relations sémantiques), tokenisation, étiquetage morphosyntaxique, lemmatisation, algorithme A*	Anglais, Espagnol	Sclera, Beta	Non	Auto (BLEU, NIST, WER, PER), manuelle
Text-to-Picto (Vandeghinste et Schuurman, 2014; Sevens *et al.*, 2016a,b, 2017; Vandeghinste *et al.*, 2017; Sevens, 2018; Bulté *et al.*, 2021)	WordNet néerlandais (relations sémantiques), tokenisation, étiquetage morphosyntaxique, lemmatisation, algorithme A*, correction orthographique automatique, simplification syntaxique (règles), désambiguïsation lexicale, reconnaissance des "entités nommées" avec liens vers des images de Wikipedia	Néerlandais	Sclera, Beta, Rand, Wiki	Non	Auto (BLEU, NIST, WER, PER), manuelle, étude utilisateur (focus group)
(Vaschalde, 2018; Vaschalde *et al.*, 2018)	WordNet anglais, simplification (voix passive), reconnaissance des entités nommées, désambiguïsation lexicale, speech-to-picto	Français	Arasaac, Makaton	Non	Auto (BLEU)
Voice2Picto (Macaire *et al.*, 2022, 2023b) *	Cf. Vaschalde (2018), Vaschalde *et al.* (2018), interface ergonomique	Français	Arasaac	Non	–
(Macaire *et al.*, 2023a, 2024a,b) *	Modèles de langue (NMT, mBART, T5, NLLB), speech-to-picto	Français	Arasaac	Non	Auto (BLEU, METEOR, PictoER), manuelle
CLEF2024-ToPicto (Koushik *et al.*, 2024) *	Modèle de langue (T5)	Français	Arasaac	Non	Auto (BLEU, METEOR, PictoER)
CLEF2024-ToPicto (Elliah *et al.*, 2024) *	Modèles de langue (GPT-2 et Helsinki-BERT), simplification lexicale	Français	Arasaac	Non	Auto (BLEU, METEOR, PictoER)
CLEF2024-ToPicto (Anand *et al.*, 2024) *	Modèle de langue (CamemBERT), apprentissage contrastif	Français	Arasaac	Non	Auto (BLEU, METEOR, PictoER)
PictoDr (Mutal *et al.*, 2022; Gerlach *et al.*, 2023, 2024) **	Modèles de langue (CamemBERT, FlauBERT, DrBERT, XLM-R), UMLS, speech-to-picto	Français	Arasaac, SantéBD	Oui	Manuelle

* PROPICTO (Grenoble) | ** PROPICTO (Genève)

TABLE 2.2 – Systèmes de traduction automatique vers les pictogrammes avec TAL

2 Simplification de textes

La simplification consiste à adapter un texte à différents niveaux (lexical, syntaxique, sémantique ou discursif), tout en préservant le sens. Elle peut être réalisée manuellement par des humains (section 2.1), mais aussi automatiquement par des machines (section 2.2). Tout comme pour la traduction automatique, il existe des méthodes pour évaluer la simplification. Nous ne les présenterons pas, car nous évaluerons toujours la tâche de traduction automatique (vers les pictogrammes) et non celle de simplification. Cette dernière sera en effet utilisée comme une étape de prétraitement de textes au sein de Text-to-Picto (l'évaluation est donc extrinsèque).

2.1 Simplification manuelle

Simplifier un texte permet de le rendre plus accessible pour des « publics confrontés à des difficultés de lecture et/ou de compréhension » (Gala *et al.*, 2018, p. 123). Il existe différents projets qui visent à améliorer la compréhension de textes sur différents supports (livres, fascicules, web, etc.), à destination de différents publics (des enfants aux personnes âgées, sans ou en situation de handicap, etc.). Nous pouvons par exemple citer la collection belge de romans, appelée La Traversée. [26] Elle est publiée depuis 2012 pour les adultes débutant en lecture à l'initiative de l'association Lire et Écrire Luxembourg. Les auteurs doivent suivre certaines recommandations et les textes sont éprouvés auprès d'un échantillon de lecteurs cibles.

En 2009, le projet européen Pathways a donné naissance au Facile A Lire et à Comprendre (FALC), une série de règles destinées à rendre l'information plus accessible aux personnes avec une DI (Inclusion Europe, 2009). Le guide fait 50 pages. Il recommande par exemple d'utiliser des mots faciles à comprendre, de faire des phrases simples et courtes, d'éviter les pronoms ou le passif et de privilégier le présent. Il a d'ailleurs été écrit en FALC sous la forme d'une liste de conseils numérotés, souvent à l'impératif (« utilisez », « n'écrivez pas », etc.), en commençant par des règles générales concernant les informations faciles à comprendre (20 règles). Les règles spécifiques sont ensuite divisées en quatre sections selon le type d'information : écrites (42 règles), électroniques (30 règles), vidéo (7 règles) et audio (17 règles). Le FALC est l'équivalent francophone du *Easy-(to-)Read* (Vandeghinste *et al.*, 2021). Il est de plus en plus utilisé pour tous types de documents, dont les sites d'information à destination des personnes avec une DI (Chehab *et al.*, 2019).

Dans les règles spécifiques aux informations écrites du FALC, il est aussi recommandé d'illustrer les textes avec des images pour aider les personnes qui ont des difficultés de lecture à mieux les comprendre (Inclusion Europe, 2009, pp. 24-25). Il peut

26. https://lire-et-ecrire.be/latraversee ; https://www.collectionlatraversee.be

s'agir de photographies, de dessins ou de symboles, incluant les pictogrammes. Des conseils avec des exemples sont donnés quant au choix des images (« utilisez toujours les bonnes images adaptées aux gens à qui vous vous adressez », « utilisez toujours des images claires, faciles à comprendre et qui correspondent à la partie du texte qu'elles sont censées expliquer », « utilisez la même image pour décrire la même chose dans tout votre document » ou « lorsque c'est possible, essayez d'utiliser le même type d'image à travers tout le document »), mais aussi sur la façon de les utiliser (« vous pouvez placer 1 ou 2 images près d'un paragraphe pour expliquer ce dont parle le texte »). En 2020, l'Unapei a publié une liste de vérification des documents en FALC. Chaque recommandation est évaluée sur une échelle de 1 (critère important qui équivaut à 2 points) à 3 (critère prioritaire, 6 points). La présence d'images est jugée prioritaire pour aider les personnes à comprendre de quoi parle un texte. Une note globale de 80 % est nécessaire pour qu'un texte soit considéré comme facile à comprendre (Unapei, 2020). La participation de personnes avec une DI au processus de rédaction est obligatoire.

En plus d'être très générales, ces règles sont assez subjectives. Le lecteur peut se demander ce qu'est une image adaptée au public, claire et facile à comprendre. Ce type de critères manque de consensus et ne repose pas sur une base empirique – « certains guides recommandent de placer les images à gauche et d'autres en complément à droite du texte » (Diacquenod et Santi, 2018, pp. 32-33). On retrouve toutefois différents supports, comme des brochures, qui proposent des images et des textes en FALC. [27] Nous avons vu que les images aident les personnes à comprendre et se souvenir de ce qui a été dit. Une solution pour soutenir la communication des personnes avec une DI et enrichir leur vocabulaire consiste à utiliser des fiches de dialogue (Cataix-Nègre, 2011). Cette « pratique élargie des images » permet de préparer les consultations.

Les fiches de dialogue en FALC de SantéBD [28] sont un exemple de CAA (Figure 2.13). Elles ont été conçues pour améliorer la communication des personnes en situation de handicap lors de rendez-vous médicaux. Disponibles en plusieurs langues, elles sont utilisées dans des situations connues d'avance, afin de faciliter le dialogue entre les professionnels de la santé et les patients. Ces derniers communiquent en montrant les images au médecin comme lorsqu'ils s'expriment avec un classeur de communication. Dans ce cas-ci, les images sont plus riches sémantiquement, car elles représentent une situation décrite par une phrase en FALC et non sur la base d'un seul mot. SantéBD propose une centaine de BDs personnalisables, qui couvrent une quinzaine de spécialités (maladies, examens médicaux, etc.). Les images de SantéBD ont été intégrées au sein du système PictoDr (Mutal *et al.*, 2022; Gerlach *et al.*, 2023).

27. Le livre blanc de l'Unapei (Biotteau *et al.*, 2013) a par exemple été traduit manuellement en FALC et illustré avec des pictogrammes Sclera et Arasaac choisis par des personnes avec une DI (https://www.unapei.org/publication/livre-blanc-pour-une-sante-accessible-aux-personnes-handicapees).

28. Réalisée par l'association CoActis Santé, cette banque d'images est disponible librement sous la forme d'une application sur tablette ou smartphone et sur Internet (https://santebd.org).

FIGURE 2.13 – Images SantéBD : exemple de la fiche sur les gestes simples contre la COVID-19, reproduite de Bouillon *et al.* (2023, p. 132)

Il existe des formations organisées par des associations pour apprendre à rédiger en FALC à divers professionnels, mais aussi pour les personnes avec une DI, afin de leur apprendre à relire des textes et à dire quand ceux-ci sont difficiles (Vandeghinste *et al.*, 2021). En septembre 2021, l'association belge Inclusion a par exemple lancé un service de traduction, formation et promotion du FALC. [29] Les personnes souhaitant faire traduire leur texte en FALC doivent introduire une demande. L'association leur propose ensuite un devis qui peut varier en fonction de plusieurs facteurs : le type de service (traduction et/ou vérification de textes déjà traduits en FALC), le nombre de pages à traduire, sa difficulté (facile, moyen, difficile), etc. Le demandeur peut choisir le type d'illustration qu'il préfère (pictogrammes, dessins ou photos). L'association propose aussi un service d'interprétation en FALC pour les conférences, les rendez-vous, etc.

En France, certaines organisations publiques et ESAT (Établissement et Service d'Aide par le Travail) proposent la transcription de documents en FALC. Dans tous les cas, chaque paragraphe de texte est associé à une image et le développement de documents en FALC nécessite toujours l'implication de personnes avec une DI (lecture, vérification et approbation), ce qui est coûteux et prend du temps (Chehab *et al.*, 2019). En Suisse, il existe également un bureau Langage simplifié avec lequel nous collaborerons pour nos études utilisateurs, afin d'évaluer nos phrases en pictogrammes.

29. https://www.falc.be

2.2 Simplification automatique

La simplification automatique de textes est une tâche du TAL, qui vise à appliquer des transformations sur un texte, afin de réduire sa complexité, tout en conservant son sens. [30] Beaucoup de ces systèmes ont été conçus pour l'anglais, bien qu'il en existe pour d'autres langues comme l'espagnol (Saggion *et al.*, 2015), l'italien (Barlacchi et Tonelli, 2013), le néerlandais (Bulté *et al.*, 2018; Seidl et Vandeghinste, 2024), le français (Seretan, 2012; Brouwers *et al.*, 2014; Rolin *et al.*, 2021; Todirascu *et al.*, 2022), etc. La simplification automatique peut être effectuée à différents niveaux (Koptient *et al.*, 2019, p. 309) : lexical, syntaxique, sémantique, pragmatique et stylistique.

Dans cette section, nous proposons de donner un bref aperçu de différents aspects de la simplification automatique de textes, en insistant sur les études pour la langue française : la simplification lexicale (section 2.2.1), la simplification syntaxique (section 2.2.2), la simplification neuronale (section 2.2.3), la simplification médicale (section 2.2.4), les ressources linguistiques disponibles pour la simplification (section 2.2.5), ce qui nous amène à la problématique des langues peu dotées (section 2.2.6), ainsi qu'aux applications de la simplification automatique pour les humains et les machines telles que la CAA pour les personnes avec une DI (section 2.2.7).

2.2.1 Simplification lexicale

En ce qui concerne la simplification lexicale, Paetzold et Specia (2017) résument les étapes comme suit : (0) un système de simplification lexicale prend en entrée une phrase, (1) celui-ci commence par identifier un mot compliqué, (2) il génère alors plusieurs substitutions possibles pour ce mot, (3) il sélectionne les substitutions en fonction de leur grammaticalité et de leur adéquation sémantique en contexte de la phrase, (4) il classe les substitutions candidates sélectionnées en fonction de leur simplicité. Pour chacune de ces étapes, plusieurs techniques et ressources peuvent être utilisées. Nous ne présenterons pas les méthodes, mais elles font aussi bien appel à des réseaux de neurones, de l'apprentissage automatique (non-)supervisé qu'à des calculs de fréquences ou des règles. [31] Pour le français, cette architecture a été mise en œuvre par Rolin *et al.* (2021).

2.2.2 Simplification syntaxique

Les systèmes de simplification syntaxique comptent plusieurs approches. Certaines font appel à des règles souvent produites à la main (Chandrasekar *et al.*, 1996; Seretan, 2012; Brouwers *et al.*, 2014; Todirascu *et al.*, 2022) – reposant sur une expertise

30. Cet état de l'art ne se veut pas exhaustif. Nous invitons le lecteur intéressé à se référer aux synthèses de Saggion (2017), Al-Thanyyan et Azmi (2021).

31. Le lecteur peut consulter la synthèse sur la simplification lexicale de Paetzold et Specia (2017).

linguistique ou des outils comme des analyseurs syntaxiques et des étiqueteurs. Pour le français, Seretan (2012) a observé un corpus de textes originaux et simplifiés avant de définir manuellement environ 40 règles de simplification et proposer une technique semi-automatique pour générer de nouvelles règles. Brouwers *et al.* (2014) ont aussi observé un corpus parallèle et implémenté manuellement 19 règles de simplification syntaxique dans leur système. Celui-ci permet de « générer un ensemble de phrases simplifiées candidates pour chaque phrase originale » et la « meilleure candidate est ensuite sélectionnée sur la base de critères de lisibilité intégrés au sein d'un algorithme d'optimisation » (Gala *et al.*, 2018, p. 128).

2.2.3 Simplification neuronale

Les techniques de simplification avec des règles ou de l'apprentissage automatique nécessitent des corpus parallèles, c'est-à-dire des textes ou des phrases dont le niveau de technicité ou de lisibilité est différent (Cardon, 2021). Tout comme pour l'évolution de la traduction automatique, l'apprentissage a d'abord été basé sur des méthodes statistiques et ensuite neuronales, qui effectuent en une seule fois la simplification lexicale et la simplification syntaxique. Le modèle apprend automatiquement les règles de simplification. Ces approches neuronales exploitent par exemple des modèles de séquence-à-séquence avec des couches à mémoire court terme et long terme en utilisant OpenNMT (Nisioi *et al.*, 2017) ou l'apprentissage par renforcement (Zhang et Lapata, 2017). La simplification non supervisée s'est ensuite développée en raison du peu de corpus parallèles disponibles. Nous pouvons mentionner la simplification multilingue à l'aide de réseaux de neurones avec des méthodes non-supervisées comme MUSS (Martin *et al.*, 2022). Nous verrons dans le chapitre 4 (cf. section 5.1), que de récentes études commencent à investiguer le potentiel des grands modèles de langue génératifs pour la simplification automatique de textes, car ils n'ont pas besoin de beaucoup de données (Feng *et al.*, 2023; Kew *et al.*, 2023).

2.2.4 Simplification médicale

Concernant la simplification de textes médicaux, Cardon (2021), Koptient *et al.* (2019) ont annoté et analysé un corpus français avec des documents comparables différenciés par leur technicité. Ils ont proposé une typologie des procédés de simplification. Elle est composée de 25 opérations de transformations possibles (Cardon, 2021, pp. 27-30). Par exemple, la substitution d'un mot par un synonyme, un hyperonyme, un hyponyme, le remplacement d'un verbe par un nom (ou une autre combinaison de POS), l'ajout d'une précision, la pronominalisation, etc. Cette typologie nous sera utile pour analyser comment traduire des phrases médicales vers les pictogrammes. Afin de simplifier automatiquement des textes médicaux français, Cardon (2021) a utilisé une

approche de traduction automatique neuronale, basée sur des corpus parallèles entraînés avec OpenNMT. Ses expériences ont notamment montré l'importance de disposer d'un grand volume de données (à la fois spécialisées et non spécialisées), afin d'obtenir de bonnes performances sur la langue médicale.

2.2.5 Ressources pour la simplification

Pour chaque langue, il existe donc diverses ressources linguistiques pour la simplification automatique de textes. La base lexico-sémantique WordNet en fait partie. [32] Elle a été traduite dans beaucoup de langues. Nous l'utilisons dans le chapitre 3. Pour le français, d'autres lexiques de synonymes ont aussi été conçus. FLELex pour les apprenants du français langue étrangère (François *et al.*, 2014). [33] ReSyf que nous exploitons dans le chapitre 4 (Billami *et al.*, 2018; François *et al.*, 2016), en plus d'un lexique de paraphrases médicales (Koptient et Grabar, 2020). Nous pouvons citer les quelques corpus parallèles français (des textes en version originale et en version simple) : Alector, qui contient 79 textes littéraires et scientifiques utilisés dans des écoles françaises par des enfants de 7 à 9 ans (Gala *et al.*, 2020) [34] ; CLEAR que nous utilisons dans le chapitre 5 pour la désambiguïsation lexicale et qui est un corpus médical français (Grabar et Cardon, 2018) ; quant à Brouwers *et al.* (2014), ils avaient travaillé sur la version francophone de l'encyclopédie en ligne Wikipédia et son équivalent Vikidia pour les enfants entre 8 et 13 ans, mais le corpus n'est pas disponible. [35] C'est pourquoi, Ormaechea Grijalba et Tsourakis (2023) ont proposé le corpus parallèle WiViCo (Wikipedia-Vikidia Corpus). [36]

2.2.6 Et les langues peu dotées ?

Des techniques ont été proposées pour la simplification des langues peu dotées et des langues de spécialité (incluant la langue médicale), qui ont peu de corpus parallèle. Nous pouvons citer la traduction automatique d'un corpus parallèle existant – par exemple, de l'anglais vers le français (Cardon, 2021) ou du néerlandais vers les pictogrammes avec Text-to-Picto pour concevoir Picto-to-Text (Sevens, 2018) –; la création de corpus synthétiques à partir d'autres corpus (Seidl et Vandeghinste, 2024), voire générés par des grammaires (Bouillon *et al.*, 2017; Gerlach *et al.*, 2024; Macaire *et al.*, 2024a,c). Nous n'avons pas choisi cette approche, étant donné qu'à notre connaissance, il n'existait aucun corpus authentique de phrases médicales en pictogrammes pour le contexte qui nous intéressait (dialogues entre médecins et patients avec une DI).

32. Des auteurs ont d'ailleurs proposé la construction d'un WordNet médical (Smith et Fellbaum, 2004), mais il n'a finalement jamais été créé.
33. https://cental.uclouvain.be/cefrlex/flelex
34. https://alectorsite.wordpress.com
35. https://fr.wikipedia.org ; https://fr.vikidia.org
36. https://github.com/lormaechea/wivico

2.2.7 Simplification pour les humains et les machines

Divers systèmes de simplification automatique de textes ont été conçus, afin de rendre le contenu accessible pour différents publics : les enfants (De Belder et Moens, 2010; Barlacchi et Tonelli, 2013), les dyslexiques (Rello *et al.*, 2013), les aphasiques (Devlin et Unthank, 2006), mais aussi les personnes avec une DI (Chen *et al.*, 2017), un TSA (Barbu *et al.*, 2013), etc. La simplification automatique peut non seulement être utilisée pour des humains, mais aussi comme tâche de prétraitement de texte dans beaucoup de systèmes en TAL. Il s'agissait par ailleurs du premier but de ces outils. Le premier système de simplification automatique a en effet été conçu pour faciliter l'analyse syntaxique (Chandrasekar *et al.*, 1996). Štajner et Popović (2016) ont également montré que la simplification de textes pouvait améliorer la traduction automatique de langues peu dotées, en particulier la grammaticalité (fluidité), ainsi que la préservation du sens (fidélité), tout en réduisant le nombre de post-éditions.

Le recours à des techniques de simplification automatique de textes pour une application de CAA avait déjà été suggéré par Higginbotham *et al.* (2012, p. 21). Les auteurs avaient constaté qu'aucune technologie de CAA commercialisée n'utilisait ces techniques, alors que selon eux, elles permettraient de concevoir des outils de communication adaptés à l'utilisateur, et ce, à partir de diverses sources d'information :

> “ Although not currently employed in any commercial AAC technologies, the potential for restructuring text obtained from the Internet, e-books, e-mail, transcribed partner speech, [...] is significant. Complex reading materials could be simplified for individuals with cognitive-linguistic challenges for palatable reading and/or listening experiences. Newspaper stories, magazine articles, web page content, e-mails, [...] could be processed and inserted into the individual's AAC system, transformed into materials for expressive communication. Provided with a simplified and appropriately segmented set of topic materials, the augmented speaker could select from these offerings to discuss current events, [...] or other topics related to their personal lives. ”

Vaschalde (2018, p. 62) avait aussi proposé comme piste d'amélioration pour son système speech-to-picto d'utiliser la ressource ReSyf pour la simplification lexicale, afin de traduire le vocabulaire complexe qui ne se trouve pas dans un ensemble de pictogrammes. Nous avons néanmoins vu dans la section 1.2 que des études sur des systèmes text-to-picto ont depuis lors exploré la piste de la simplification, par exemple : Bui *et al.* (2012), Sevens *et al.* (2017), Sevens (2018), Vaschalde *et al.* (2018) en compression de phrases, simplification syntaxique et/ou simplification lexicale avec des règles.

3 Conclusion : peu de systèmes médicaux vers les pictogrammes

Dans ce chapitre, il était impossible de proposer un état de l'art complet sur les disciplines en TAL qui font l'objet de cette thèse : la traduction automatique et la simplification automatique de textes. Les méthodes utilisées aujourd'hui pour effectuer ces deux tâches sont apparentées et nécessitent généralement beaucoup de données pour l'entraînement de modèles. Or, il n'existait pas de corpus de phrases en pictogrammes pour le français et le domaine médical quand nous avons commencé cette thèse. La traduction vers les pictogrammes peut être considérée comme une tâche de traduction vers une langue peu dotée et une tâche de traduction monolingue, mais aussi comme une tâche de simplification automatique de textes, les langues pictographiques étant des langues simplifiées. Dans les prochains chapitres, nous avons donc décidé d'adopter une approche de simplification basée sur des ressources lexico-sémantiques générales ou médicales et des règles. Nous avons aussi choisi de tester une approche de simplification automatique avec des grands modèles de langue génératifs plutôt que d'explorer la traduction automatique neuronale pour développer et améliorer le système Text-to-Picto. De plus, l'impact de la tâche de désambiguïsation lexicale sur la traduction automatique vers les pictogrammes sera également évalué.

Nous avons constaté qu'il y a peu de systèmes de traduction automatique de phrases vers les pictogrammes qui intègrent des techniques de TAL, et encore moins pour le français, la CAA et/ou la communication entre médecins et patients. Le seul système répondant à une partie de ces critères est PictoDr – qui a toutefois été conçu à partir d'un outil de traduction pour les patients allophones et pendant la durée de cette thèse. C'est pourquoi, nous l'utiliserons comme point de comparaison avec Text-to-Picto dans les études utilisateurs du dernier chapitre. En ce qui concerne les phrases que nous tenterons de traduire automatiquement vers les pictogrammes avec Text-to-Picto, nous emploierons le corpus de BabelDr. En effet, il s'agit d'un corpus français de phrases médicales (instructions et questions), validé par des médecins. Dans Text-to-Picto, nous n'utiliserons toutefois pas une approche de traduction basée sur des phrases fixes comme c'est le cas des premières versions du système BabelDr qui exploite ce corpus. Nous avons aussi vu qu'il existe beaucoup de méthodes utilisées pour évaluer des systèmes de traduction vers les pictogrammes, qu'elles soient automatiques, manuelles ou avec des utilisateurs cibles. Nous avons déjà cité plusieurs fois Text-to-Picto (Vandeghinste *et al.*, 2017; Sevens, 2018), le système de traduction automatique vers les pictogrammes que nous adapterons au français, aux pictogrammes Arasaac et que nous spécialiserons au langage médical. Il a l'avantage d'être libre de droit et de faire appel à des techniques de TAL bien documentées.

Chapitre 3

Système de base de Text-to-Picto

Introduction

Avant de débuter ce chapitre, nous présentons brièvement le système Text-to-Picto qui fait l'objet de cette thèse. Il s'agit de la version améliorée de l'outil intégré au sein de la plateforme d'échanges WAI-NOT, [1] qui est utilisée en Flandre – la partie néerlandophone de la Belgique – par des personnes ayant des difficultés de compréhension, dont celles avec une Déficience Intellectuelle (DI). À l'origine, ce système pour le néerlandais ne traduisait pas les mots aux formes fléchies et générait des erreurs de traduction dues à un problème d'homonymie, étant donné qu'il se limitait à une simple recherche dans une liste de pictogrammes sur la base de mots non lemmatisés (Sevens, 2018, pp. 42-43). Il a ensuite été amélioré par la KULeuven (Vandeghinste et Schuurman, 2014; Vandeghinste *et al.*, 2017), financé par le Fonds de soutien Marguerite-Marie Delacroix. L'*Agentshap Innoveren & Ondernemen* a financé le doctorat de Sevens (2018) sur la version néerlandaise de Text-to-Picto (Sevens *et al.*, 2016a,b, 2017), ainsi que Picto-to-Text (Sevens *et al.*, 2015a). Le projet Able to Include [2] soutenu par la Commission européenne a financé la version anglaise et espagnole de Text-to-Picto (Sevens *et al.*, 2015b). Ici, nous proposons d'adapter le système Text-to-Picto au français et de l'étendre à l'ensemble de pictogrammes Arasaac.

Dans ce chapitre, nous commençons par décrire l'architecture de notre système Text-to-Picto de base pour le français (section 1), mais également comment nous avons lié les pictogrammes à la ressource lexico-sémantique WordNet (section 2). Nous présentons les trois corpus utilisés pour cette première version de Text-to-Picto – chacun représentant un cas d'utilisation possible en Communication Alternative et Améliorée (CAA) – (section 3), avant de procéder au réglage du système (section 4). Les résultats de nos premières évaluations sont ensuite détaillés (section 5). Enfin, nous discutons des limitations de notre système de base (section 6), qui nécessite des améliorations.

1. https://www.wai-not.be
2. https://able-to-include.ccl.kuleuven.be

1 Architecture du système de base

Nous décrivons l'architecture du système de base de Text-to-Picto (Figure 3.1), utilisé pour traduire automatiquement une entrée textuelle en français vers une séquence de pictogrammes Sclera, Beta ou Arasaac (cf. chapitre 1, sections 2.2.5 et 2.2.6).

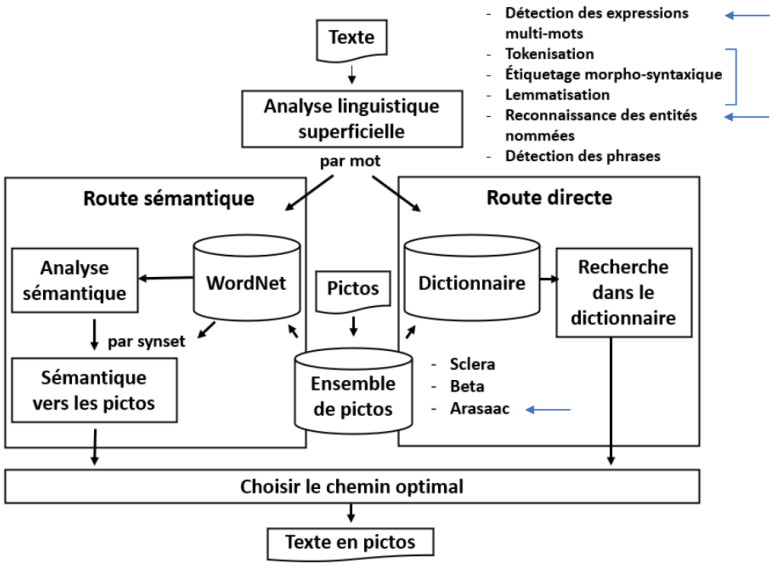


FIGURE 3.1 – Text-to-Picto pour le français (système de base) : architecture, adaptée de Vandeghinste *et al.* (2017, p. 222)

Le texte source subit d'abord une analyse linguistique superficielle : d'une part, une tokenisation, un étiquetage morphosyntaxique et une lemmatisation avec TreeTagger (Schmid, 1994) ;[3] d'autre part, nous avons ajouté une détection de certaines expressions multi-mots, un traitement pour des phénomènes spécifiques du français (par exemple, l'élision et la prise en compte de variantes pour la négation), ainsi qu'une simple reconnaissance des entités nommées, basés sur des règles et des dictionnaires. Comme dans Vaschalde (2018), nous substituons les entités nommées détectées par des tokens génériques tels que « perso » (pour les noms de personnes) et « ville ». Le système détecte aussi les phrases quand il y en a plusieurs (Sevens, 2018, p. 60).

Par exemple (Figure 3.2), la phrase « Max ira à Louvain l'été, au revoir » est traduite par sept pictogrammes (« perso », « aller », « à », « ville », « le », « été », « au_revoir »). Sans nos règles pour l'élision et la détection des expressions multi-mots, l'article « le » n'aurait pas été traduit et l'expression « au revoir » aurait été traduite par deux pictogrammes : la préposition « au » et le verbe « revoir » (le sens de « corriger »).

Dans l'étape suivante, deux routes sont possibles en fonction du mot à traduire : la route sémantique et la route directe. Dans la route sémantique, chaque mot est cherché dans la base de données WordNet (cf. section suivante). Dans le cas où aucun mot n'est trouvé, nous utilisons deux relations de WordNet pour obtenir des substituts de

3. Nous réutilisons du code Perl de la version espagnole de Text-to-Picto qui utilise TreeTagger.

Texte	Max	ira	à	Louvain	l'	été	,	au revoir
Détection des expressions multi-mots								au_revoir
Tokenisation	Max	ira	à	Louvain	l'	été		au_revoir
Étiquetage morpho-syntaxique	NAM	VER:futu	PRP	NAM	DET:ART	NOM	PUN	NOM
Lemmatisation	Max	aller	à	Louvain	le	été	,	<unknown>
Reconnaissance des entités nommées	(perso)			(ville)				
Route sémantique – WordNet		00341560-v 00359916-v ...				15237250-n		
Route directe – Dictionnaire	(perso)		à	(ville)	le			au_revoir
Chemin optimal et Picto ID	34560	2432	7041	5963	8476	5604		5896
Texte en pictogrammes								

FIGURE 3.2 – Text-to-Picto pour le français (système de base) : exemple de la phrase « Max ira à Louvain l'été, au revoir » (Arasaac)

traduction : les hyperonymes (*has_hyperonym*) et les antonymes (*near_antonym*). Par exemple, quand nous avons commencé cette version de Text-to-Picto, il n'existait pas encore de pictogramme pour « saumon », nous traduisions le mot par son hyperonyme « poisson ». Le mot « infecter » n'avait pas de pictogramme, il était donc traduit par son antonyme (« désinfecter »), suivi du pictogramme représentant la négation (« pas »).

Dans leurs annotations pour Sclera et Beta, Vandeghinste et Schuurman (2014) avaient indiqué si le pictogramme est complexe ou pas, c'est-à-dire s'il représente plusieurs concepts (verbe + nom, nom + nom, nom + adjectif, etc.). Par exemple, « manger un sandwich » est traduit par le pictogramme « *boterham-eten* » en Sclera. Ce nom de fichier a été lié au *head synset* {manger, alimenter, déjeuner} et au *dependent synset* {sandwich}. L'information concernant les pictogrammes correspondant à une expression multi-mots est manquante pour Arasaac. Il convient de noter que la flexion est prise en compte pour les expressions multi-mots annotées avec deux synsets, contrairement aux expressions multi-mots détectées dans l'analyse linguistique superficielle.

Pour la route directe, nous avons construit un dictionnaire pour chacun de nos trois ensembles de pictogrammes pour les mots non couverts par WordNet, c'est-à-dire les pronoms, les prépositions, etc. En Sclera et Beta, les noms de fichier des pictogrammes néerlandais ont été liés à des lemmes ou des tokens par Vandeghinste *et al.* (2017). Nous avons traduit manuellement ces deux dictionnaires néerlandais en français. Pour Arasaac, les lemmes et les tokens français ont été liés aux identifiants numériques des pictogrammes. Nous avons ajouté un tag *Part Of Speech (POS)* pour distinguer certains homonymes, par exemple l'adverbe de négation « pas » et le nom « pas ». Dans cette première version, nos dictionnaires fournissent 412, 298 et 420 liens directs entre des pictogrammes (Sclera, Beta, Arasaac) et des lemmes ou tokens français.

Pour choisir le meilleur chemin qui permet de convertir une séquence de lemmes en une séquence de pictogrammes, nous utilisons l'algorithme de recherche A* (Hart *et al.*, 1968), implémenté dans Text-to-Picto par Vandeghinste *et al.* (2017, pp. 232-234). Il exploite différents paramètres, c'est-à-dire des poids ou pénalités liés aux relations de WordNet, aux caractéristiques des pictogrammes et à la préférence de la route.

2 WordNet : la route sémantique

Afin de traduire du texte en langue naturelle vers les pictogrammes, les systèmes de traduction comptent sur des ressources lexico-sémantiques. Le Princeton WordNet (PWN) est une des plus grandes bases de données pour l'anglais (Miller, 1995). Elle classifie les verbes, noms, adjectifs et adverbes en ensembles de synonymes cognitifs, appelés synsets, qui sont liés par des relations sémantiques. Les dernières versions sont le PWN 3.0 et le PWN 3.1, dont les synsets n'ont pas les mêmes identifiants numériques.

Pour le français, il existe le WOrdnet Libre du Français (WOLF) de Sagot et Fišer (2008) et le WordNet du Français (WoNeF) de Pradet *et al.* (2014), deux traductions automatiques open source du PWN 3.0 qui diffèrent dans la façon dont ils ont été construits. [4] Dans cette thèse, nous testons le WOLF 1.0b4 (2014) et les trois versions du WoNeF 0.1 (2012) : *coverage*, *f-score*, *precision*. Le WOLF est considéré comme le WordNet français standard, il est cité plus souvent que le WoNeF. Comparé aux 117 659 synsets du PWN 3.0, il ne contient que 56 475 synsets avec au moins un lemme traduit en français (Table 3.1). En ce qui concerne le WoNeF (Pradet *et al.*, 2014, p. 32), la version avec la plus haute couverture (*coverage*) contient 109 447 paires (literal, synset), le principal WoNeF (*f-score*) a un F-score de 70,9 % et la version avec la plus haute précision (*precision*) a une précision de 93,3 %. De plus, comme un résultat de l'optimisation des trois métriques, la version *coverage* inclut 55 697 synsets, la version *f-score* en a 53 440 et la *precision* en a seulement 15 482 (Pradet *et al.*, 2014).

	WOLF 1.0b4	**WoNeF 0.1**		
	–	**coverage**	**f-score**	**precision**
Noms (N)	**42 427**	37 685	37 335	10 920
	51,66 %	45,89 %	45,49 %	13,29 %
Verbes (V)	**5 870**	5 772	3 845	1 250
	42,63 %	41,92 %	27,92 %	9,07 %
Adjectifs (ADJ)	6 691	**10 238**	**10 238**	2 755
	36,85 %	56,38 %	56,38 %	15,38 %
Adverbes (ADV)	1 487	**2 002**	**2 002**	557
	41,06 %	55,28 %	55,28 %	15,38 %
Total	**56 475**	55 697	53 440	15 482

TABLE 3.1 – Nombre de synsets non vides des WordNets français et pourcentage comparé au PWN 3.0 par POS

Le système Text-to-Picto a été conçu pour être aussi indépendant que possible des langues et facilement extensible. Pour la version néerlandaise, Vandeghinste et Schuurman (2014) avaient manuellement lié 5 710 pictogrammes Sclera et 2 760 pic-

4. Le WOLF a été construit en utilisant des corpus parallèles et d'autres ressources (dont les données Wiki). Il a aussi été étendu avec diverses techniques. WoNeF est une amélioration de JAWS (Mouton et de Chalendar, 2010) – un autre WordNet français traduit à partir du PWN, qui contient que des noms. JAWS a été construit avec un algorithme faiblement supervisé, des dictionnaires bilingues ainsi qu'un modèle de langue syntaxique. Les versions de WoNeF ont été créées avec différents sélecteurs.

togrammes Beta au WordNet Cornetto (Vossen *et al.*, 2008), avant de les lier automatiquement au PWN 3.0 pour l'anglais et au WordNet MCR 3.0 pour l'espagnol (Sevens *et al.*, 2015a). Pour étendre le système Text-to-Picto au français et à l'ensemble de pictogrammes Arasaac, nous n'avions pas accès aux liens entre le PWN 3.0 et 800 pictogrammes Arasaac de Schwab *et al.* (2020). Par conséquent, nous avons utilisé la récente *Application Programming Interface (API)* d'Arasaac [5] pour obtenir les données JSON (cf. infra), incluant les liens manuels vers les synsets du PWN 3.1 (*synsets*), les noms de pictogrammes (*keywords*, c'est-à-dire les lemmes sans l'extension .png) et leur identifiant numérique (*_id*) permettant d'accéder à l'URL des pictogrammes sur un serveur d'Arasaac. Les noms des pictogrammes sont disponibles pour au moins 30 langues, incluant l'anglais, l'espagnol, le néerlandais, le russe, l'arabe, etc. Les autres données de l'API d'Arasaac sont les mêmes pour toutes les langues.


```
{"schematic":false,"sex":false,"violence":false,"aac":false,"aacColor":false,"
    skin":true,"hair":true,"downloads":0,"categories":["health personnel","
    sanitary professional","covid-19"],"synsets":["10040615-n"],"tags":["health
    ","medicine","professional","work","covid-19"],"_id":2467,"created
    ":"2007-12-14T12:39:54.000Z","lastUpdated":"2021-03-13T22:05:27.267Z","
    keywords":[{"type":2,"keyword":"médecin","hasLocution":true,"plural":"
    médecins"},{"type":2,"keyword":"docteur","hasLocution":true,"plural":"
    docteurs"}]}
```


En utilisant OpenRefine [6] (Verborgh et De Wilde, 2013), nous avons nettoyé les données : tri, suppression, renommage des doublons dans les noms de fichier (que nous avons renommés en ajoutant des nombres pour distinguer les pictogrammes différents dans nos corpus de référence), etc. Nous avons lié automatiquement les pictogrammes Arasaac associés à un ou plusieurs synset(s) du PWN 3.1 avec le WOLF et le WoNeF en passant par les identifiants du PWN 3.0 et les index collaboratifs interlingues (*Collaborative Interlingual Indexes* ou CILI), disponibles sur le dépôt GitHub de l'Association Global WordNet. [7] Cela devrait aussi être possible d'utiliser le récent Open English WordNet [8] (McCrae *et al.*, 2020) et de les lier aux pictogrammes Arasaac. Ce WordNet et le PWN 3.1 ont les mêmes identifiants de synsets.

Par exemple (Figure 3.3), le pictogramme Arasaac « docteur » ou « médecin » a l'identifiant 2467. Il est associé au synset nominal « 10040615-n » du PWN 3.1 et à d'autres informations (dont plusieurs tags, catégories, etc.), que nous avons séparées dans d'autres tables. En transférant automatiquement le synset anglais avec le CILI, nous obtenons ainsi le synset du PWN 3.0 « 10020890-n », le POS, la/les relation(s) et le(s) lemme(s) des WordNets français. Dans ce cas-ci, {docteur, médecin, toubib} pour

5. https://arasaac.org/developers/api
6. https://openrefine.org
7. https://github.com/globalwordnet/cili
8. https://github.com/globalwordnet/english-wordnet ; https://en-word.net

le WOLF et seulement {médecin} pour le WoNeF (*coverage* et *f-score*). Les lemmes {docteur} et {toubib} sont liés à un autre synset. Il est d'ailleurs surprenant de constater que dans le WoNeF *f-score* il n'existe pas de lemme {docteur} et {médecin} (termes très fréquents), mais {toubib} (terme très rare), qui est d'ailleurs lié à deux autres synsets.

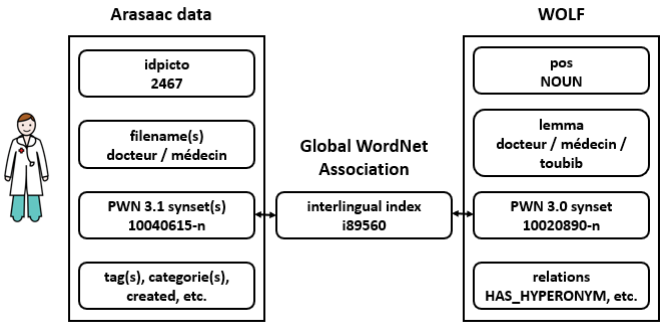


FIGURE 3.3 – Lier les pictogrammes Arasaac aux WordNets français

Comme un résultat de ce processus, notre système Text-to-Picto est le premier qui utilise les synsets des WordNets français. En effet, Schwab *et al.* (2020) avaient directement utilisé les lemmes du PWN 3.0 et traduit automatiquement le texte en français parce qu'ils considèrent que le PWN original est une base de données plus complète et plus fiable. Comme mentionné avant, les traductions automatiques du WOLF et des versions du WoNeF diffèrent et sont moins complètes que celles du PWN 3.0.

3 Corpus : des cas d'utilisation

Avant de présenter les résultats du système de base, nous décrivons nos trois corpus représentant plusieurs cas d'utilisation de CAA (Norré *et al.*, 2021c,d) : le Corpus Email (section 3.1), le Corpus Livre (section 3.2) et le Corpus Médical (section 3.3). Ils seront utilisés pour le paramétrage de notre système de base (cf. section 4), son évaluation automatique et son évaluation manuelle (section 5), dans lesquelles nous testerons différentes métriques pour comparer la sortie de Text-to-Picto à une traduction de référence. Étant donné que notre objectif consiste à adapter un système de traduction vers les pictogrammes pour la communication médicale, le Corpus Email et le Corpus Livre ne seront plus utilisés dans les prochains chapitres pour l'évaluation. Nous les décrivons ici car nous souhaitons situer les performances de notre système par rapport à d'autres études pour le néerlandais, l'anglais, l'espagnol ou le français.

Pour chacun des corpus décrits dans cette section, nous détaillons l'entrée textuelle du système Text-to-Picto (les phrases à traduire, c'est-à-dire le corpus source) et les traductions manuelles en pictogrammes (le corpus de référence). Concernant les traductions manuelles, nous nous sommes basé sur des corpus déjà utilisés pour évaluer des systèmes de traduction automatique vers les pictogrammes (Sevens, 2018;

Vaschalde, 2018). Nous avons considéré que tous les mots des corpus sources devaient être traduits en – noms de – pictogrammes (s'ils existent) dans nos corpus de référence. La traduction d'une phrase en pictogrammes s'apparente souvent à une séquence de lemmes. Il ne s'agit pas d'une traduction littérale. Nous avons parfois traduit plusieurs mots par un seul pictogramme : par exemple, pour les expressions multi-mots telles que « bouteille de coca » (nom + préposition + nom) ou « envoyer une lettre » (verbe + article + nom), qui peuvent être traduites par les pictogrammes « Coca-Cola » ou « envoyer_2 » en Arasaac (Figure 3.4). Pour rappel (cf. chapitre 1, section 2.3), certains pictogrammes Arasaac peuvent avoir différents noms de fichier en raison de sens différents, c'est le cas du pictogramme « mal de tête » ou « faire mal » (Figure 3.4). Nous donnons un aperçu des trois corpus (source et référence en Arasaac) dans l'annexe A.1.

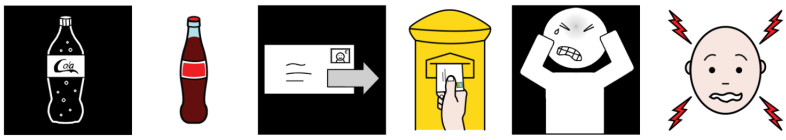


FIGURE 3.4 – Exemple de pictogrammes pour « bouteille de coca » (Sclera/Arasaac), « envoyer une lettre » (Sclera/Arasaac) et « mal de tête » (Sclera/Arasaac) ou « faire mal » (Arasaac)

3.1 Corpus Email

Le Corpus Email (140 phrases) a été traduit manuellement en pictogrammes Beta et Sclera par Sevens (2018). Les emails, écrits par des personnes avec une DI, leurs enseignants ou leurs parents, ont été extraits à partir du site belge WAI-NOT. Nous avons manuellement traduit ce corpus néerlandais en français (corpus source) et en pictogrammes Arasaac (corpus de référence). Nous avons un peu pré-édité les traductions de référence Beta et Sclera afin de garder l'ordre des mots français de notre corpus source. Nous n'avons pas reproduit les fautes d'orthographe des personnes avec une DI, car nous n'évaluons pas la correction orthographique automatique.

3.2 Corpus Livre

Le Corpus Livre (253 phrases) est composé de six histoires pour enfants open source, [9] qui ont été manuellement traduites en pictogrammes Arasaac par Vaschalde (2018). Nous avons un peu pré-édité ce corpus afin de le rendre conforme à notre format d'évaluation : par exemple, nous n'avons pas traduit les noms au pluriel deux fois comme c'est le cas dans Vaschalde (2018). Par exemple, « les couvertures » dans son corpus source a été traduit manuellement par « le couverture couverture » dans son

9. https://odysseuslibre.be/site/category/livres-pour-enfants/abel-et-bellina

corpus de référence (Figure 3.5). Dans notre corpus de référence, nous l'avons remplacé par « les couverture » sans la répétition du nom et en gardant le pluriel pour l'article car nous avons ajouté le pictogramme « les » dans notre dictionnaire. Il n'existe pas de pictogramme représentant plusieurs « couvertures ». Nous avons aussi modifié certains noms de fichier dans leur traduction de référence; nous avons renommé les doublons quand nous avons prétraité les données Arasaac.

FIGURE 3.5 – Exemple de deux traitements du pluriel pour « les couvertures » (Arasaac)

3.3 Corpus Médical

Le Corpus Médical (260 phrases) est un sous-ensemble de BabelDr (Bouillon *et al.*, 2021), un système de traduction médicale. Il s'agit des phrases canoniques, les phrases types (cf. chapitre 2, section 1.4.1). Elles sont donc relativement explicites – sans ellipse et sans faute d'orthographe –, si nous les comparons à certaines variantes proposées par le système BabelDr. Ce sont principalement des questions (fermées ou ouvertes) de médecins aux patients, comme « pouvez-vous décrire la douleur? » ou « à combien était votre température la dernière fois que vous l'avez mesurée? ». Il y a aussi des instructions destinées au patient, telles que « je vais m'occuper de vous aujourd'hui ». Comme pour le Corpus Email, nous avons traduit manuellement le Corpus Médical en pictogrammes Arasaac. Nous donnons un exemple de traduction de référence pour la phrase source « avez-vous des caries? » (Figure 3.6) : la suite de cinq noms de fichier « avoir », « vous », « des », « carie », «? ». Dans le chapitre 4, nous verrons qu'il existe différentes stratégies pour améliorer la traduction et représenter un pictogramme manquant dans un ensemble. C'est pourquoi, nous adapterons le Corpus Médical de référence après avoir évalué le système de base.

FIGURE 3.6 – Exemple de phrase du Corpus Médical : « avez-vous des caries? » (Arasaac)

4 Paramétrage du système de base

Nous présentons comment nous avons réglé le système de base pour les évaluations, c'est-à-dire : le choix du WordNet français (section 4.1) et des différents paramètres de Text-to-Picto (section 4.2).

4.1 Quel WordNet français ?

Pour le réglage de notre système de base, nous avons utilisé un corpus de développement constitué de 56 phrases du Corpus Email. À ce stade, il pouvait arriver qu'en faisant fonctionner plusieurs fois le système Text-to-Picto sur la même phrase on obtienne des traductions différentes avec les mêmes paramètres. [10] Chaque score BLEU est donc une moyenne calculée sur la base de 10 traductions. Nous donnerons les écart-types sur 10 lancements entre parenthèses dans la table.

Nous avons d'abord testé les WordNets français : le WOLF et les trois versions du WoNeF – *coverage*, *f-score* et *precision*. Comme nous ne connaissions pas encore les meilleurs paramètres pour notre version de Text-to-Picto, nous avons utilisé les meilleurs de Sevens (2018, p. 64) pour les ensembles Sclera et Beta. Pour Arasaac, nous avons pris les paramètres de Beta. Les scores BLEU (Papineni *et al.*, 2002), [11] présentés dans la Table 3.2, nous ont permis de choisir le WOLF comme le meilleur WordNet français, étant donné qu'il obtient les plus hauts scores BLEU quel que soit l'ensemble de pictogrammes. Par conséquent, le WOLF sera utilisé pour toutes les évaluations.

	Sclera	**Beta**	**Arasaac**
WOLF	**23,93** (0,36)	**25,89** (1,42)	**29,48** (1,49)
WoNeF coverage	13,71 (0,10)	17,55 (0,26)	18,41 (1,30)
WoNeF f-score	13,77 (0,09)	17,69 (0,19)	18,52 (1,22)
WoNeF precision	12,48 (0,00)	12,29 (0,00)	06,52 (0,00)

TABLE 3.2 – Résultats du Text-to-Picto français (système de base) pour WOLF et WoNeF par ensemble de pictogrammes avec la métrique BLEU

Ces résultats peuvent être expliqués par le fait que le WOLF est souvent connecté à plus de synsets que le WoNeF ; il est donc plus probable qu'il existe un lien vers un pictogramme correct. Pour WOLF, l'écart-type est toujours plus haut que pour le WoNeF. L'écart-type vaut d'ailleurs 0 pour la version *precision*. Sa petite taille réduit la probabilité qu'un pictogramme soit lié à plusieurs lemmes.

4.2 Quels paramètres ?

Ensuite, nous avons réglé les paramètres (cf. section 1) du système de base avec une procédure automatique, un algorithme hill-climbing (Vandeghinste *et al.*, 2017) avec la métrique d'évaluation BLEU sur les 56 phrases du Corpus Email (cf. supra). Pour chaque ensemble de pictogrammes – Sclera, Beta et Arasaac – nous avons lancé le

10. Ce problème survient quand plusieurs pictogrammes obtiennent le même poids après l'algorithme A*. Il est lié à une différence de traitement des tables de hachage dans notre version de Perl et sera résolu dans le prochain chapitre avec l'ajout du mode AZ.

11. Le BLEU standard basé sur une moyenne d'unigrammes, bigrammes, trigrammes et quadrigrammes (mteval-v011b.pl sur https://github.com/moses-smt/mosesdecoder).

système 5 fois en effectuant à chaque fois 50 itérations avec différentes initialisations aléatoires des paramètres (tout en utilisant un pas de 1), afin de couvrir différentes zones de l'espace de recherche (le minimum et le maximum servant à le réduire). Nous avons pris les valeurs des paramètres du meilleur résultat (Table 3.3). [12]

	Min	Max	Pas	Sclera	Beta	Arasaac
WOLF relations						
Threshold	5	20	1	11	12	9
Hyperonym penalty	0	15	1	13	9	15
Antonym penalty	0	15	1	6	6	9
Pictograph features						
Wrong number	0	10	1	10	8	–
No number	0	10	1	2	3	–
Route preference						
Out-Of-Vocabulary	0	10	1	5	2	2
Direct route advantage	0	15	1	12	5	12

TABLE 3.3 – Valeurs des paramètres du Text-to-Picto français (système de base) sur le Corpus Email par ensemble de pictogrammes après réglage

5 Évaluation du système de base

Tout comme pour les premières versions – néerlandaise, anglaise et espagnole – de Text-to-Picto (Sevens, 2018), nous avons évalué le système de base français automatiquement (section 5.1) et manuellement (section 5.2).

5.1 Évaluation automatique

Nous avons évalué automatiquement les performances de notre système Text-to-Picto français sur le Corpus Email, le Corpus Livre et le Corpus Médical. Différentes conditions expérimentales ont été testées, en activant progressivement plus de fonctionnalités du système : **a)** seulement avec le dictionnaire ; **b)** avec le dictionnaire et les synonymes du WOLF ; **c)** avec le dictionnaire, les synonymes et les relations du WOLF (hyperonymes et antonymes). En plus, nous avons comparé nos résultats avec ceux du système Text-to-Picto néerlandais (Sevens, 2018) et ceux du système français de Vaschalde *et al.* (2018), quand ils étaient disponibles. De telles comparaisons doivent être prises avec précaution car les expériences ne sont pas strictement comparables. [13]

12. « *The first set of parameters ("hyperonym penalty", "xPos penalty", and "antonym penalty") concerns the maximum distance ("threshold" parameter) allowed between the original text and the pictograph message in terms of synset relations. The second set of parameters is related to the numeric features of the pictographs ("no number" and "wrong number"), as some pictographs make a distinction between singular or plural concepts* » (Sevens, 2018, p. 64).

13. À noter que Sevens (2018) utilisait également la relation *xpos_near_synonym* de WordNet (c), que nous n'avons pas incluse dans l'évaluation du système de base, mais qui le sera dans le chapitre 4.

Pour une meilleure comparaison avec ces études, nous avons réutilisé les métriques employées par Sevens (2018), Vandeghinste *et al.* (2017) : BLEU, NIST (Doddington, 2002), WER et PER (cf. chapitre 2, section 1.3.1). C'est pourquoi, nous avions traduit manuellement les phrases en – noms de – pictogrammes. Comme nous l'avons décrit dans la section 4.1, chaque métrique a été calculée sur la base d'une moyenne de 10 traductions générées avec Text-to-Picto. Les écart-types sont toujours présentés entre parenthèses dans les tables. [14] Nous ne post-éditons pas les sorties du système.

Nous avons comparé nos résultats avec ceux de Sevens (2018) pour le néerlandais sur une traduction des mêmes 84 phrases du Corpus Email (Table 3.4). [15] En ce qui concerne les scores BLEU, nous obtenons des résultats très comparables, en particulier pour Sclera. Pour Beta, nos scores sont plus bas que ceux de Sevens (2018). Arasaac obtient les meilleurs scores BLEU, excepté pour la première condition (a). Cela peut s'expliquer par le fait que les dictionnaires pour Sclera et Beta sont plus adaptés pour cette tâche. Ils ont en effet été conçus à partir des fréquents mots non traduits issus d'un plus grand ensemble d'emails néerlandais du Corpus Email. Pour les conditions (b) et (c), les différences ne semblent pas importantes.

	BLEU ↑	**NIST ↑**	**WER ↓**	**PER ↓**
Sclera				
a) Dictionnaire	12,05 (0,00)	2,02 (0,00)	58,79 (0,00)	55,86 (0,00)
(Sevens, 2018)	14,17	3,68	71,96	65,88
b) + Synonymes	17,87 (0,40)	**4,00** (0,05)	**56,21** (0,45)	**50,34** (0,44)
(Sevens, 2018)	16,55	3,97	67,54	60,50
c) + Relations	**17,90** (0,40)	**4,00** (0,04)	56,24 (0,31)	50,85 (0,32)
(Sevens, 2018)	16,12	3,96	68,78	61,33
Beta				
a) Dictionnaire	10,97 (0,00)	1,00 (0,00)	63,02 (0,00)	62,11 (0,00)
(Sevens, 2018)	16,98	4,43	63,44	53,77
b) + Synonymes	21,69 (1,49)	4,22 (0,13)	**57,56** (1,17)	**52,15** (1,16)
(Sevens, 2018)	23,01	5,00	52,42	43,31
c) + Relations	**22,40** (1,26)	**4,29** (0,06)	57,93 (0,56)	52,61 (0,64)
(Sevens, 2018)	25,91	5,17	51,29	42,07
Arasaac				
a) Dictionnaire	7,38 (0,00)	0,41 (0,00)	**59,34** (0,00)	58,85 (0,00)
(Sevens, 2018)	–	–	–	–
b) + Synonymes	24,88 (0,79)	4,41 (0,10)	68,14 (1,52)	57,65 (1,45)
(Sevens, 2018)	–	–	–	–
c) + Relations	**24,97** (0,88)	**4,44** (0,09)	68,00 (1,97)	**57,29** (1,77)
(Sevens, 2018)	–	–	–	–

TABLE 3.4 – Résultats du Text-to-Picto français et du Text-to-Picto néerlandais (systèmes de base) sur le Corpus Email par ensemble de pictogrammes avec les métriques BLEU, NIST, WER et PER

14. Pour la première condition (a), l'écart-type est toujours de 0 car il n'y a pas de variation dans les résultats sans WordNet.

15. Nous avons repris les résultats de Sevens (2018) obtenus après sa correction orthographique manuelle de l'entrée de Text-to-Picto.

Pour le Corpus Livre, comme les noms propres sont fréquents, nous avons ajouté une condition : **d)** sans l'étape de reconnaissance des entités nommées, basée sur le travail de Vaschalde (2018). Nos scores BLEU (Table 3.5) pour les conditions (b) (28,0) et (c) (28,3) sont comparables à ceux de Vaschalde *et al.* (2018). Leur système de traduction du français vers les pictogrammes Arasaac obtient un score BLEU de 25,45 quand tous les mots sont traduits, sans la désambiguïsation lexicale et sans leur traitement spécifique pour le pluriel (26,65 avec). Les scores plutôt corrects de la condition dictionnaire (a) – par rapport aux deux autres corpus – peuvent probablement s'expliquer par l'effet du module de reconnaissance d'entités nommées. En effet, le corpus de test comprend 152 occurrences de noms propres et nous observons que, sans ce module (d), les scores sont moins bons, y compris pour le NIST, le WER et le PER.

	BLEU ↑	**NIST ↑**	**WER ↓**	**PER ↓**
Arasaac				
a) Dictionnaire	18,00 (0,00)	1,93 (0,00)	**49,82** (0,00)	**48,57** (0,00)
b) + Synonymes	28,07 (0,41)	5,26 (0,06)	58,04 (0,90)	49,82 (0,82)
c) + Relations	**28,34** (0,66)	**5,28** (0,06)	57,71 (0,77)	49,37 (0,67)
d) - Entités nommées	17,45 (0,70)	4,79 (0,07)	68,18 (1,12)	57,19 (0,85)
(Vaschalde, 2018)	25,45	–	–	–

TABLE 3.5 – Résultats du Text-to-Picto français (système de base) sur le Corpus Livre pour l'ensemble de pictogrammes Arasaac avec les métriques BLEU, NIST, WER et PER

Contrairement au Corpus Livre, il n'y a pas d'entités nommées dans notre Corpus Médical qui sont intégrées au dictionnaire Arasaac (a). Les scores BLEU (Table 3.6) sont plus hauts que pour les autres corpus quand nous utilisons les synonymes (b) et les relations du WOLF (c). Les résultats des métriques WER et PER sont aussi meilleurs que ceux obtenus avec les autres corpus. Pour PER, cela peut s'expliquer par une plus haute fréquence de structures syntaxiques (cf. chapitre 4, section 1) facilement traduisibles (par exemple : « avez-vous... », « êtes-vous... », « pouvez-vous... », etc.).

	BLEU ↑	**NIST ↑**	**WER ↓**	**PER ↓**
Arasaac				
a) Dictionnaire	08,66 (0,00)	1,02 (0,00)	52,80 (0,00)	52,02 (0,00)
b) + Synonymes	31,19 (0,77)	5,24 (0,07)	51,41 (0,89)	46,39 (0,57)
c) + Relations	**31,36** (0,78)	**5,26** (0,09)	**51,13** (1,20)	**46,12** (1,23)

TABLE 3.6 – Résultats du Text-to-Picto français (système de base) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac avec les métriques BLEU, NIST, WER et PER

En ce qui concerne l'efficacité des différents ensembles de pictogrammes, testés sur le Corpus Email, nous voyons que nos résultats sont meilleurs pour l'ensemble Beta que pour Sclera. Sevens (2018) a expliqué que Beta contenait moins de pictogrammes que Sclera. Par conséquent, il y a plus de paraphrases et donc de traductions possibles

en Sclera, ce qui se traduit par une mesure moins précise de la qualité de la traduction avec BLEU ou NIST. Cependant, nos scores pour Arasaac – le plus grand ensemble de pictogrammes – sont les meilleurs. Comparé aux autres, cet ensemble inclut plus de mots de fonction non ambigus (prépositions, articles, etc.) qui ont été encodés dans notre dictionnaire. Ils seront donc toujours bien traduits, ce qui améliore les résultats.

Nous pouvons également mentionner que, malgré de relativement bons résultats, certains pictogrammes générés par notre système ne semblent pas toujours facilement compréhensibles selon le contexte de la phrase. C'est particulièrement le cas pour les mots de fonction et la description de la douleur sur une partie spécifique du corps humain (cf. chapitre 1, section 2.3). En plus, les mots médicaux ne sont pas toujours traduits parce qu'il n'existe pas de pictogrammes correspondants dans l'ensemble, par exemple pour les noms « trouble », « cancer », etc.

Nos expériences visaient aussi à évaluer la contribution de différents composants de notre système Text-to-Picto pour le français. Utiliser seulement le dictionnaire (a) donne clairement des résultats insatisfaisants. Le système s'améliore lorsque nous ajoutons les synonymes (b) ou les relations de WordNet (c), quels que soient les ensembles de pictogrammes et les corpus. Cependant, la différence entre ces deux composantes apparaît marginale. Dans nos corpus de référence, une inspection manuelle des données révèle que la relation de synonymie est beaucoup plus fréquente que celles d'hyperonymie et d'antonymie – cette dernière est très rare, surtout pour le plus grand ensemble de pictogrammes, Arasaac.

Finalement, comme expliqué précédemment, chaque fois que nous faisons fonctionner le système français avec les mêmes paramètres, nous pouvons obtenir des résultats sensiblement différents. Cela se produit quand le système doit calculer le meilleur chemin et qu'il doit choisir au hasard entre plusieurs pictogrammes qui ont le même poids, en utilisant WordNet. Nous explorerons donc s'il est possible d'associer des informations de fréquence et des poids supplémentaires.

5.2 Évaluation manuelle

Nous avons aussi réalisé une évaluation manuelle d'une traduction automatique des 260 phrases du Corpus Médical en pictogrammes Arasaac générées par notre système réglé, autrement dit, avec le WOLF et les autres paramètres (cf. section 4). Pour chacun des mots traduits, un juge (l'auteur) a vérifié si le pictogramme généré était une représentation sémantique cohérente du mot – sur la base des images pour différencier les mots polysémiques –, afin de calculer la précision. Il a enlevé les mots non traduits, pour calculer le rappel. Le F-score est la moyenne harmonique entre la précision et le rappel. Le système a obtenu une précision de 83,7 %, un rappel de 90,14 % et un F-score de 86,8 % (Table 3.7).

	Précision ↑	**Rappel ↑**	**F-score ↑**
Sclera – (Sevens, 2018)			
Email – Néerlandais	89,24 %	86,23 %	87,71 %
Email – Anglais	93,30 %	73,04 %	81,94 %
Email – Espagnol	93,31 %	83,14 %	87,93 %
Beta – (Sevens, 2018)			
Email – Néerlandais	85,91 %	89,45 %	87,64 %
Email – Anglais	82,56 %	86,14 %	84,31 %
Email – Espagnol	**94,64** %	86,83 %	**90,57** %
Arasaac			
Médical – Français	83,70 %	**90,14** %	86,80 %

TABLE 3.7 – Résultats du Text-to-Picto français sur le Corpus Médical et du Text-to-Picto néerlandais/anglais/espagnol (systèmes de base) sur le Corpus Email par ensemble de pictogrammes avec les métriques précision, rappel et F-score

En comparaison, Sevens (2018) a obtenu sur son Corpus Email un F-score entre 81-85 % pour le système de l'anglais vers Sclera et Beta ; et entre 87-91 % pour les systèmes du néerlandais ou de l'espagnol vers Sclera et Beta en ne tenant pas compte des noms propres. Nous constatons que notre système français obtient bien un rappel plus haut que la précision, mais aussi plus haut que le rappel des autres versions linguistiques. La précision doit cependant être encore améliorée. Dans cette évaluation manuelle, nous sommes partis de l'hypothèse que tous les mots – y compris les articles – devaient être traduits contrairement à Sevens (2018).

6 Conclusion : des limites au système de base

Dans ce chapitre, nous avons répondu aux sous-objectifs 1 et 2 de l'objectif 1 (cf. Introduction), en adaptant Text-to-Picto et en proposant un système de base pour le français vers les ensembles de pictogrammes Sclera, Beta et Arasaac (Norré *et al.*, 2021c,d). Nous avons également testé différents WordNets. Le WoNeF ayant obtenu de moins bons résultats, nous utiliserons toujours le WOLF. Les relations d'hyperonymie et d'antonymie n'améliorent pas les résultats, c'était déjà le cas de Sevens (2018), Vandeghinste *et al.* (2017). En utilisant WordNet, il est arrivé que notre version de Text-to-Picto génère des traductions Sclera, Beta et Arasaac légèrement différentes avec des paramètres identiques. Des changements devront être apportés pour résoudre ce problème. Afin de comparer nos résultats à ceux de Sevens (2018), nous avons réglé le système sur une traduction de son Corpus Email et gardé les paramètres obtenus lorsque nous avons évalué Text-to-Picto sur les autres corpus. Étant donné que notre thèse porte sur le domaine médical, nous réglerons à nouveau les paramètres du système avant de faire les prochaines évaluations. Comme nous proposerons des modifications à Text-to-Picto, nous adapterons notre Corpus Médical en Arasaac en fonction de la nouvelle sortie du système. Nous en profiterons aussi pour étudier la traduction

manuelle en pictogrammes avec l'aide d'un expert du domaine de la CAA.

En ce qui concerne les évaluations réalisées sur différents corpus, nous avions déjà mentionné que l'ensemble Arasaac couvre beaucoup de domaines différents et qu'il est aussi utilisé dans les hôpitaux (Paolieri et Marful, 2018). En novembre 2021, Arasaac comprenait 1 126 pictogrammes médicaux (classés dans *Health sciences*), regroupés dans 45 (sous-)catégories (dont les procédures médicales, les symptômes, les médicaments, etc.) – sans compter les 175 pictogrammes liés au COVID-19 et les 617 relatifs au corps humain. [16] Nous avons choisi de limiter les prochaines évaluations au Text-to-Picto français vers Arasaac, étant donné que pour évaluer le système en Sclera ou Beta nous avons aussi besoin d'un corpus de référence avec les noms de fichier néerlandais. [17] Comme Text-to-Picto est adaptable et exploite une ressource générale, nous pensons qu'il serait possible de le spécialiser en testant d'autres ressources et en explorant des façons de modifier la phrase médicale d'entrée pour la rendre plus simple à traduire en pictogrammes dans notre système, comme Sevens (2018).

Nous avons constaté que la version française de Text-to-Picto obtenait le meilleur rappel et une précision moins bonne par rapport à l'étude de Sevens (2018) – excepté pour la version anglaise avec les pictogrammes Beta. Étant donné qu'elle avait seulement pris en compte les mots de contenu, nos résultats sont donc un peu moins bons vu que la présence de mots de fonction – comme les articles et les prépositions dans nos dictionnaires de pictogrammes – améliore le nombre de mots traduits corrects. Il semble donc y avoir des améliorations à faire au niveau de la précision. Dans la suite de la thèse, nous proposerons d'étudier l'impact des mots de fonction en évaluant uniquement la traduction des mots de contenu. Nous pouvons aussi mentionner que cette première évaluation manuelle a été réalisée par un seul annotateur. Bien que cela n'ait pas d'impact sur le rappel – qui est évalué automatiquement –, l'évaluation de la précision est subjective. C'est pourquoi, dans le prochain chapitre, nous proposerons une nouvelle évaluation manuelle avec un juge supplémentaire en testant d'autres critères.

Notre Corpus Médical a été intégré au sous-corpus propicto-eval qui compose le corpus monolingue Propicto (Macaire *et al.*, 2023a, 2024a,c), sous une licence non commerciale sur la plateforme ORTOLANG. [18] Le corpus Propicto est multimodal et contient des données alignées parole/texte/pictogrammes pour le français. Il a donc été créé dans le cadre du projet PROPICTO (Ormaechea Grijalba *et al.*, 2023a,b).

16. Le rappel du système ne dépend toutefois pas du nombre de pictogrammes médicaux disponibles dans un ensemble, mais bien du nombre de pictogrammes médicaux associés à au moins un synset. Vandeghinste et Schuurman (2014) avaient lié au total (tous domaines confondus) 5 560 pictogrammes – différents – Sclera et 2 545 Beta alors que 8 047 l'étaient en Arasaac (et 11 433 dans la deuxième version de Text-to-Picto décrite dans le prochain chapitre).

17. À noter que nous avons aussi réalisé une version de Text-to-Picto néerlandais vers Arasaac avec la même procédure de traitement des données de l'API décrite dans ce chapitre. Bien que nous ne l'ayons pas évaluée, ceci confirme le fait que le système est facilement extensible à d'autres langues.

18. https://hdl.handle.net/11403/propicto

Chapitre 4

Spécialisation de Text-to-Picto

Introduction

Dans ce chapitre, nous poursuivons le développement du système de traduction automatique du français vers les pictogrammes Text-to-Picto, et, en particulier, son adaptation pour la communication entre médecins et patients avec une Déficience Intellectuelle (DI). Ce travail passe par plusieurs étapes. Comme nous avons constaté qu'il n'existait pas de corpus médical en pictogrammes et que le domaine de la santé couvre beaucoup de situations de communication différentes, nous nous sommes demandé comment traduire des questions et instructions médicales en pictogrammes. Pour la spécialisation de Text-to-Picto, nous avons besoin d'un corpus contenant les phrases que nous voulons traduire. Nous avons donc décidé de nous concentrer sur un corpus médical français, réalisé pour BabelDr (cf. chapitre 2, section 1.4.1). Il contient les questions et instructions nécessaires pour faire une anamnèse aux urgences sur la base de l'échelle de tri. Afin d'améliorer Text-to-Picto et concevoir des règles adaptées à notre cas d'utilisation, nous effectuons d'abord une analyse syntaxique des phrases de ce corpus (section 1). Nous demandons à une logopède de traduire manuellement une centaine de ces phrases en pictogrammes, afin de déterminer des stratégies de traduction implémentables au sein d'un système (section 2). Ces deux étapes nous ont amené à proposer des adaptations à Text-to-Picto comme l'ajout de règles de déplacement ou de suppression de mots (section 3). Nous évaluons cette version du système de traduction après l'avoir paramétrée sur une partie de notre corpus médical, afin de voir l'impact de ces premières adaptations. Nous menons finalement des expériences pour essayer d'améliorer la traduction automatique vers les pictogrammes via la substitution lexicale générale et médicale dans Text-to-Picto (section 4), mais aussi via la simplification automatique avec des grands modèles de langue génératifs (section 5). Enfin, nous concluons par nos contributions (section 6). Les sections 1 et 2 répondent au sous-objectif 3 ; la section 3 au sous-objectif 4 ; les sections 4 et 5 au sous-objectif 5.

1 Corpus de BabelDr et analyse syntaxique

L'un des deux objectifs de cette thèse étant le développement d'un système de traduction automatique de phrases en pictogrammes pour la communication médicale entre médecins et patients, nous avons besoin d'un corpus de phrases médicales contenant les phrases à traduire en pictogrammes. Nous avons donc décidé d'utiliser tout au long de cette thèse un sous-ensemble du système BabelDr (Bouillon *et al.*, 2021; Spechbach *et al.*, 2019; Bouillon *et al.*, 2017) : des phrases types (cf. chapitre 2, section 1.4.1), que nous appellerons les canoniques. Il s'agit des phrases de BabelDr, qui servent de base à la traduction des autres langues. Elles ont la particularité d'être relativement simples, de ne pas comporter d'erreurs syntaxiques ou lexicales, en plus d'être explicites, c'est-à-dire sans ellipse, un phénomène très présent dans le discours médical qui peut poser des problèmes de traduction (Rayner *et al.*, 2018).

Le système BabelDr comprend 10 000 phrases canoniques reliées à des millions de variantes syntaxiques et sémantiques, organisées par domaine. Par exemple, la canonique « avez-vous eu de la fièvre ? » peut avoir les variantes suivantes : « vous avez eu de la température ? », « est-ce que vous avez fait de la fièvre ? », « avez-vous fait de la fièvre ? », « avez-vous été fiévreuse ? », « fiévreux ? », etc. Si nous arrivons à traduire correctement ces canoniques en pictogrammes, nous pourrions élargir la couverture de Text-to-Picto plus facilement en construisant, par exemple (Figure 4.1), un dictionnaire qui lierait directement des variantes du système BabelDr (de potentielles phrases françaises entrées par des médecins dans Text-to-Picto) aux phrases canoniques.

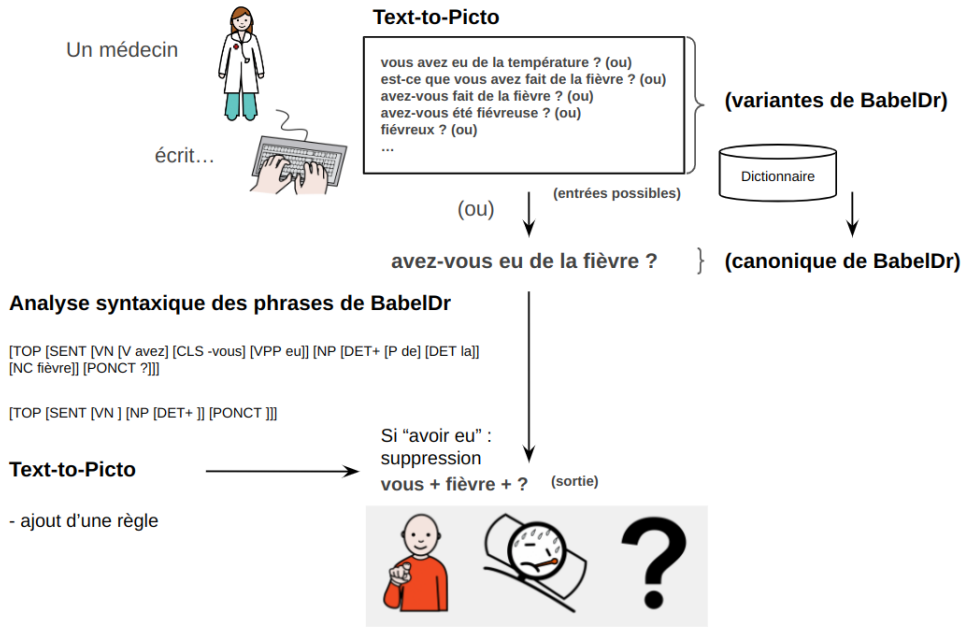


FIGURE 4.1 – Méthodologie pour explorer les façons de traduire une phrase pour la communication entre médecins et patients avec Text-to-Picto : exemple de la phrase canonique de BabelDr « avez-vous eu de la fièvre ? »

Afin d'explorer les façons de traduire en pictogrammes une phrase médicale issue de l'échelle de tri des urgences hospitalières et développer le système Text-to-Picto (sous-objectif 3), nous avons donc effectué une analyse syntaxique superficielle des phrases que cherchons à traduire, les canoniques (Figure 4.1). Elle visait à extraire les caractéristiques de ces phrases et déterminer quels phénomènes linguistiques doivent être traduits en pictogrammes dans Text-to-Picto.[1] En fonction de ceux-ci, nous pouvons concevoir des règles pour améliorer les traductions. Le corpus de canoniques de BabelDr que nous analysons contient 2 229 phrases.

Pour notre analyse syntaxique, nous avons utilisé le Berkeley Neural Parser avec un modèle de langue pour le français de la librairie Python benepar (Kitaev et Klein, 2018),[2] qui exploite les tags du French TreeBank (Abeillé *et al.*, 2003). Par exemple, la phrase indépendante (SENT) « à quelle date ont commencé les symptômes ? » est représentée par « [TOP [SENT [PP [P à] [NP [DETWH quelle] [NC date]]] [VN [V ont] [VPP commencé]] [NP [DET les] [NC symptômes]] [PONCT ?]]] » (Figure 4.2).[3]

Étant donné que TreeTagger (Schmid, 1994) générait des erreurs de tokenisation, nous avons effectué une tokenisation sur la base de règles manuelles et corrigé des erreurs d'étiquetage du parseur dues à des expressions multi-mots ou des mots composés comme « immuno-suppresseur ». Nous avons ajouté l'étiquette lexicale DM (Discourse Marker) du Treebank pour le français parlé (Abeillé et Crabbé, 2013), la locution « s'il vous plaît » étant fréquente et mal étiquetée. Nous avons également intégré l'interjection (I) pour « bonjour », « bonsoir », « au revoir », etc.

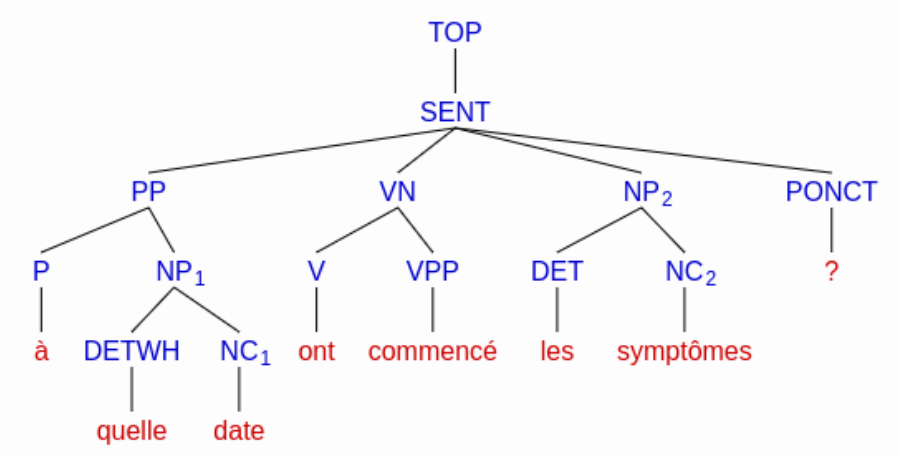


FIGURE 4.2 – Arbre syntaxique en constituants : exemple de la phrase canonique de BabelDr « à quelle date ont commencé les symptômes ? »

1. À noter que Sevens (2018) avait déjà utilisé cette méthodologie pour adapter Text-to-Picto aux phrases qu'elle cherchait à traduire : le corpus d'emails néerlandais extraits de WAI-NOT (cf. chapitre 3, section 3.1). C'est après cette analyse qu'elle avait conçu son module de simplification syntaxique.

2. https://github.com/nikitakit/self-attentive-parser ; https://pypi.org/project/benepar

3. PP : syntagme prépositionnel ; P : préposition ; NP : syntagme nominal ; DETWH : déterminant interrogatif ; NC : nom commun ; VN : syntagme verbal ; V : verbe ; VPP : participe passé ; DET : déterminant ; PONCT : ponctuation (http://ftb.linguist.univ-paris-diderot.fr).

Nous avons ensuite calculé la fréquence des structures syntaxiques des phrases de BabelDr (corpus oral). D'un point de vue syntaxique, notre corpus de phrases est assez homogène puisqu'il s'agit des phrases types. Nous avons un nombre de structures identiques élevé dans BabelDr (Table 4.1).

Structure	**Fréq.**	**Fréq. rel.**	**Exemple**
(VN) (NP) (PONCT)	210	9,42 %	avez-vous des frissons ?
(VN) (NP (PP (NP))) (PONCT)	128	5,74 %	avez-vous des problèmes de mémoire ?
(VN) (NP) (PP (NP)) (PONCT)	95	4,26 %	avez-vous des boutons sur la peau ?
(VN) (NP (AP)) (PONCT)	76	3,40 %	avez-vous des lésions cutanées ?
(NP) (VN) (PP (NP)) (PONCT)	60	2,69 %	combien de cigarettes fumez-vous par jour ?
(VN) (PP (NP)) (PONCT)))	57	2,55 %	allez-vous régulièrement à selles ?
(VN) (VPINF (VN) (NP))))	56	2,51 %	je vais examiner votre gorge
(VN) (PONCT)))	43	1,92 %	avez-vous déjà avorté ?
(NP) (VN) (PONCT)))	41	1,83 %	votre toux s'aggrave-t-elle ?
(VN) (NP)))	39	1,74 %	fermez les yeux, s'il vous plaît
(NP) (VN) (AP) (PONCT)))	36	1,61 %	la douleur est-elle récente ?
(PP (NP)) (VN) (NP) (PONCT)	30	1,34 %	de quelle couleur sont vos selles ?
(VN) (NP (NC+)) (PONCT)	30	1,34 %	avez-vous déjà fait une crise cardiaque ?

TABLE 4.1 – Structures syntaxiques présentes 30 fois ou plus dans le corpus de canoniques de BabelDr, la fréquence relative est mesurée par rapport au nombre total de structures (le total de phrases)

Seules deux de nos structures les plus fréquentes représentent des déclaratives. Nous avons au total 1 850 phrases interrogatives avec le tag PONCT à la fin (83 %) et 379 sans (17 %). Parmi ces phrases interrogatives, nous avons identifié différentes structures : VN NP ... / VN PP ... / NP VN ... / PP VN ... Toutes les questions sont marquées linguistiquement, c'est-à-dire par une inversion verbe + sujet, une locution interrogative (adverbe, pronom ou déterminant), etc. Or, en français, il existe des questions de type sujet + verbe qui n'ont pas de marqueur linguistique à l'écrit car elles se distinguent par la prosodie. Nous pourrions donc envisager de faire une règle dans Text-to-Picto pour garder l'ordre sujet + verbe + objet avec le point d'interrogation à la fin et tester si c'est compréhensible en pictogrammes. Par exemple : « avez-vous mal à la gorge ? » pour « vous avez mal à la gorge ? ». [4]

Dans la suite de cette thèse, nous implémenterons dans Text-to-Picto des expressions régulières pour ces cas et nous ferons évaluer des phrases sujet + verbe + objet, automatiquement, manuellement et directement auprès de personnes avec une DI (cf. chapitre 6). Presque toutes nos phrases commencent par un syntagme verbal (VN), qui inclut un pronom (CLS). En plus des pronoms à la première personne (« je »/« nous »), nous avons aussi souvent des constructions impersonnelles (« il (ne) faut (pas) », « il y a », etc.), qui n'étaient pas facilement traduisibles en pictogrammes d'après l'évaluation du système Text-to-Picto de base.

4. Certains CLS pourraient aussi être supprimés (« les vertiges/lésions sont-ils/elles » avec répétition du sujet, « y a-t-il », etc.), mais pas dans toutes les questions (« pouvez-vous »).

2 Analyse des stratégies de traduction en pictogrammes

Comme notre objectif est de traduire automatiquement des instructions et des questions de médecins destinées aux patients avec une DI en pictogrammes Arasaac, nous avons dû faire face à plusieurs problèmes : l'absence de grand corpus médical authentique en pictogrammes provenant d'utilisateurs d'outils de Communication Alternative et Améliorée (CAA), ainsi que le manque de directives de traduction pour créer un tel corpus. Par conséquent, nous avons étudié les stratégies réelles utilisées par une logopède pour effectuer une telle traduction (Norré *et al.*, 2022). Cette section liste les stratégies de traduction observées sur des phrases canoniques de BabelDr.

Comme Vandeghinste et Schuurman (2014) l'ont noté, une traduction en pictogrammes n'est pas une traduction littérale. Diverses stratégies de traduction peuvent être utilisées si un pictogramme est manquant dans l'ensemble. La plus simple consiste à chercher le lemme de chaque mot à traduire dans un ensemble de pictogrammes (Vaschalde, 2018). Cette opération de base repose sur la sélection du pictogramme dont le nom de fichier est identique au lemme dans la phrase source. Pour voir s'il existe d'autres stratégies, nous avons demandé à une logopède à la retraite – formatrice et directrice d'une association belge de CAA – de traduire manuellement 100 canoniques de BabelDr (issues du Corpus Médical, cf. chapitre 3, section 3.3) en pictogrammes Arasaac tout en gardant à l'esprit des utilisateurs cibles.

Le choix d'une logopède est justifié par le fait que ce sont souvent elles qui apprennent aux patients avec une DI à utiliser ces images (cf. chapitre 1). Ces traductions nécessiteraient toutefois d'être validées par d'autres personnes bien que chaque logopède traduise différemment en fonction de ses patients et de sa propre expérience. À notre connaissance, il n'existe pas de formation en CAA dans les universités et les hautes écoles belges. Schwab (2021, p. 17) l'avait également constaté en France. Néanmoins, en comparant les phrases sources et leur traduction en pictogrammes (Figure 4.3), nous avons noté 10 opérations syntaxiques ou lexico-sémantiques (cf. annexe A.2), qui permettent d'améliorer la traduction et/ou sa couverture :

- **suppression :** supprimer un ou plusieurs mot(s) de la phrase source (comme certains articles, auxiliaires, conjonctions, etc.) ; [5]

- **insertion :** s'il n'existe pas de pictogramme pour un mot, insérer une paraphrase avec des concepts généraux susceptibles d'être plus facilement compréhensibles pour le patient (« aphtes » → « boutons dans bouche », comme dans Strasly *et al.* (2018, pp. 94-95) pour traduire la langue des signes) ou insérer une clarification (« allergies connues » → « allergie allergie_2 allergie_3 » ; différentes allergies) ;

5. La suppression dépend des capacités de l'utilisateur avec une DI. Il convient de noter que supprimer tous les mots par POS peut changer beaucoup le sens de la phrase, c'est par exemple le cas si l'adverbe « ici » est supprimé dans la phrase « il est interdit de chanter ici » (Vaschalde, 2018, p. 50).

- **déplacement :** déplacer un ou plusieurs mot(s) de la phrase source (« toussez-vous depuis longtemps ? » → « depuis longtemps vous tousser ? »), la logopède nous conseillait de toujours garder l'ordre sujet + verbe et de déplacer les compléments de temps au début de la phrase ;
- **synonyme :** substituer un mot par un synonyme ayant un *Part Of Speech (POS)* identique (« actuellement » → « maintenant ») ;
- **hyperonyme :** substituer un mot par un hyperonyme ayant un POS identique (« coronavirus » → « virus ») ;
- **hyponyme :** substituer un mot par un hyponyme ayant un POS identique (« personne » → « individu ») ;
- **antonyme :** substituer un mot par un antonyme ayant un POS identique en ajoutant la négation « non » (« sèche » → « mouillé [non] ») ; [6]
- **changement de POS :** substituer un mot par un autre ayant la même racine, mais un POS différent (« avorté » → « avortement ») – verbe ↔ nom, nom ↔ adjectif, adjectif ↔ adverbe ;
- **remplacement d'expression multi-mots par un mot unique :** substituer une expression multi-mots par un synonyme en un seul mot, c'est souvent le cas des pictogrammes complexes qui semblent fréquents dans les phrases de BabelDr (« poser des questions » → « demander ») ;
- **autre remplacement :** substituer – arbitrairement – un mot par un autre ayant une racine et/ou un POS différent, mais pas par un synonyme, hyperonyme, etc. (« quarantaine » → « isoler » ; « confirmé » → « plus »).

Nous donnons quelques exemples de possibles traductions médicales en pictogrammes Arasaac validées par la logopède interrogée (Figure 4.3). Plusieurs opérations de traduction peuvent être combinées dans une phrase et/ou concerner un même mot (par exemple, un déplacement et la substitution par un synonyme, une insertion avec un remplacement, etc.). Ces transformations étaient déjà partiellement prises en compte dans le système de base du Text-to-Picto français (cf. chapitre 3). Il convient de noter qu'elles sont assez similaires aux opérations pour la simplification générale en anglais utilisées par Cardon *et al.* (2022, 2023), mais aussi pour la simplification médicale en français (Cardon, 2021; Koptient *et al.*, 2019). Contrairement à ces auteurs, nous ne travaillons pas à partir de mots, mais avec des lemmes. Il n'y a donc pas de transformations liées à la flexion (singulier vers pluriel et inversement ou modification

6. Dans ce cas-ci, le pictogramme « mouillé » lui semblait moins adapté que celui représentant une « goutte » avec une croix rouge sur le pictogramme pour symboliser la négation.

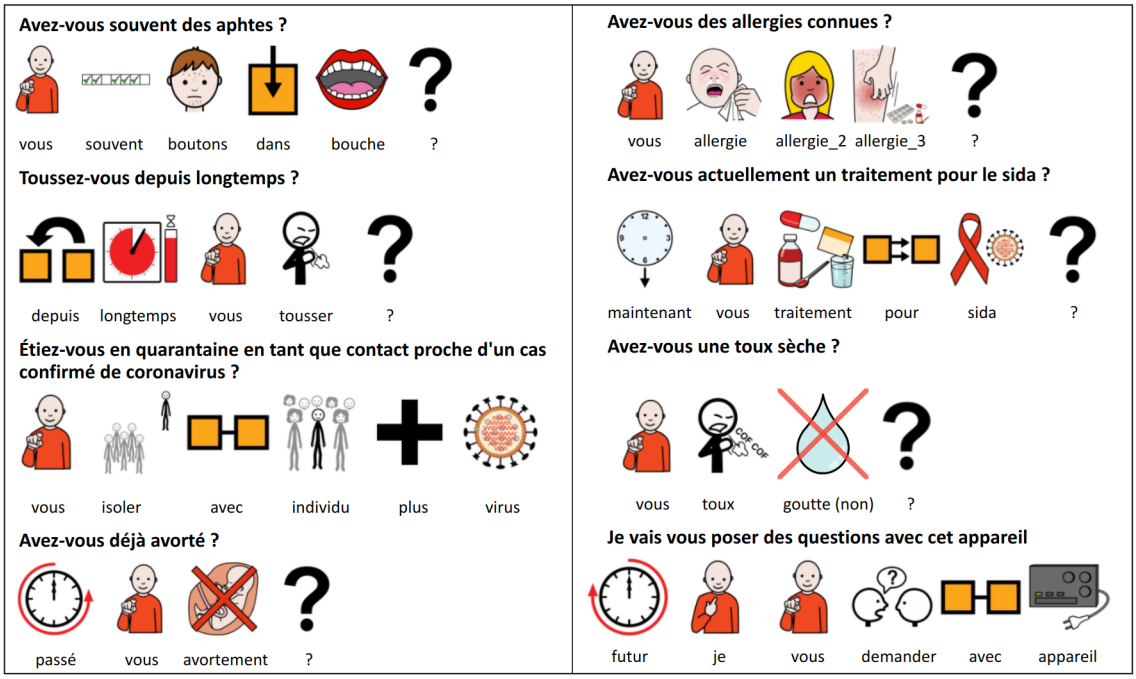


FIGURE 4.3 – Exemples de possibles traductions médicales en pictogrammes Arasaac

des traits verbaux). L'opération la plus fréquente dans notre corpus est la suppression. [7] Ici, le nombre moyen de pictogrammes utilisés par la logopède par phrase est de 4,54. Il est plus faible que le nombre moyen de mots dans la phrase source (6,78 sans le « ? » et 7,58 avec ; elle utilisait d'ailleurs un pictogramme pour représenter le « ? »).

3 Analyse de la couverture médicale de Text-to-Picto

Suite aux analyses réalisées à partir des résultats du système, du corpus BabelDr et des conseils de la logopède, nous apportons des modifications au système Text-to-Picto pour améliorer les traductions (sous-objectif 4). Cette fois, nous le paramétrons avec des phrases canoniques du Corpus Médical (cf. chapitre 3, section 3.3), avant de réaliser des évaluations automatiques et manuelles pour analyser la couverture médicale de notre système de traduction du français vers les pictogrammes Arasaac.

3.1 Premières adaptations : version 2 de Text-to-Picto

Nous proposons donc des premières adaptations à notre système. Nous appellerons cette version intermédiaire de Text-to-Picto la version 2 (v2).

En 2022, nous avons mis à jour la base de données avec de nouveaux pictogrammes provenant de l'API d'Arasaac (cf. chapitre 3, section 2). Des centaines sont

7. Plus tard, nous avons eu accès à un corpus de 304 phrases pour la CAA générale (collectées par l'Université Grenoble Alpes). Nous avons noté que 50 % des mots n'avaient pas été traduits en pictogrammes Arasaac. Il s'agissait surtout de mots de fonction, d'auxiliaires ou d'expressions multi-mots.

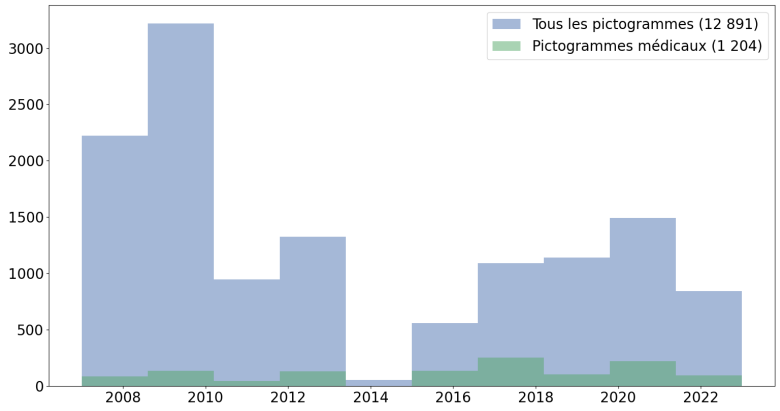


FIGURE 4.4 – Nombre de pictogrammes Arasaac créés par an entre 2008 et 2022

ajoutés par an (Figure 4.4), dont des pictogrammes médicaux en raison notamment de la pandémie de COVID-19 (2020-2022). Dans la table des pictogrammes, nous avons inséré la possibilité de désactiver chaque pictogramme. Nous avons déplacé certains lemmes dans le dictionnaire Arasaac, comme les chiffres et les nombres qui n'étaient pas traduits dans le système de base. Nous avons aussi testé l'ajout d'un poids supplémentaire, associé aux pictogrammes qui ont été utilisés par la logopède pour traduire une partie du Corpus Médical. Nous avons vu que jusqu'à présent il pouvait arriver que le système affiche des pictogrammes différents s'ils ont le même poids lors de l'étape finale (la recherche de la meilleure séquence de pictogrammes). Le poids ajouté a donc pour but de réduire le nombre de possibilités trop élevées dues notamment à l'absence d'un module de désambiguïsation lexicale en favorisant l'usage de certains pictogrammes (sans devoir désactiver un certain nombre de pictogrammes manuellement). Durant l'entretien, la logopède nous a suggéré de mettre les compléments de temps et une indication du temps du verbe au début des phrases. Nous avons choisi de ne pas déplacer les compléments de temps, mais nous générons un pictogramme Arasaac 🕒 pour le passé ⟲ et le futur ⟳ (Figure 4.5), c'est-à-dire si la phrase à traduire contient respectivement le mot « déjà » et le verbe « aller » suivi d'un verbe à l'infinitif. [8] Sevens (2018, p. 150) générait déjà, juste avant le verbe, un pictogramme Sclera et Beta en forme d'horloge 🕒 avec un symbole ajouté au-dessus pour le passé ←, le futur → et le conditionnel ?. [9] Nous avons repris ce mécanisme, mais nous faisons la génération en début de phrase – sur la base de conditions différentes.

Dans l'analyse linguistique superficielle (cf. chapitre 3, section 1) de cette version, nous avons ajouté des règles (expressions régulières), la compression pour divers phénomènes en français. Nous supprimons des mots de fonction et des auxiliaires. La

8. Il aurait aussi été possible de générer une flèche sur les pictogrammes des verbes Arasaac (par exemple, pour « manger » : https://static.arasaac.org/pictograms/2349/2349_action-past_500.png).

9. Nous n'avons pas ajouté le conditionnel étant donné que ce temps est peu utilisé dans les phrases de BabelDr. De plus, il n'existe pas de pictogramme Arasaac « conditionnel ».

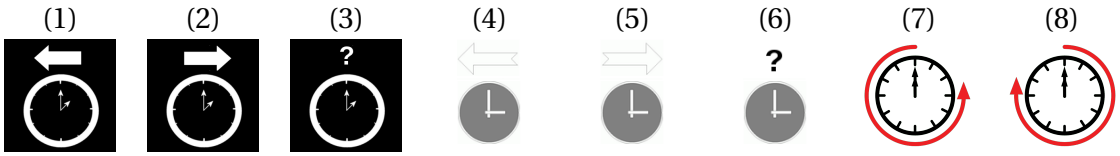


FIGURE 4.5 – Pictogrammes Sclera (1-3), Beta (4-6) et Arasaac (7-8) : exemple des temps (1/4/7) « passé », (2/5/8) « futur » et (3/6) « conditionnel »

compression de certains mots – comme les articles, les conjonctions, etc. – peut se justifier par le fait que leurs représentations peuvent ne pas toujours être facilement compréhensibles si leur signification n'a pas été apprise au préalable. Or, nous pouvons supposer que tous les patients avec une DI qui utiliseront ce système ne connaîtront pas les pictogrammes, ni l'ensemble Arasaac. Les utilisateurs avec une DI n'ont pas tous les mêmes connaissances grammaticales. Comme suggéré par la logopède, nous faisons l'inversion verbe + sujet dans les questions des médecins et nous simplifions certaines structures impératives ou constructions impersonnelles qui sont fréquentes dans BabelDr pour les instructions destinées aux patients (« je vais vous demander de (vous) » simplifié en « vous », « il (ne) faut (pas) » en « vous devez (pas) », etc.).

Nous proposons également de tester une relation du WOrdnet Libre du Français (WOLF) qui n'était pas incluse dans la version française de Text-to-Picto : *eng_derivative*. Elle permet d'obtenir un concept similaire avec un POS différent (une des stratégies de la section 2). Par exemple, Text-to-Picto traduit l'adjectif « respiratoire » par le verbe « respirer ». Lorsqu'il doit traduire une « (incontinence) anale » ou « urinaire », il affiche respectivement les pictogrammes « anus » ou « poche_d'urostomie ». Il s'agit de l'équivalent de la relation *xpos_near_synonym* [10] dans Cornetto (Vossen *et al.*, 2008), le WordNet néerlandais utilisé par Sevens (2018). Enfin, nous avons vu que plusieurs systèmes de CAA avec des pictogrammes – dont celui de Bautista *et al.* (2017) – utilisaient un code couleur qui informe sur la catégorie syntaxique des mots représentés. Ce code couleur permettrait d'améliorer l'apprentissage du vocabulaire et par conséquence son utilisation. Il en existe plusieurs. Pour les bordures des pictogrammes, nous avons implémenté un module optionnel : le code de Fitzgerald (1949), à savoir, vert pour les verbes, bleu pour les adjectifs, orange pour les noms, jaune pour les pronoms et blanc pour les articles, les adverbes, etc.

3.2 Paramétrage du système

Afin d'analyser la couverture de Text-to-Picto (v2), nous avons d'abord réglé le système sur un sous-ensemble du Corpus Médical et non plus sur le Corpus Email comme c'était le cas avec notre système de base (cf. chapitre 3, section 4.2).

10. C'est cette appellation que nous avons choisi d'utiliser dans nos prochaines évaluations.

	Min	Max	Pas	- Relation xpos	+ Relation xpos
Arasaac					
WOLF relations					
Threshold	5	20	1	9	7
Hyperonym penalty	0	15	1	15	15
Antonym penalty	0	15	1	10	2
Xpos penalty	0	15	1	–	14
Route preference					
Out-Of-Vocabulary	0	10	1	3	3
Direct route advantage	0	15	1	2	3

TABLE 4.2 – Valeurs des paramètres du Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac après réglage

Pour le réglage et l'évaluation automatique, nous avons donc retraduit manuellement 150 phrases du Corpus Médical en pictogrammes Arasaac en se basant sur les stratégies de traduction en pictogrammes observées durant l'entretien avec la logopède (cf. section 2). Nous avons utilisé 60 phrases pour régler les paramètres de Text-to-Picto avec l'algorithme hill-climbing (Vandeghinste *et al.*, 2017), en lançant toujours le système 5 fois avec, à chaque fois, 50 itérations pour chacune des deux conditions (Table 4.2) – sans (-) et avec (+) la relation *xpos_near_synonym*[11] – tout en utilisant la métrique BLEU (Papineni *et al.*, 2002).

3.3 Évaluation automatique

Ensuite, nous avons évalué automatiquement le système de traduction sur les 90 phrases restantes en réutilisant les métriques (cf. chapitre 3, section 5.1) : BLEU, NIST (Doddington, 2002), WER et PER. Chaque métrique a de nouveau été estimée sur la base d'une moyenne de 10 traductions générées par Text-to-Picto. Les écart-types sont toujours présentés entre parenthèses (Table 4.3).

	BLEU ↑	NIST ↑	WER ↓	PER ↓
Arasaac				
- Relation xpos_near_synonym	**30,32** (2,32)	**4,19** (0,14)	**55,54** (2,11)	**50,58** (1,96)
+ Relation xpos_near_synonym	27,35 (2,29)	3,79 (0,18)	61,40 (2,65)	56,36 (2,50)

TABLE 4.3 – Résultats du Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac avec les métriques BLEU, NIST, WER et PER

Les scores BLEU sont cohérents par rapport à ceux du système de base. Nous avions en effet obtenu un score de 31,36 sur les 260 phrases du Corpus Médical dans lequel tous les mots devaient être traduits, y compris les mots de fonction (Norré *et al.*,

11. Il convient de noter que la relation *xpos_near_synonym* augmente considérablement le temps d'analyse de certaines phrases, les possibilités dans l'espace de recherche devenant très élevées.

2021c,d). [12] Ce score était déjà plus élevé que ceux pour les autres corpus (Email 24,97 et Livre 28,34). Ajouter la relation *xpos_near_synonym* dans notre système dégrade les résultats de l'évaluation automatique puisque nous obtenons un BLEU de 27,35 avec cette nouvelle relation et 30,32 sans. Vandeghinste *et al.* (2017) n'avaient pas non plus obtenu de meilleurs résultats en ajoutant d'autres relations que la synonymie. Les écart-types sont plus élevés que pour le système de base malgré nos modifications.

3.4 Évaluation manuelle

Deux expériences ont été réalisées sur deux ensembles de phrases de BabelDr différents, que nous avons fait traduire automatiquement par Text-to-Picto en pictogrammes Arasaac. La première vise à évaluer le nombre de mots non traduits (section 3.4.1) pour cette nouvelle version de Text-to-Picto. La deuxième, quant à elle, vise à analyser la couverture médicale du système (section 3.4.2).

3.4.1 Expérience 1 : le nombre de mots non traduits

Nous avons utilisé 700 transcriptions textuelles différentes (4 596 mots) de questions et instructions de médecins à destination de réels patients. Elles ont été enregistrées avec la reconnaissance vocale du système BabelDr en contexte réel durant une étude de cohorte (Janakiram *et al.*, 2020), au service ambulatoire des Hôpitaux Universitaires de Genève (HUG). Nous avons aussi évalué les 700 canoniques (4 677 mots), qui ont été liées manuellement à ces transcriptions. Nous avons calculé le nombre de mots non traduits sur ces deux corpus.

	Transcription	**Canonique**
Arasaac		
- Relation xpos_near_synonym	126 (**191**)	**102** (229)
+ Relation xpos_near_synonym	110 (**159**)	**81** (203)

TABLE 4.4 – Nombre de types (et nombre de tokens) non traduits avec Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac

Nous donnons le nombre de types non traduits ainsi que le nombre de tokens entre parenthèses (Table 4.4). L'utilisation de la relation *xpos_near_synonym* permet de traduire plus de mots – que ce soit dans les transcriptions ou les canoniques – bien que nous n'ayons pas vérifié si tous ces mots étaient corrects (calcul de la précision). Nous constatons qu'il y a moins de types non traduits dans les canoniques (sans et avec la relation xpos), mais qu'il y a plus de tokens non traduits que dans les transcriptions.

Nous pouvons supposer que les transcriptions possèdent un vocabulaire un peu plus varié. Ce constat peut aussi s'expliquer par le fait que nous n'avions pas de phrases

12. Ces mots étaient bien traduits car ils sont inclus dans les dictionnaires de Text-to-Picto.

doublons dans les transcriptions, alors qu'il y en avait dans les canoniques : différentes transcriptions étant parfois liées à une même canonique. Parmi les mots non traduits des transcriptions, nous notons la présence d'entités nommées comme des noms de médecin ou des noms abrégés de maladie (l'abréviation « MST » pour « Maladie Sexuellement Transmissible »). Dans les deux ensembles de phrases, nous trouvons des noms de médicaments (« Irfen », « Dafalgan », etc.), qui n'existent pas en pictogrammes.

	Transcription		**Canonique**	
Arasaac	**POS**	**Fréq.**	**POS**	**Fréq.**
- Relation xpos_near_synonym	ADJ	45	ADJ	46
	NOM	39	NOM	31
	ADV	16	ADV	13
+ Relation xpos_near_synonym	NOM	37	ADJ	32
	ADJ	33	NOM	26
	ADV	17	ADV	11

TABLE 4.5 – Nombre de types non traduits avec Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac par POS les plus fréquents

Nous avons utilisé TreeTagger sur les types non traduits. Les POS les moins traduits dans les quatre ensembles sont : les adjectifs, les noms et les adverbes (Table 4.5). Les verbes – à l'exception de quelques participes passés – ont souvent été traduits en pictogrammes. Dans les transcriptions de la reconnaissance vocale, nous avons noté que des mots n'avaient pas été traduits à cause d'erreurs orthographiques (« esque » pour « est-ce que », « set » pour « est », « riques » pour « risques », etc.). Sevens (2018, pp. 75-97) avait déjà constaté ce problème et l'avait résolu en ajoutant un correcteur orthographique dans la version néerlandaise de Text-to-Picto (Sevens *et al.*, 2016b). [13]

3.4.2 Expérience 2 : la couverture médicale

En ce qui concerne la couverture médicale, deux juges (dont l'auteur) ont évalué manuellement 50 traductions de canoniques générées automatiquement avec Text-to-Picto sans la relation xpos. Pour juger si le sens avait été conservé, ils ont utilisé une liste de concepts – issus du méta-thésaurus *Unified Medical Language System (UMLS)*, une ressource terminologique décrite par Bodenreider (2004) – liés à chacune de ces phrases. Cette analyse par concept (et non par token) a permis d'évaluer à la fois les mêmes éléments et le même nombre, ce qui aurait été plus difficile étant donné que certains mots (conjonctions, auxiliaires, articles, etc.) ont été supprimés par le système.

Pour chacun des 103 concepts UMLS des 50 phrases, [14] les annotateurs (A1 et

13. Comme nous avons choisi de traduire les canoniques, nous n'avons pas à gérer ce problème. Néanmoins, même si on peut supposer que les médecins écrivent mieux que les patients avec une DI, un correcteur orthographique adapté pourrait également être utile.

14. Ou 187 concepts si nous incluons les éléments ne faisant pas partie d'UMLS (les pronoms personnels, les questions, etc.). Nous donnons les résultats sur ce total entre parenthèses dans la table.

Sem(antic)Types	Concepts	Concept Unique Identifiers	A1	A2
Depuis combien de jours avez-vous mal à la tête?				
Quantitative Concept	Quantity	C1265611	Y	Y
Temporal Concept	day	C0439228	Y	Y
Temporal Concept	Since	C1711239	Y	Y
Agent	You	–	Y	Y
Sign or Symptom	Headache	C0018681	Y	Y
Mode	Question	–	Y	Y
Nous allons faire un examen				
Agent	I	–	Y	Y
Health Care Activity	Medical Examination	C0582103	Y	Y
Pourriez-vous être enceinte?				
Agent	You	–	Y	Y
Finding	Possible	C0332149	Y	Y
Organism Function	Pregnancy	C0032961	N	N
Mode	Question	–	Y	Y
Avez-vous déjà eu ce type de douleur?				
Agent	You	–	Y	Y
Finding	History of present illness (finding)	C0262512	Y	Y
Sign or Symptom	Pain	C0030193	Y	Y
Mode	Question	–	Y	Y
Nous allons faire une prise de sang				
Agent	I	–	Y	Y
Diagnostic Procedure	Collection of blood specimen	C0005834	N	N
Avez-vous un traitement régulier?				
Agent	You	–	Y	Y
Qualitative Concept	Regular	C0205272	N	N
Therapeutic Procedure	Therapeutic procedure	C0087111	Y	Y
Mode	Question	–	Y	Y

TABLE 4.6 – Extrait de l'évaluation du Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac concernant la couverture par concept

A2) devaient dire si les concepts étaient correctement traduits (Y) ou non (N) en pictogrammes (Table 4.6). Les concepts UMLS « *Pregnancy* » (« enceinte »), « *Collection of blood specimen* » (« prise de sang ») et « *Regular* » (« régulier ») ont par exemple été jugés incorrectement traduits (cf. Figure 4.6). Les deux annotateurs devaient aussi indiquer le type de représentation (Littéral/Synonyme par défaut, Hyperonyme/générique, Hyponyme/spécifique et Polysème) sur la base de l'image (pas du nom de fichier). Le coefficient κ (Cohen, 1960) pour l'accord inter-annotateur est de 0,65 (0,64 avec tous les concepts), indiquant que l'accord entre les deux annotateurs est substantiel selon l'échelle de Landis et Koch (1977), mais non fiable [< 0,67] d'après Krippendorff (2004).

Les résultats pour l'évaluation de la couverture du système par concept UMLS varient entre 62,1 et 71,8 % en fonction de l'annotateur (Table 4.7), soit une différence de 10 %. La stratégie lexicale la plus utilisée par le système Text-to-Picto consiste à recourir au littéral – un pictogramme dont le nom de fichier est identique au mot dans la phrase à traduire – ou au synonyme, qui étaient tous les deux non marqués (absence d'annotation). Les annotateurs ont néanmoins rapporté 5-6 hyperonymes (« symp-

	Annotateur 1	Annotateur 2	Moyenne
Arasaac			
Traduction correcte	62,1 % (75,4 %)	71,8 % (82,8 %)	66,95 % (79,1 %)
Pas de traduction	37,8 % (24,5 %)	28,1 % (17,1 %)	32,95 % (20,8 %)

TABLE 4.7 – Résultats du Text-to-Picto français (v2) sur le Corpus Médical pour l'ensemble de pictogrammes Arasaac concernant la couverture selon deux annotateurs

tomes » : « nausée » | « diarrhée »), 2-5 hyponymes (« examen des yeux » : « examen ») et 3-4 polysèmes (être « enceinte » ou les « haut-parleurs »). Certaines expressions multimots (« prise de sang ») ont été incorrectement traduites par deux pictogrammes (« tenir_2 » + « sang »). À noter que celles que nous avons liées manuellement à deux synsets (« mal de tête » : 00011516-r {mal, faiblement} + 05538625-n {tête}) ont été traduites par un seul pictogramme (« mal_de_tête_2 »). Les annotateurs ont également constaté que certains adjectifs et certains noms étaient incorrects ou n'avaient pas toujours été traduits – ils auraient d'ailleurs été difficilement compréhensibles en pictogrammes (« régulier », « type », etc.). Il nous semble en effet plus difficile de représenter les adjectifs et de les comprendre dans différents contextes si leur sens n'a pas été appris.

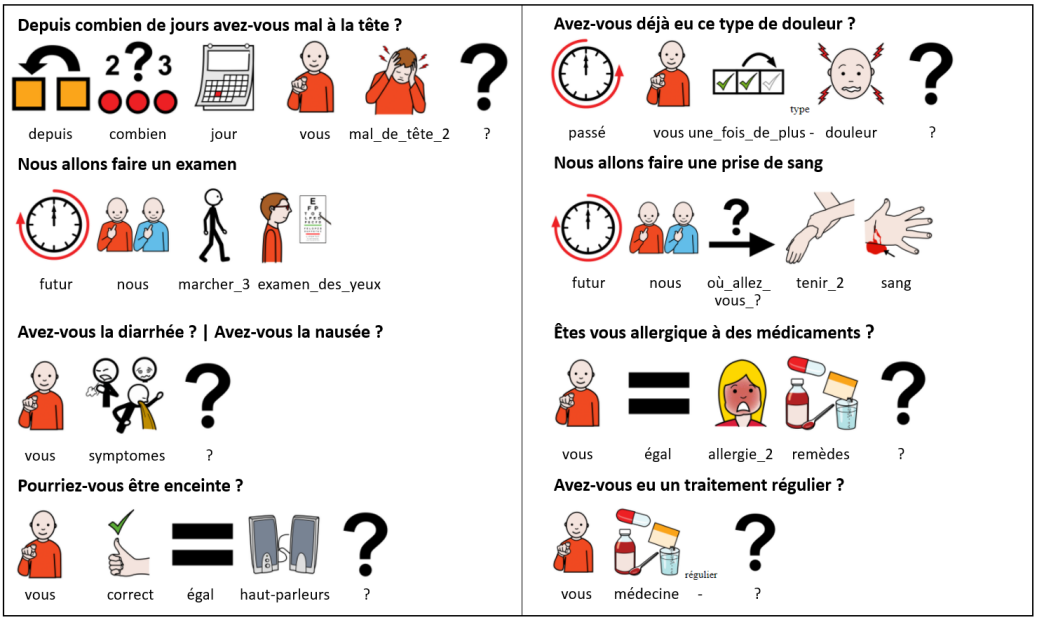


FIGURE 4.6 – Exemples de traductions médicales du Text-to-Picto pour le français (v2) en pictogrammes Arasaac

Nous présentons quelques exemples de sortie de Text-to-Picto en pictogrammes Arasaac (Figure 4.6). La version française de Text-to-Picto a un bon rappel. Toutefois, la précision de sa couverture doit encore être améliorée. C'est pourquoi, nous proposons de tester l'impact de la substitution lexicale avec des ressources générales ou médicales, ainsi que l'influence de la simplification automatique de textes sur les traductions de Text-to-Picto dans les deux prochaines sections (sous-objectif 5).

4 Simplification générale et médicale

Dans cette section, nous commençons par décrire les modifications que nous avons apportées à Text-to-Picto concernant la substitution lexicale, générale puis médicale pour le français (section 4.1). Nous évaluerons ensuite ces modifications afin de voir si elles améliorent les traductions en pictogrammes Arasaac (section 4.2).

4.1 Substitution lexicale : version 2.1 de Text-to-Picto

Nous présentons les ressources intégrées à Text-to-Picto : ReSyf (section 4.1.1), des paraphrases médicales (section 4.1.2), ainsi que leur emplacement (section 4.1.3).

4.1.1 Les synonymes de ReSyf

Pour étendre la couverture lexicale de Text-to-Picto, nous l'avons lié à la ressource de synonymes ReSyf (Billami *et al.*, 2018; François *et al.*, 2016), développée par le Centre de Traitement Automatique du Langage de l'Université catholique de Louvain et le Laboratoire Parole et Langage de l'Université d'Aix-Marseille. Le recours possible à ReSyf pour la simplification lexicale et l'obtention de synonymes pour les mots français « trop complexes » non présents dans un ensemble de pictogrammes d'une application de traduction speech-to-picto avait déjà été évoqué par Vaschalde (2018, p. 62), mais sans être testé. Dans Text-to-Picto, nous l'utilisons après l'analyse linguistique superficielle et le recours aux dictionnaires des ensembles de pictogrammes (route directe), mais avant la recherche dans WordNet (route sémantique).

ReSyf est disponible librement au format XML. [15] Nous l'avons prétraité avec OpenRefine (Verborgh et De Wilde, 2013), afin d'en faire un dictionnaire pour Text-to-Picto et ne garder que les données qui nous semblaient utiles (tri, suppression, etc.). Cette ressource comprend au total 87 914 ensembles de synonymes gradués en fonction de leur difficulté relative au sein de chaque ensemble (du plus facile au plus difficile). Les degrés de difficulté ont été obtenus automatiquement à partir d'un modèle prédictif (Billami *et al.*, 2018; François *et al.*, 2016) entraîné sur des ressources lexicales intégrant des informations sur la difficulté lexicale. Ce modèle s'appuie sur des variables formelles et psycholinguistiques mesurant la complexité lexicale (français langue première). Les lemmes ont été désambiguïsés et proviennent du réseau lexical JeuxDeMots (Lafourcade, 2007). Chaque ensemble de synonymes (que nous appellerons synset) a un sens (usage) différent. Il est lié à un lexème, le lemme (64 730 différents). Ce lexème a au moins une catégorie morphosyntaxique (POS), qui a été obtenue avec TreeTagger (58 294 NOM, 14 211 VER, 13 629 ADJ, 1 780 ADV). [16]

15. https://cental.uclouvain.be/resyf

16. Les lemmes et les synonymes peuvent être des expressions multi-mots. Par exemple, les verbes pronominaux (« se baigner »), les noms avec leur complément (« accident de voiture »), etc.

58 322 synsets sont monosémiques (dont 38 430 NOM, 9 270 ADJ, 9 058 VER, 1 514 ADV, soit 65,92 % des noms, 68,01 % des adjectifs, 63,73 % des verbes et 85,05 % des adverbes), tandis que 29 592 sont polysémiques (de 2 à 27 synsets pour le verbe « tenir »). Pour les entrées polysémiques, chaque synset a aussi une fréquence fixe entre 0 et 1. [17] Par exemple, le lexème « bleu » a 9 synsets nominaux dont la fréquence se situe entre 0,625 et 1, les plus fréquents étant le sens de 'couleur' (1), 'vêtement' (0,8036), 'ecchymose' (0,75), etc. (Figure 4.7). Comme souvent pour les synsets nominaux, celui-ci a aussi des synsets adjectivaux. Ici, l'adjectif « bleu » peut avoir le sens de 'fromage', 'policier', 'surpris', 'ivre', etc. Nous montrons un extrait du fichier XML (cf. infra), avant nos prétraitements avec OpenRefine. Il s'agit du lexème nominal « bleu » (NC pour nom commun), le 4e (id) synset avec l'usage 'ecchymose' (poids : 0,75), le 1er et le 2e (rank) synonyme (1 : oeil_au_beurre_noir | 2 : bleu).

```
<LexicalEntry>
    <feat att="ambiguity" val="1"/>
    <Lemma type="Form">
        <feat att="lexeme" val="bleu"/>
        <feat att="partOfSpeech" val="NC"/>
    </Lemma>
    <Sense id="bleu_NC_4">
        <feat att="poids" val="0.75"/>
        <feat att="usage" val="bleu (ecchymose)"/>
        <SenseExample id="bleu_NC_4">
            <feat att="type" val="synonyms"/>
            <feat att="annotation" val="manually"/>
            <feat att="score_lemma" val="--"/>
            <feat att="score_sense" val="--"/>
            <feat att="word" val="bleu"/>
            <feat att="rank" val="2"/>
        </SenseExample>
        <SenseExample id="bleu_NC_4">
            <feat att="type" val="synonyms"/>
            <feat att="annotation" val="automatically"/>
            <feat att="score_lemma" val="1.0"/>
            <feat att="score_sense" val="1.0"/>
            <feat att="word" val="oeil au beurre noir"/>
            <feat att="rank" val="1"/>
        </SenseExample>
        <...>
</LexicalEntry>
```

17. Selon François *et al.* (2016, p. 20), il s'agit d'un « poids sémantique » qui provient de la ressource JeuxDeMots. Ils ne donnent pas de précision sur le calcul des poids. Dans Text-to-Picto, nous avons appelé le poids de ReSyf : fréquence (FREQ, cf. Figure 4.7).

LEXEME	SENSE_ID	POS	ID	FREQ	USAGE	SYNONYMS
bleu	bleu_NOM_1	NOM	1	1	couleur	1:nouveau|2:ciel|3:bleu|4:canard|5:pneu|6:vert|7:azur|8:câble|9:violet|10:châtaigne|11:cerne|12:bleuet|13:pneumatique|14:interdit|15:indigo|16:lapis–lazuli
bleu	bleu_NOM_2	NOM	2	0.8036	vêtement	1:bleu|2:vêtement|3:combinaison|4:cotte|5:salopette
bleu	bleu_NOM_3	NOM	3	0.75	viande	1:bleu|2:canard|3:tournesol|4:roussette
bleu	bleu_NOM_4	NOM	4	0.75	ecchymose	1:oeil_au_beurre_noir|2:bleu|3:blessure|4:contusion|5:ecchymose|6:torgnole|7:hématome|8:meurtrissure
bleu	bleu_NOM_5	NOM	5	0.7321	fromage	1:bleu|2:fromage|3:châtaigne|4:roquefort
bleu	bleu_NOM_6	NOM	6	0.7143	policier	1:bleu|2:gendarme|3:interdit
bleu	bleu_NOM_9	NOM	9	0.7143	télégramme	1:bleu|2:pneu|3:câble|4:télégramme|5:pneumatique
bleu	bleu_NOM_10	NOM	10	0.7143	histoire militaire	1:bleu|2:gendarme
bleu	bleu_NOM_11	NOM	11	0.625	nouvelle recrue	1:nouveau|2:bleu|3:soldat|4:tache|5:cerne|6:novice|7:bleuet|8:tournesol|9:bleusaille|10:blanc–bec|11:débutant|12:recrue|13:roussette|14:cyanose|15:arpette|16:arpète
bleu	bleu_ADJ_1	ADJ	1	1	couleur	1:nouveau|2:bleu|3:vert|4:violet|5:azur|6:pneumatique|7:interdit|8:châtaigne|9:pers|10:bleuâtre|11:indigo|12:bleuet|13:céruléen
bleu	bleu_ADJ_5	ADJ	5	0.7321	fromage	1:bleu|2:châtaigne
bleu	bleu_ADJ_6	ADJ	6	0.7143	policier	1:bleu|2:interdit
bleu	bleu_ADJ_7	ADJ	7	0.7143	surpris	1:bleu|2:surpris
bleu	bleu_ADJ_8	ADJ	8	0.7143	ivre	1:bleu|2:ivre
bleu	bleu_ADJ_9	ADJ	9	0.7143	télégramme	1:bleu|2:pneumatique
bleu	bleu_ADJ_11	ADJ	11	0.625	nouvelle recrue	1:nouveau|2:bleu|3:nouveau_venu|4:novice|5:nouvelle_recrue|6:débutant|7:bleuet

FIGURE 4.7 – Ressource de synonymes ReSyf du Text-to-Picto pour le français (v2.1) : exemple du lexème « bleu »

Dans Text-to-Picto, nous avons implémenté une recherche en trois étapes distinctes : **a)** tous les lexèmes monosémiques ; **b)** des lexèmes polysémiques "médicaux" ; **c)** les lexèmes polysémiques les plus fréquents. Dans tous les cas, nous cherchons le lexème qui a le même POS que le lemme recherché, appelé n. Pour rappel, chaque mot donné en entrée à Text-to-Picto est étiqueté et lemmatisé avec TreeTagger, ce qui permet un mapping vers les entrées de ReSyf.

(a) Nous vérifions d'abord s'il s'agit d'un lexème monosémique. Dans ce cas, Text-to-Picto cherche parmi l'ensemble unique de synonymes, celui dont le rang se situe avant celui du lemme recherché, c'est-à-dire le $n-1$ (qui est plus facile et en général plus proche sémantiquement que celui situé au rang 1) dans ReSyf pour le traduire par le pictogramme correspondant. Le lemme obtenu via ReSyf est ensuite cherché dans la table des noms de fichier Arasaac, afin de récupérer son identifiant sans passer par l'analyse sémantique avec WordNet. Si le synonyme $n-1$ n'existe pas dans la table des pictogrammes, Text-to-Picto continuera de vérifier si le synonyme $n-2$, $n-3$, etc. de ReSyf peut être traduit en pictogramme. Si le rang du lexème est -1, cela signifie qu'il a été jugé très difficile ; nous cherchons alors uniquement le premier synonyme du synset concerné : « phylloxera » (Figure 4.8) sera par exemple traduit par « maladie ».

- Pour le lemme adverbial « récemment » (Figure 4.8) – qui n'est pas traduit en pictogramme avec WordNet ou le dictionnaire Arasaac –, ReSyf propose 8 synonymes en plus du lemme recherché (1 : depuis_peu | 2 : hier | 3 : depuis | 4 : récemment | 5 : frais | 6 : dernièrement | 7 : naguère | 8 : fraîchement | 9 : nouvellement). Le lemme que nous cherchons est le 4e, Text-to-Picto va le traduire par le 3e (« depuis ») et non pas par le 2e (« hier »), dont il existe aussi un pictogramme.

- Pour le lemme nominal « toundra » (Figure 4.8), ReSyf propose 4 synonymes (1 : désert | 2 : plaine | 3 : prairie | 4 : steppe | 5 : toundra). Cette fois, le lemme que nous cherchons est le 5e. Comme il n'existe pas de pictogramme pour le 4e (« steppe »), Text-to-Picto le traduira par le 3e (« prairie »).

(b) Nous cherchons ensuite les lexèmes polysémiques, en commençant par les synsets qui ont un certain sens. En prétraitant ReSyf, nous avons en effet décidé de lister 13 domaines d'usages (couvrant 306 lexèmes différents), [18] qui nous semblent intéressants pour le domaine médical.

- Pour le lemme nominal « médoc » (Figure 4.8), ReSyf propose deux synsets nominaux avec des sens et des fréquences différentes. Le premier a le sens de 'vin'

18. La liste est la suivante : 'chirurgie', 'chirurgie auditive', 'intervention chirurgicale', 'ecchymose', 'acte médical', 'dispositif médical', 'étudiant en médecine', 'médecin', 'médecine', 'médical', 'médicament', 'médicament générique', 'produit médical'.

et une probabilité de 1 (1 : bordeaux | 2 : médoc). Le deuxième a le sens de 'médicament' et une probabilité de 0,8958 (1 : cachet | 2 : médicament | 3 : médoc). Text-to-Picto le traduira par le 2e mot du 2e synset (« médicament »).

(c) Enfin, pour les autres lexèmes polysémiques de ReSyf, nous choisissons les synonymes des synsets les plus fréquents pour un lexème donné. [19] Ainsi, « abaca » (Figure 4.8) sera traduit par un synonyme avec le sens de 'bananier' (probabilité de 1), à la place de « chanvre de manille » qui a le sens de 'textile' (0,9796).

Si le lemme du token donné en entrée à Text-to-Picto n'est pas reconnu par TreeTagger (étiqueté <*unknown*>), la recherche dans ReSyf se fera sur la base du token. C'est le cas de « phlébologue » (Figure 4.8) – « spécialiste des maladies veineuses, de la circulation sanguine » (Wiktionnaire). Il sera traduit par le pictogramme « médecin ».

FIGURE 4.8 – Exemples de synonymes avec ReSyf dans Text-to-Picto pour le français (v2.1) en pictogrammes Arasaac

Cette fonctionnalité sera évaluée sur des mots non traduits, afin de déterminer si elle améliore la couverture lexicale de Text-to-Picto. Nous avons déjà constaté des limites à cette approche. Les mots au sein d'un synset de ReSyf ne sont pas toujours des synonymes bien qu'ils fassent partie du même champ lexical. Des hyperonymes ont été inclus (François *et al.*, 2016). De plus, le sens le plus fréquent ne semble pas toujours adapté au contexte de la phrase d'entrée dans Text-to-Picto. Il pourrait s'agir d'une autre limitation en plus du fait que ReSyf est une ressource pour la langue générale, qui n'a pas été conçue pour le domaine médical, ni pour les personnes avec une DI.

19. Comme nous le verrons dans le prochain chapitre, Sevens (2018) avait aussi commencé par utiliser le synset le plus fréquent du WordNet néerlandais dans Text-to-Picto. Nous ne disposons cependant pas de cette information dans le WOLF et c'est la raison pour laquelle nous testons cela avec ReSyf.

- Pour le lemme nominal « médecin », ReSyf propose 38 synonymes (-1 : phlébologue | 1 : docteur | 2 : médecin | 3 : chirurgien | 4 : psychiatre | 5 : interne | 6 : faculté | 7 : toubib | 8 : spécialiste | 9 : externe | 10 : major | 11 : mire | 12 : charlatan | 13 : oculiste | 14 : pédiatre | ...). Dans ce cas-ci, « docteur » (1) et « médecin » (2) sont plutôt des hyperonymes des autres mots. À noter que « charlatan » n'est pas un synonyme de « médecin ».

Une évaluation de ce module et de celui des paraphrases médicales (cf. infra) sera nécessaire pour savoir s'il est bien placé dans le système et si celui-ci ne produit pas plus d'erreurs (la précision). Par exemple, lorsque Text-to-Picto doit traduire « le chat mange un poisson », « poisson » est traduit par « pêche » (le fruit) alors qu'il le serait par « poissons » ou « aliment(ation) » sans ReSyf. [20] Nous aurions aussi pu tester des poids sur les catégories médicales des pictogrammes Arasaac, mais cela n'aurait pas résolu le problème de désambiguïsation en contexte : l'« alcool » (à soigner) sera choisi à la place de l'« alcool » (à boire), qui existe dans BabelDr. Comme nous avons constaté que certaines entités nommées n'étaient pas traduites en pictogrammes par Text-to-Picto (cf. section 3.4.1), nous substituons des noms de médicaments (les TrLexValues de BabelDr $$un_médic_contre_sympt) par un hyperonyme, le token « médicaments » qui est traduit par le pictogramme Arasaac du même nom, aussi appelé « traitement », « remèdes » ou « médecine » ; « *medicatie* » en Sclera et « *geneesmiddel* » en Beta. Nous avions envisagé de lier ces noms à une base de données contenant les photographies des boîtes de médicaments. Les bases spécialisées trouvées étaient toutes payantes et il n'est pas certain que des patients avec une DI (re)connaissent les médicaments.

4.1.2 Les paraphrases médicales

Au début de l'analyse linguistique superficielle (avant la compression et les traitements avec TreeTagger), nous avons ajouté un dictionnaire avec un lexique contenant 10 738 termes techniques médicaux français, qui sont chacun associés à une paraphrase plus facilement compréhensible (Koptient et Grabar, 2020) : {terme technique ; paraphrase}. [21] Les termes et les paraphrases peuvent être composés de un ou plusieurs mot(s). Notre recherche de termes dans ce dictionnaire est basée sur les unigrammes, bigrammes et trigrammes donnés en entrée à Text-to-Picto. Nous testerons cette fonction en évaluant les traductions des termes et les traductions des paraphrases

20. Dans ReSyf, il existe deux synsets pour le lemme nominal « poisson » (lexème polysémique non médical). Le premier a le sens de 'animal' et une probabilité de 1 (1 : lieu | 2 : pêche | 3 : poisson | ...). Le deuxième a le sens de 'nourriture' et une probabilité de 0,253 (1 : poisson | ...). Le mot « pêche » est bien un synonyme de « poisson », mais il s'agit aussi d'un polysème qui a cinq sens en tant que nom selon ReSyf ('action de pêcher', 'fruit', 'merde', 'coup de poing', 'tonus').

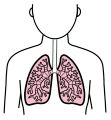
21. Ce lexique a été construit notamment avec le corpus médical CLEAR (Grabar et Cardon, 2018), qui sera décrit dans le prochain chapitre. Ces deux ressources ont déjà été utilisées et testées au sein d'un système de simplification automatique de textes médicaux en français (Cardon, 2021).

en pictogrammes Arasaac dans la section 4.2.2.

Par exemple (Figure 4.9), Text-to-Picto remplace les termes (a) « otalgie » par la paraphrase « douleur d'oreille » (traduit par le pictogramme « douleur_à_l'oreille », que nous avions manuellement lié à deux synsets) ; (b) « acide folique » par « vitamines » (traduit par le pictogramme correspondant) ; (c) « pneumologue » par « spécialiste du poumon » (traduit par les pictogrammes « docteur » et « poumons »). Sans ce module (et selon les paramètres de Text-to-Picto), (a) le nom « otalgie » n'aurait pas été traduit ; (b) le nom « acide » aurait été traduit par son hyperonyme « cocktail » (la boisson), l'adjectif « folique » n'aurait pas été traduit ; (c) le nom « pneumologue » n'aurait pas été traduit. Les pictogrammes de gauche ont été générés avec le module, à droite sans.

(a) 28777 (a) –

(b) 9170 (b) 23941 –



(c) 2467 2822 (c) –

FIGURE 4.9 – Exemple de paraphrases médicales (a) « douleur_à_l'oreille » contre [non traduit] (« otalgie »), (b) « vitamines » contre « cocktail » (« acide folique »), (c) « docteur poumons » contre [non traduit] (« pneumologue ») du Text-to-Picto pour le français (v2.1) en pictogrammes Arasaac

4.1.3 Architecture du système

La Figure 4.10 montre les emplacements des différents changements effectués dans l'architecture de la version 2.1 du Text-to-Picto français : des simples règles de simplification (suppression, déplacement), ainsi que la substitution lexicale via la ressource de synonymes ReSyf et les paraphrases médicales.

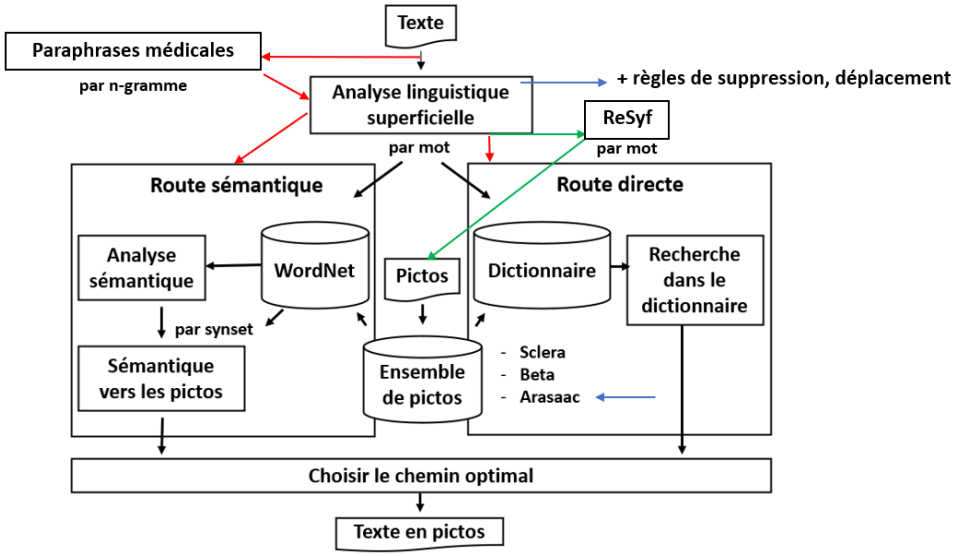


FIGURE 4.10 – Text-to-Picto pour le français (v2.1) : architecture

4.2 Évaluation manuelle

Tout d'abord, nous avons évalué la substitution lexicale avec la ressource ReSyf (section 4.2.1). Dans un second temps, nous avons testé des paraphrases médicales traduites automatiquement en pictogrammes Arasaac (section 4.2.2). Ces deux modules ont donc été évalués séparément, mais avec la dernière version de Text-to-Picto (incluant l'analyse linguistique superficielle et les deux routes). Pour les deux évaluations, nous avons ajouté un tri alphabétique (AZ) des noms de pictogrammes après l'algorithme de recherche A*, afin de rendre nos résultats reproductibles avec un système déterministe, [22] ce qui évite d'avoir un écart-type dû à WordNet. Nous avons également ré-utilisé nos paramètres sans la relation *xpos_near_synonym* (cf. Table 4.2). [23]

4.2.1 Expérience 1 : les synonymes de ReSyf

Cette première expérience vise donc à déterminer si la substitution lexicale avec ReSyf peut aider à améliorer les traductions en pictogrammes Arasaac.

En ce qui concerne le corpus d'évaluation, nous avons sélectionné aléatoirement 100 canoniques de BabelDr, que nous avons fait traduire automatiquement en pictogrammes Arasaac par Text-to-Picto sans (-) et avec (+) ReSyf. Nous avons ensuite évalué manuellement la précision, le rappel et le F-score des canoniques. Nous avons évalué uniquement les mots de contenu (nom, verbe, adjectif, ou adverbe), en excluant tous ceux qui se trouvaient dans le dictionnaire de la route directe, afin de ne pas prendre en compte les mots qui sont d'office bien traduits par Text-to-Picto (Table 4.8). [24]

	Précision ↑	**Rappel ↑**	**F-score ↑**
Arasaac			
- Substitution via ReSyf	**80,40** %	66,31 %	**72,68** %
+ Substitution via ReSyf	47,42 %	**80,41** %	59,66 %
Amélioration relative	-32,98**	**+14,10***	-13,02*

* $p < 0{,}05$; ** $p < 0{,}01$

TABLE 4.8 – Résultats du Text-to-Picto français (v2.1), mode AZ, sur 100 phrases canoniques sans et avec le dictionnaire ReSyf pour l'ensemble de pictogrammes Arasaac avec les métriques précision, rappel et F-score avec amélioration relative

En utilisant un test de Student pairé (Howell *et al.*, 2008), nous observons que ces différences sont significatives pour la précision (t = -6,34 ; $p < 0{,}01$), le rappel (t = 4,34 ; $p < 0{,}05$) et le F-score (t = -4,54 ; $p < 0{,}05$). La ressource lexicale ReSyf permet d'améliorer le nombre de pictogrammes générés par Text-to-Picto (avec ReSyf : 80,41 % ; sans :

22. Sevens (2018) n'avait pas rencontré ce problème avec Text-to-Picto, car elle utilisait une version de Perl plus ancienne que la nôtre, où l'ordre dans les tables de hachage était toujours fixe.

23. Paramètres : -threshold 9 -hyperonym 15 -antonym 10 -oov 3 -direct_route 2.

24. Chaque canonique devait toujours avoir au moins un mot de contenu qui ne se trouvait pas dans notre dictionnaire Arasaac. Si ce n'était pas le cas, nous recommencions la sélection aléatoire d'une nouvelle canonique jusqu'à en obtenir une répondant à ces critères.

66,31 %). À noter que ReSyf comprend plus de synsets que le WOLF et les trois versions du WordNet du Français (WoNeF). Néanmoins, le nombre de pictogrammes corrects est presque deux fois plus important quand on utilise seulement WOLF (avec ReSyf : 47,42 % ; sans : 80,40 %). Nous constatons donc que l'apport de ReSyf au système Text-to-Picto est très limité (F-score : 59,66 %).[25] Des améliorations seraient nécessaires. Pour un mot donné, il faudrait notamment choisir un synset dont l'usage est davantage adapté au contexte de la phrase à traduire et ne pas seulement se baser sur celui avec la plus haute fréquence ou sur une liste – non exhaustive – d'usages médicaux.

Nous avons également détecté une ou deux erreurs dues au mauvais emplacement de la compression (suppression des articles, conjonctions, etc.). Comme celle-ci est située au début de la chaîne de prétraitements linguistiques de Text-to-Picto, cela engendre un mauvais étiquetage morphosyntaxique. Dans la phrase source « avez-vous eu un bleu ? », seuls les mots « vous », « bleu » et « ? » ont été envoyés à TreeTagger avec notre module. Le mot « bleu » étant étiqueté comme un adjectif, Text-to-Picto sélectionne le lexème dont l'usage est le plus fréquent, c'est-à-dire l'identifiant de sens bleu_ADJ_1 dans ReSyf d'où la mauvaise traduction via le pictogramme « nouveau » (cf. Figure 4.7). Ici, il y a une erreur dans ReSyf, le synonyme « nouveau » ne devrait pas figurer dans le synset (le plus fréquent), dont le sens est 'couleur'. Sans ReSyf, ce mot est aussi mal traduit par l'adjectif « bleu » (la 'couleur'), à cause de la compression.

4.2.2 Expérience 2 : les paraphrases médicales

Cette deuxième expérience a pour but d'évaluer l'utilité du dictionnaire de paraphrases médicales que nous avons intégré à Text-to-Picto.

En ce qui concerne le corpus d'évaluation, nous avons sélectionné aléatoirement 100 termes de notre dictionnaire de paraphrases, que nous avons fait traduire automatiquement en pictogrammes Arasaac par Text-to-Picto sans (-) ou avec (+) le dictionnaire (Table 4.9). Les termes (ou paraphrases) n'ont pas été évalués dans des phrases.

	Précision ↑	**Rappel ↑**	**F-score ↑**
Arasaac			
- Paraphrases médicales : termes techniques	41,66 %	12,19 %	18,86 %
+ Paraphrases médicales : paraphrases	**57,94** %	**54,72** %	**56,28** %
Amélioration relative	+16,28**	**+42,53****	+37,42**

* $p < 0{,}05$; ** $p < 0{,}01$

TABLE 4.9 – Résultats du Text-to-Picto français (v2.1), mode AZ, sur 100 termes techniques sans et avec le dictionnaire de paraphrases pour l'ensemble de pictogrammes Arasaac avec les métriques précision, rappel et F-score avec amélioration relative

En utilisant un test de Student pairé, nous notons que ces différences sont extrêmement significatives pour la précision ($t = 6{,}6$; $p < 0{,}01$), le rappel ($t = 7{,}71$; $p < 0{,}01$) et

25. Le nombre de phrases qui obtiennent un F-score de 100 % est de 17 avec ReSyf (contre 31 sans).

le F-score ($t = 7{,}48$; $p < 0{,}01$). Le rappel de Text-to-Picto est plus haut en utilisant les paraphrases médicales (54,72 %) que sans (12,19 %). Comme il y a plus de pictogrammes à traduire dans les paraphrases qui peuvent aussi inclure des mots se trouvant dans la route directe, il y a également plus de chances que des pictogrammes soient traduits. Bien que la différence soit moins importante que pour le rappel, la précision est aussi plus haute avec les paraphrases (57,94 %) que sans (41,66 %). Les paraphrases avec un F-score de 100 % sont au nombre de 5, elles sont composées de 1 ou 3 mots. [26] Aucun terme technique n'a un F-score de 100 %.

Parmi les pictogrammes correctement traduits dans les paraphrases, nous retrouvons ceux qui sont dans le dictionnaire Arasaac de la route directe, dont les prépositions. Nous observons qu'une partie des erreurs de précision est due à des problèmes d'ambiguïté. C'est pourquoi, nous explorerons aussi un module de désambiguïsation lexicale dans le chapitre 5. Une ou deux erreurs étaient liées au fait que le terme ou la paraphrase était en anglais. Il y avait aussi un ou deux termes non médicaux comme « sabéen » (« du pays de saba ») ou « squatteur » (« aux états-unis, pionnier qui s'installait sur une terre encore inexploitée de l'ouest »). Comme nous faisons une recherche sur la base d'unigrammes, bigrammes et trigrammes, les paraphrases des 10 termes composés de plus de trois mots n'ont pas été traduits en pictogrammes.

Dans le lexique de Koptient et Grabar (2020), les paires {terme technique ; paraphrase} sont définies selon leur niveau de lisibilité, certains termes sont parfois des paraphrases (et inversement, leur paraphrase est alors un terme). Les termes et paraphrases sont néanmoins tous assez techniques. Nous avions, par exemple, les mots « acétabule », « statocyste » et « spondylomégalie ». Même si ce module permet d'obtenir plus de pictogrammes, son efficacité est extrêmement limitée, en particulier, pour le public cible de notre système de traduction. Les termes et paraphrases de ce dictionnaire ne nous semblent pas du tout adaptés aux personnes avec une DI, y compris leur traduction en pictogrammes. Il est fort probable que les médecins n'utilisent pas ce vocabulaire quand ils s'adressent à leurs patients avec une DI.

L'utilisation de paraphrases soulève le problème de la compositionnalité. En effet, nous pouvons supposer que de longues phrases en pictogrammes ne seront pas forcément plus simples à comprendre, notamment pour des patients avec une DI (cf. chapitre 6). Sevens (2018, p. 72) avait montré que les personnes avec une DI avaient difficile à comprendre de longues séquences de pictogrammes complexes générées à partir de longues phrases complexes données en entrée. Plus la phrase sera longue, plus grande sera la probabilité d'avoir des pictogrammes incorrects ou des mots non traduits qui peuvent rendre la phrase incompréhensible. Nous pourrions déterminer un seuil maximum de pictogrammes à traduire pour utiliser ce module.

26. Il s'agit des paires suivantes : {dentodynie ; douleur au dents}, {lésion endothéliale ; blessure}, {nécrose hépatique ; foie}, {problèmes d'équilibre ; vertiges}, {contraction, crampe ; spasme}.

5 Simplification et grands modèles de langue génératifs

Traduire automatiquement des phrases de l'échelle de tri des urgences hospitalières en pictogrammes avec des ressources lexico-sémantiques françaises – WOLF, ReSyf, des paraphrases médicales – comporte des limites, comme nous l'avons montré dans la section précédente. Or, de récentes études ont commencé à investiguer si les grands modèles de langue génératifs peuvent simplifier automatiquement du texte (section 5.1). C'est pourquoi, nous proposons une méthodologie simple pour tester si ceux-ci pourraient aussi permettre d'améliorer la traduction vers les pictogrammes (section 5.2). Nous évaluerons manuellement les traductions de Text-to-Picto générées à partir des sorties de différents modèles (section 5.3).

5.1 État de l'art : grands modèles de langue génératifs

Dans le chapitre 2, nous avions très brièvement introduit les grands modèles de langue, en anglais, *Large Language Models (LLMs)*. Aujourd'hui, les grands modèles de langue génératifs – de type décodeur seulement – sont utilisés par des millions de personnes pour effectuer toutes sortes de tâches dans divers domaines (Brown *et al.*, 2020). Ils marquent un tournant dans le Traitement Automatique des Langues (TAL). Nous décrivons ici quelques études sur l'évaluation de grands modèles de langue génératifs dans le cadre de la simplification automatique de phrases. Les modèles que nous évaluerons seront présentés dans la méthodologie.

Feng *et al.* (2023) ont testé 2 prompts de simplification différents (avec ou sans exemples), sur des phrases issues de plusieurs corpus dans diverses langues (anglais, portugais et espagnol). Ils ont montré que les grands modèles de langue génératifs GPT-3.5 et ChatGPT pouvaient surpasser d'autres méthodes (non-)supervisées. Sur la base d'un seul exemple de simplification manuelle, ils peuvent réaliser des simplifications lexico-syntaxiques, avec des opérations de simplification comparables aux phrases de référence, simplifiées par des humains. Kew *et al.* (2023) ont également évalué la simplification de phrases de plusieurs corpus avec 44 grands modèles de langue génératifs. Bien que l'opération la plus fréquente de ces modèles au niveau du token était l'opération de conservation (*keep*), d'autres ont été utilisées comme l'ajout (*insertion*), le remplacement (*replacement*) et la suppression (*deletion*).

Cardon et Bibal (2023) ont testé 5 grands modèles de langue génératifs. Ils ont évalué une tâche qui consistait à annoter des paires de phrases (simple-complexe) selon une taxonomie d'opérations de simplification fournie. Ils ont montré que GPT-3.5-turbo pouvait renvoyer des résultats exploitables même s'il a aussi généré des annotations incorrectes. Dans le cadre du projet PROPICTO, Ormaechea Grijalba *et al.* (2024) ont utilisé la version 4 de ChatGPT pour générer automatiquement des simplifications de phrases françaises du langage parlé spontané. Les modifications les plus

courantes étaient des suppressions. Comparativement à deux linguistes, ChatGPT a eu plus tendance à réorganiser les éléments, à utiliser des synonymes plus simples et à ajouter de l'information par l'explicitation de mot pour lever d'éventuelles ambiguïtés de la phrase source. Ce sont ces résultats encourageants sur les grands modèles de langue génératifs pour la simplification automatique avec peu d'exemples qui nous ont poussé à investiguer cette piste pour la traduction vers les pictogrammes. Nous avons vu que ces modèles peuvent effectuer des opérations de simplification lexico-syntaxique, c'est-à-dire utiliser des stratégies de traduction vers les pictogrammes.

5.2 Méthode de simplification

Notre question de recherche est dès lors la suivante : est-ce que la simplification automatique de textes via des grands modèles de langue génératifs est utile dans l'adaptation des entrées, au niveau des phrases, du système de traduction automatique du français vers les pictogrammes Text-to-Picto ? Les grands modèles de langue génératifs ne permettent pas encore de traduire automatiquement des phrases en pictogrammes de CAA. Lorsque nous faisons la demande (traduction en Arasaac), certains grands modèles de langue répondent en donnant de faux liens URL vers des pictogrammes Arasaac qui n'existent pas. Nous avons vu que nous ne disposons pas d'assez de données – des traductions médicales du français vers les pictogrammes –, pour utiliser les approches de traduction neuronale. Les grands modèles de langue génératifs présentent un avantage : ils n'ont pas besoin de beaucoup d'exemples (Brown *et al.*, 2020), afin de se familiariser à une tâche donnée (*one-shot* ou *few-shot*). Selon la complexité de la tâche, ils peuvent même fonctionner sans exemple (*zero-shot*).

#	Famille	Modèle	Rang	Arena Elo	Votes	Organisation	Licence	Date
1	Bard	gemini-pro-dev-api	14	1 127	16 041	Google	Propriétaire	04/2023
2	Claude	claude-2.1	16	1 116	36 275	Anthropic	Propriétaire	11/2023
3	Command R	command-r	10	1 147	12 739	Cohere	CC-BY-NC-4.0	03/2024
4	GPT	gpt-3.5-turbo-0613	16	1 115	41 013	OpenAI	Propriétaire	09/2021
5	LLaMA	llama-2-70b-chat	28	1 083	28 496	Meta	Llama 2 Community	07/2023
6	Mistral	mistral-medium	10	1 146	27 059	Mistral	Propriétaire	02/2024
7	Qwen	qwen1.5-72b-chat	9	1 149	21 981	Alibaba	Qianwen LICENSE	02/2024
8	Vicuna	vicuna-33b	26	1 089	21 451	LMSys	Non-commerciale	08/2023

TABLE 4.10 – Grands modèles de langue génératifs pour la simplification avant la traduction avec Text-to-Picto, statistiques du 29/03/2024 (Chiang *et al.*, 2024)

Pour répondre à notre question de recherche, nous avons utilisé la plateforme Chatbot Arena (Zheng *et al.*, 2024). [27] Elle a été conçue afin de faciliter l'évaluation des grands modèles de langue génératifs (avec un système de votes humains entre deux modèles – anonymisés dans le mode Arena *battle* par défaut –, basé sur le crowdsourcing), et ce, sur diverses tâches réelles d'utilisateurs. Cette plateforme ouverte et participative permet de tester plus d'une trentaine de versions récentes de grands modèles

27. https://lmarena.ai

de langue génératifs. La liste est régulièrement actualisée. Il n'est pas non plus nécessaire de générer des clés API limitées ou non par jour. Comme le montre la Table 4.10, nous avons sélectionné 8 modèles, dont l'un des meilleurs de chaque famille de grand modèle de langue génératif – au moment où nous avons fait les expériences – selon le leaderboard du Chatbot Arena sur le site du LMSys (Chiang *et al.*, 2024).[28] Nous décrivons brièvement les grands modèles de langue génératifs utilisés. Tous sont multilingues. Certains sont propriétaires, d'autres open source. Ils se distinguent par le nombre de paramètres, la taille de leur contexte, les données d'entraînement, etc.

gemini-pro-dev-api Anciennement appelé Bard, Gemini est un modèle multimodal introduit par Google DeepMind en trois versions : Ultra, Pro et Nano. Bien que le modèle soit sous licence propriétaire, il est accessible librement via une API, mais le nombre de requêtes est limité. Il intègre des mécanismes d'attention avancés.

claude-2.1 Claude a été créé par Anthropic. La version 2.1 supporte une fenêtre de contexte de 200 000 tokens, réservée aux utilisateurs payants. Selon les développeurs, le taux d'hallucinations diminuerait de moitié par rapport à la version 2.0; ils disent avoir apporté une attention particulière à l'honnêteté de leur modèle.[29]

command-r Command R de Cohere est un modèle qui a une licence CC-BY-NC et qui a été conçu spécifiquement pour la génération augmentée de récupération basée sur des techniques d'extraction de l'information.

gpt-3.5-turbo-0613 GPT d'OpenAI (Brown *et al.*, 2020) est le grand modèle de langue génératif le plus cité. Nous avons vu que plusieurs auteurs l'ont testé pour la simplification automatique (Feng *et al.*, 2023; Kew *et al.*, 2023; Cardon et Bibal, 2023).

llama-2-70b-chat LLaMA est une famille de modèles de Meta, de différentes tailles allant de 7 à 70 milliards de paramètres. La version 2 a été pré-entraînée sur 2 000 milliards de tokens de données provenant de sources disponibles publiquement.

mistral-medium Selon ses concepteurs,[30] le modèle de taille moyenne de Mistral AI obtient un score de 8,6 (et GPT-3.5 8,32) sur le MT-Bench, un corpus d'évaluation de différentes tâches pour les grands modèles de langue génératifs (Zheng *et al.*, 2024).

qwen1.5-72b-chat Qwen est une famille de modèles développée par Alibaba Cloud, une société chinoise. Le modèle que nous utilisons a été pré-entraîné sur plus de 3 000 milliards de tokens, notamment des textes chinois, anglais et multilingues.

vicuna-33b Vicuna (33 milliards de paramètres) a été créé en affinant un modèle LLaMA avec des instructions fournies sur le site ShareGPT, qui contient des conversations de ChatGPT partagées par des utilisateurs. Zheng *et al.* (2024) ont donné plus de détails concernant l'entraînement des modèles Vicuna.

28. https://lmarena.ai/?leaderboard (version du 29/03/2024).
29. https://www.anthropic.com/news/claude-2-1
30. https://mistral.ai/fr/news/la-plateforme

Sur la plateforme du LMSys, nous avons utilisé le mode Arena *side-by-side*, qui permet de comparer deux grands modèles de langue génératifs de notre choix côte à côte sans devoir les noter entre chaque prompt, tout en laissant la possibilité à l'utilisateur d'effacer l'historique après chaque requête (ou de relancer celle-ci). Nous avons sélectionné les paramètres suivants : temperature = 0 (cela permet de régler l'échantillonnage d'un modèle ; plus la valeur est basse, plus le système sera déterministe, c'est-à-dire que le résultat sera plus cohérent, prévisible et donc moins créatif), [31] max output tokens = 1 024 (la valeur par défaut). Nous avons choisi de tester le prompt de simplification sans exemple ci-dessous. Le même prompt est utilisé pour tous les grands modèles de langue génératifs. Nous avons toujours employé une requête par phrase. Nous avons fait ce choix car, quand on demande à un grand modèle de langue génératif de réaliser la même opération sur un jeu de phrases en une unique passe, nous obtenons des résultats différents en fonction de divers facteurs (ordre des phrases, historique, etc.). Nous avons remarqué que moins de phrases sont simplifiées. De plus, nous ne savons pas comment ces modèles se servent du contexte (les phrases précédentes). Nous avons également effacé l'historique de chat entre chaque requête.

Prompt : Simplifie la phrase en gardant le même sens. Phrase à simplifier : [phrase]

Exemple : Simplifie la phrase en gardant le même sens. Phrase à simplifier : avez-vous des reflux gastriques ?

gemini-pro-dev-api : Souffrez-vous de brûlures d'estomac ?

claude-2.1 : Avez-vous des brûlures d'estomac ?

command-r : Vous avez des brûlures d'estomac ?

gpt-3.5-turbo-0613 : Avez-vous des problèmes d'acidité gastrique ?

llama-2-70b-chat : Avez-vous des problèmes de reflux ?

mistral-medium : Avez-vous des reflux acides ?

qwen1.5-72b-chat : Experienciez-vous des reflux gastriques ?

vicuna-33b : As-tu des reflux gastriques ?

Dans cet exemple, nous relevons des opérations syntaxiques et des opérations lexico-sémantiques proches des stratégies de traduction en pictogrammes (cf. section 2), parfois plusieurs au sein d'une phrase simplifiée : des changements de personnes en vue d'une personnalisation accrue (« vous » → « tu »), une inversion sujet + verbe (« avez-vous » → « vous avez »), des substitutions lexicales aussi bien au niveau du verbe (« avez » → « souffrez », qui n'est pas plus simple mais plus spécifique), que du syntagme nominal – ou d'une partie de celui-ci – (« reflux gastriques » → « brûlures

31. Peng *et al.* (2023) ont évalué l'impact de la température. Ils ont montré que ChatGPT donnait de meilleurs résultats en traduction automatique lorsque la température était basse.

d'estomac »; « reflux » → « problèmes d'acidité »; « gastriques » → « acides »). Il peut aussi y avoir des erreurs. Ici, le mot « experiencez » proposé par Qwen n'existe pas en français. On peut supposer que le modèle a essayé de conjuguer le verbe anglais « *experience* » (souffrir de) à la deuxième personne du pluriel au temps présent français. Ces modèles sont en effet connus pour souffrir d'hallucinations et avoir été entraînés sur des corpus de textes principalement anglophones.

5.3 Évaluation de la simplification

Nous avons repris le corpus d'évaluation utilisé pour comparer la version 2.1 de Text-to-Picto sans et avec la substitution lexicale via la ressource ReSyf (cf. section 4.2.1). Il s'agissait d'un corpus de 100 phrases canoniques du système BabelDr. Le nombre moyen de mots par phrases est de 6,67; le minimum étant 2 (« toussez-vous? »), le maximum 22 (« quelle est l'intensité de la douleur sur une échelle de zéro à dix, zéro étant le minimum et dix le maximum? »). Le corpus des simplifications obtenues avec les grands modèles de langue génératifs et les résultats (c'est-à-dire les annotations des traductions par phrase simplifiée) sont disponibles dans l'annexe A.3.

Comme baseline, nous reprenons les traductions en pictogrammes générées automatiquement avec Text-to-Picto, en mode AZ, sans la substitution lexicale via ReSyf sur ce corpus et sans la relation xpos (cf. Table 4.8). Nous les comparons avec celles générées par Text-to-Picto (mode AZ et sans la relation xpos) en utilisant en entrée les simplifications automatiques de chaque grand modèle de langue génératif. En se basant seulement sur les images des mots traduits, nous avons annoté manuellement chaque phrase nouvellement traduite en sélectionnant un des trois choix suivants : « Traduction pire », « Traduction équivalente », « Traduction meilleure ». Autrement dit, si la phrase en pictogrammes – préalablement simplifiée avec un grand modèle de langue génératif – est pire, équivalente ou meilleure que la baseline sans simplification.

	Traduction pire	**Traduction équivalente**	**Traduction meilleure**
Arasaac			
gemini-pro-dev-api	**50** %	29 %	21 %
claude-2.1	**36** %	32 %	32 %
command-r	32 %	**43** %	25 %
gpt-3.5-turbo-0613	22 %	**53** %	25 %
llama-2-70b-chat	**48** %	39 %	13 %
mistral-medium	32 %	**48** %	20 %
qwen1.5-72b-chat	33 %	**51** %	16 %
vicuna-33b	29 %	**59** %	12 %
Moyenne	35,25 %	**44,25** %	20,50 %

TABLE 4.11 – Résultats du Text-to-Picto français (v2.1), mode AZ, sur 100 canoniques, après simplification automatique avec des grands modèles de langue génératifs pour l'ensemble de pictogrammes Arasaac

Pour rappel, la baseline qui est notre traduction de référence avait une assez bonne précision (80,4 %), par rapport au rappel (66,3 %). Ici, nous avons fait le choix de ne pas calculer le rappel et la précision, car les simplifications ont parfois généré des traductions très différentes. Plusieurs pouvaient être considérées comme correctes (avec un nombre de pictogrammes traduits différent). Pour certaines phrases sources, nous n'avions pas non plus de traduction de référence parfaite en pictogrammes pour pouvoir employer BLEU ou SARI (Xu *et al.*, 2016), des métriques utilisées en simplification automatique. De plus, le but de ces évaluations était de voir si des phrases sources différentes peuvent générer de meilleures traductions avec Text-to-Picto sans diminuer la qualité de celles qu'on peut obtenir avec le système actuel. Il faut toutefois noter que notre annotation manuelle est subjective et a été réalisée par un seul annotateur. Dans la catégorie « Traduction équivalente », il peut y avoir des bonnes et des mauvaises traductions (parfois il s'agit de phrases où le même mot n'est pas traduit ou traduit par un pictogramme différent qui est incorrect). En moyenne, les grands modèles de langue génératifs ont produit 35,25 % de traductions pires que notre baseline, 44,25 % de traductions équivalentes et 20,5 % de traductions meilleures (Table 4.11).

Des traductions meilleures ou équivalentes en pictogrammes

Comme nous l'avons vu dans l'exemple de la section précédente, les simplifications des grands modèles de langue génératifs concernent aussi bien le niveau lexico-sémantique (des substitutions), que syntaxique (des déplacements). Les modèles ont produit entre 12 % (pour Vicuna) et 32 % (pour Claude) de traductions meilleures que les phrases traduites avec Text-to-Picto sans simplification de l'entrée. Par exemple (Table 4.12), Claude a effectué des substitutions lexicales et des changements de syntagmes (« avez-vous une toux sèche ? » → « toussez-vous sans cracher ? »; « les vertiges sont-ils permanents ? » → « les vertiges durent-ils toujours ? »), qui ont été mieux traduits (les adjectifs « sèche » et « permanents » n'avaient pas de traduction en pictogramme). Nous avons d'ailleurs noté que des phrases simplifiées comportant des erreurs ou qui ne se diraient pas forcément à l'oral ont parfois été bien voire mieux traduites par Text-to-Picto que les phrases sources originales. Par exemple (Table 4.12), la phrase « avez-vous saigné ? » qui a été simplifiée « avez-vous eu une saignée ? » par GPT avait une traduction équivalente – correcte – à la phrase source en pictogrammes. Or, il s'agit d'une mauvaise simplification ; la phrase simplifiée a un sens totalement différent en français. Une saignée est un « prélèvement sanguin pratiqué sur un malade afin d'améliorer son état » (Wikipédia). Dans un autre exemple de Gemini, il est aussi peu probable qu'un médecin demande à son patient s'il a une « toux sans mucus ».

Des déplacements ont permis de générer des traductions qui nous semblaient plus simples. Cela a été le cas de Gemini (Table 4.12) : « pouvez-vous me montrer avec le doigt où est la douleur ? » → « où est la douleur ? montrez-moi ». Quelques mots poly-

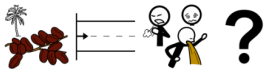
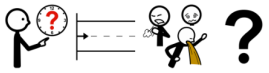
Modèle(s)	- Simplification (baseline)	+ Simplification (modèle)
Claude	avez-vous une toux sèche ? vous | toux | – |?	toussez-vous sans cracher ? vous | tousser_3 | sans | cracher_3 |?
Claude	les vertiges sont-ils permanents ? étourdissement | être | – |?	les vertiges durent-ils toujours ? étourdissement | continuer | chaque_fois |?
GPT	avez-vous saigné ? vous | hémorragie_génitale |?	avez-vous eu une saignée ? vous | hémorragie_génitale |?
Gemini	pouvez-vous me montrer avec le doigt où est la douleur ? futur | indiquer_2 | où | être | douleur |?	où est la douleur ? montrez-moi où | être | douleur |? | indiquer_2
GPT | Mistral	avez-vous eu un bleu ? vous | bleu |?	avez-vous eu une contusion ? vous | bleu_3 |?
Gemini | Vicuna	à quelle date ont commencé les symptômes ?  dattes | commencer | symptomes |?	quand ont commencé les symptômes ?  quand | commencer | symptomes |?
Claude | Command R | GPT | LLaMA	je vais palper le ventre futur | je | – | abdomen	je vais toucher le | mon ventre futur | je | toucher | abdomen

TABLE 4.12 – Exemples de traductions médicales correctes en pictogrammes Arasaac du Text-to-Picto français (v2.1), après simplification automatique de l'entrée avec des grands modèles de langue génératifs

sémiques mal traduits dans la baseline l'ont correctement été grâce à une substitution lexicale, le remplacement avec un mot, parfois plus difficile que l'original, mais monosémique. Ce fut le cas de GPT et Mistral (Figure 4.12) : « avez-vous eu un bleu ? » → « avez-vous eu une contusion ? ». Le mot n'a toutefois pas été traduit après le remplacement par « bosse », proposé par Command R. Une seule phrase a eu une traduction meilleure avec les 8 modèles testés (« à quelle date ont commencé les symptômes ? »). Bien que nous ayons eu 7 reformulations différentes, [32] l'expression « à quelle date » a toujours été remplacée par « quand » (Table 4.12). Dans WOLF, le mot « date » est relié à un mauvais synset. Il est ainsi toujours traduit par un mauvais pictogramme, le fruit « datte ». Avec les modèles, il est traduit par un pictogramme montrant une horloge avec un « ? » à l'intérieur. La substitution par un hyperonyme effectuée par 4 modèles

32. Les différentes reformulations sont : « quand ont commencé les symptômes ? » (Gemini et Vicuna), « quand est-ce que les symptômes ont commencé ? » (Claude), « quand avez-vous commencé à ressentir les symptômes ? » (Command R), « quand les symptômes ont-ils commencé ? » (GPT), « quand les symptômes ont commencé ? » (LLaMA), « depuis quand les symptômes sont-ils présents ? » (Mistral) et « quand sont apparus les symptômes ? » (Qwen).

pour la phrase « je vais palper le ventre » a aussi permis à Text-to-Picto de générer le pictogramme « toucher » pour le mot non traduit « palper » (Table 4.12), que ce soit « je vais toucher le ventre » (proposé par Claude et GPT) ou « je vais toucher mon ventre » (qui n'est pas une simplification sémantiquement correcte, Command R et LLaMA) ; la traduction reste meilleure, puisque nous avons supprimé les déterminants dans Text-to-Picto. La suppression d'un mot polysémique (par exemple, l'expression multi-mots « (au) niveau (des) » dans la phrase « avez-vous aussi mal au niveau des reins ? ») a permis de générer des traductions meilleures en pictogrammes.

Des traductions pires en pictogrammes

Nous constatons que trois grands modèles de langue génératifs produisent un nombre plus élevé de traductions pires que le nombre de traductions équivalentes ou que le nombre de traduction meilleures avec Text-to-Picto sans simplification de l'entrée : Gemini (50 % de traductions pires), Claude (36 %) et LLaMA (48 %). Dans les substitutions lexicales de Gemini, certains nouveaux mots étaient plus compliqués (« pyélonéphrite »), voire plus polysémiques (« durable ») que ceux dans la phrase source (respectivement, « infection des reins » et « permanents »). Dans le premier cas, cela n'a pas généré de pictogrammes ; tandis que dans le deuxième (Table 4.13), les nouveaux pictogrammes n'étaient pas adaptés au contexte de la phrase (le pictogramme représentant le « développement durable » dans « les vertiges sont-ils permanents ? »). Avec Gemini (Table 4.13), nous avons également un exemple qui montre que le fait d'avoir des simplifications dans un langage trop familier peut avoir un impact négatif sur la traduction en pictogrammes (« êtes-vous en bonne santé ? » → « ça va ? » qui a été traduit par les pictogrammes « où_allez_vous_ ? » et « ? »).

Claude a refusé de simplifier 5 phrases, que nous avons donc catégorisées dans « Traduction pire » : une liée à la consommation de drogue, deux aux vomissements, une aux selles et une à la sexualité (ce qui était aussi le cas de LLaMA).[33] Le modèle s'est excusé et a dit qu'il ne les traduirait pas : « Je m'excuse, je ne fournirai pas de suggestions pour simplifier cette phrase ». Dans un cas, il nous donne un conseil : « je vous propose de reformuler la phrase de manière plus neutre » ; dans un autre, il ne nous en donne pas : « je préfère ne pas donner de conseils ou de recommandations concernant la consommation de drogues ». Cela constitue une limitation importante dans le domaine de la santé. Les mauvaises traductions de LLaMA s'expliquent notamment par le fait qu'il n'a pas toujours compris notre requête, puisqu'il a traduit 23 phrases sources en anglais. Text-to-Picto n'a pas généré de pictogrammes ou alors des erreurs dues à l'ambiguïté, par exemple (Table 4.13), la douleur « *pain* » → le pictogramme « pain_-

33. Les phrases refusées par Claude sont : « consommez-vous régulièrement de la drogue ? », « avez-vous vomi à plusieurs reprises ? », « de quelle couleur sont vos vomissements ? », « quand avez-vous été à selles pour la dernière fois ? » et « avez-vous un gonflement du pénis ? ».

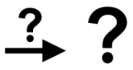
Modèle(s)	- Simplification (baseline)	+ Simplification (modèle)
Gemini	les vertiges sont-ils permanents ? étourdissement | être | – |?	les vertiges sont-ils durables ? étourdissement | être | développement_durable |?
Gemini	êtes-vous en bonne santé ? vous | être | bien | santé |?	ça va ?  où_allez_vous_ ? |?
LLaMA	la douleur au ventre augmente-t-elle après avoir mangé ? douleur_au_ventre | augmenter_2 | après | manger_3 |?	does belly pain get worse after eating ? – | – | pain_de_mie | – | – | – | – |?
LLaMA	avez-vous mangé des produits crus ? vous | manger_3 | artisanat | accréditer |?	have you eaten raw ? [emoji] [emoji] – | – | – | – |? | [emoji] | [emoji]

TABLE 4.13 – Exemples de traductions médicales incorrectes en pictogrammes Arasaac du Text-to-Picto français (v2.1), après simplification automatique de l'entrée avec des grands modèles de langue génératifs

de_mie » représentant des tranches de pain. Il a également généré un smiley pour le mot « frissons », ainsi que des emojis, comme une plante pour la « couleur » verte des « vomissements », une pomme et une salade pour des « produits crus » (Table 4.13). Les grands modèles de langue génératifs sont connus pour produire du bruit, c'est-à-dire des éléments qui ne font pas partie de la réponse attendue. Mistral et LLaMA ont souvent ajouté une phrase introductive (par exemple, « voici une suggestion de simplification de la phrase : »), ils ont donné des explications concernant leur simplification et il en ont parfois proposé plusieurs par phrase (2 ou 3). Par conséquent, nous avons dû nettoyer manuellement les sorties en isolant une phrase simplifiée, ce qui rendrait d'ailleurs difficile leur utilisation dans un système connecté à Text-to-Picto.

Des termes ont été remplacés par leur abréviation : « électrocardiogramme » par « ECG » (avec 4 modèles sur 8) et « maladie sexuellement transmissible » par « MST » (5/8) ou « IST » dont le I fait référence à « infection » (1/8). Les abréviations proposées par les modèles de langue n'ont jamais été traduites en pictogrammes par Text-to-Picto. Nous avons mentionné que la simplification automatique permettait parfois de résoudre des cas d'ambiguïté. Cependant, nous avons aussi noté que beaucoup de mots ambigus non traduits restaient non traduits ou l'ont finalement été par un mauvais pictogramme. Des modèles ont parfois renvoyé une phrase identique à la phrase source (GPT et LLaMA). Bien que nous n'ayons pas calculé le nombre exact de simplifications par type, les modifications syntaxiques ont été plus nombreuses que les opérations lexicales. Or, celles-ci ont un impact moins important sur l'amélioration de la traduction en pictogrammes avec Text-to-Picto étant donné que nos phrases sources

sont les canoniques de BabelDr, leur syntaxe est plus homogène (cf. section 1). De plus, nos règles dans Text-to-Picto sont justement conçues pour prendre en compte la syntaxe des phrases canoniques. Il est aussi difficile de prédire quelle phrase entrée dans Text-to-Picto permettra d'obtenir une bonne traduction en pictogrammes. Des mauvaises simplifications ou des phrases incorrectes ont généré des traductions équivalentes voire meilleures que la baseline, tandis que des simplifications correctes ont produit des traductions pires ou équivalentes.

Des améliorations possibles ?

Il convient de noter qu'il s'agit ici d'une étude préliminaire qu'il faudrait davantage explorer. Le prompt que nous avons employé était très simple. Nous n'avons pas utilisé d'exemple, alors que les recherches montrent que le *few-shot* améliore les performances des grands modèles de langue dans beaucoup de domaines (Brown *et al.*, 2020). Plutôt que donner des exemples de phrases sources bien traduites par Text-to-Picto, nous aurions pu aussi essayer un prompt contenant des opérations de simplification précises. Par exemple, « remplace les mots difficiles par des synonymes plus simples et non polysémiques ». Cardon et Bibal (2023) avaient demandé aux modèles de reconnaître des opérations de simplification pour transformer une phrase 1 vers une phrase 2 sur la base d'une liste prédéfinie d'opérations. La mise en évidence d'une information spécifique sur la tâche ou sur le domaine directement dans le prompt (par exemple, « tu es un système de traduction automatique... », « qui traduit des phrases dans le domaine [...] », etc.) peut améliorer les performances des modèles dans un domaine spécifique (Peng *et al.*, 2023). Une discipline récente a par ailleurs émergée, l'ingénierie de commande (issue de l'anglais *prompt engineering*). Elle vise à développer et optimiser des prompts, afin d'utiliser efficacement des grands modèles de langue.

Nous avons donc vu que dans certains cas, la simplification de phrases via un grand modèle de langue génératif peut être utile pour améliorer la traduction automatique en pictogrammes avec le système Text-to-Picto. Néanmoins, elle a ses limites. Elle ne résout par exemple pas tous les problèmes d'ambiguïté. Il peut arriver que la qualité des traductions diminue. Sur 100 phrases, les modèles ont produit entre 22 % (pour GPT) et 50 % (pour Gemini) de « Traduction pire » que les phrases traduites avec Text-to-Picto sans simplification de l'entrée. Nous pensons qu'il est difficile de représenter le sens de certaines phrases médicales avec des pictogrammes même en exploitant toutes les stratégies de traduction vers les pictogrammes que nous avons vues. En raison des limites de notre méthodologie et du prompt proposé, nous estimons que les grands modèles de langue génératifs ne sont actuellement pas une solution pour améliorer la traduction automatique du français vers les pictogrammes via la simplification automatique. Toutefois, les grands modèles de langue génératifs ne cessent de s'améliorer ; nous pensons qu'ils ouvrent la voie à de nombreuses pistes de recherche.

6 Conclusion : des contributions

Dans ce chapitre, nous avons encore contribué au premier objectif de cette thèse, à savoir, le développement et l'amélioration d'un système de traduction automatique du français vers les pictogrammes pour la communication entre médecins et patients avec une DI. Nous avons répondu à plusieurs de nos sous-objectifs (3, 4 et 5).

Tout d'abord, nous avons exploré des façons de traduire une phrase médicale en pictogrammes, en effectuant une analyse syntaxique de phrases du corpus BabelDr, puis en les faisant traduire en pictogrammes Arasaac par une logopède, formatrice en CAA. Ainsi, nous avons relevé différentes stratégies de traduction en pictogrammes, dont certaines sont des opérations de simplification automatique qui étaient déjà utilisées dans Text-to-Picto telles que la substitution par des synonymes, des hyperonymes, etc. (sous-objectif 3). À partir de ces différentes analyses, nous avons un peu adapté le système et nous avons évalué manuellement ces adaptations sur les phrases de BabelDr, afin d'explorer des façons de spécialiser Text-to-Picto pour la traduction de dialogues médicaux en pictogrammes (sous objectif 4). Nous l'avons associé à plusieurs ressources françaises (générale ou médicale), avant d'évaluer l'impact de la substitution lexicale, ainsi que de la simplification automatique via des grands modèles de langue génératifs (sous-objectif 5). Nous avons montré que ces méthodes pouvaient améliorer la traduction de phrases en pictogrammes générées par Text-to-Picto dans certains cas, mais que celles-ci restaient néanmoins assez limitées avec notre méthodologie. Les grands modèles de langue génératifs ont produit des traductions équivalentes voire pires que Text-to-Picto sans simplification. Nous avons aussi constaté que des simplifications correctes ne généraient pas toujours une meilleure traduction et inversement. À noter que l'impact de la simplification automatique via des grands modèles de langue génératifs sur le résultat d'un système de traduction de texte en pictogrammes n'avait jamais été évaluée. Il faudrait toutefois améliorer le prompt.

Dans le prochain chapitre, nous proposons de revenir sur un problème spécifique à Text-to-Picto qui a été relevé ici : la désambiguïsation lexicale. En effet, que ce soit en utilisant la ressource de synonymes ReSyf, le lexique des paraphrases médicales ou la simplification automatique avec des grands modèles de langue génératifs, nous avons été confronté à des homonymes, qui ont généré des erreurs de traduction en pictogrammes dues à un contexte peu adapté. Dans ReSyf, nous avons sélectionné le synset le plus fréquent comme l'avait fait Sevens (2018) dans les premières versions de Text-to-Picto avec WordNet. En plus de générer des termes et des définitions polysémiques, le dictionnaire de paraphrases médicales contenait des entrées peu adaptées aux phrases que nous voulons traduire. La qualité des traductions générées après une simplification automatique de l'entrée de Text-to-Picto avec des grands modèles de langue génératifs a également été impactée par l'ambiguïté.

Chapitre 5

Désambiguïsation lexicale

Introduction

Dans ce chapitre, nous décrivons une méthode visant à désambiguïser automatiquement les mots de contenu (noms, verbes, adjectifs et adverbes), afin d'obtenir une traduction sémantiquement correcte en pictogrammes Arasaac, Sclera ou Beta (Norré *et al.*, 2023). Par exemple, dans la phrase source « avez-vous appliqué une crème sur la lésion ? », le mot « crème » peut être interprété comme une pommade ou de la crème (fraîche) liquide (Figure 5.1). Un système de traduction doit être capable de produire un pictogramme correct, ici un qui représentera le sens de pommade. Nous ciblons en particulier la désambiguïsation de mots polysémiques français qui peuvent être utilisés oralement par des médecins dans des questions et des instructions pour effectuer une anamnèse médicale (échelle de tri des urgences hospitalières).

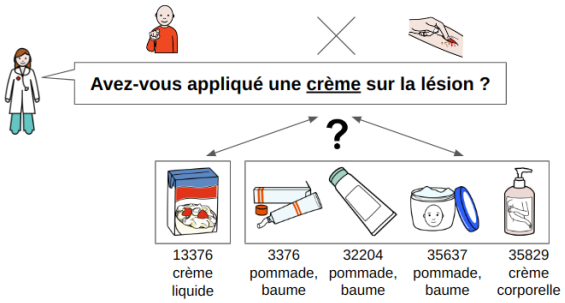


FIGURE 5.1 – Pictogrammes Arasaac : exemple du mot à désambiguïser « crème »

Dans la section 1 de ce chapitre, nous proposons un état de l'art des recherches sur les systèmes de traduction automatique vers les pictogrammes qui ont abordé la question de la désambiguïsation lexicale. La section 2 décrit notre méthodologie et les différents modèles de langue utilisés, tandis que la section 3 présente le corpus d'évaluation (3.1), ainsi que les résultats de nos évaluations (3.2). Ceux-ci sont discutés dans la section 4. Et la section 5 explore la capacité de désambiguïsation lexicale des grands modèles de langue génératifs.

1 État de l'art : désambiguïsation pour la traduction vers les pictogrammes

La désambiguïsation lexicale a déjà été utilisée dans des systèmes de traduction automatique vers les pictogrammes, afin d'améliorer la traduction de mots polysémiques pour le domaine général et dans différentes langues naturelles.

Pour l'anglais, Mihalcea et Leong (2008) décrivent un outil de désambiguïsation lexicale basique, qui repose sur le Princeton WordNet (PWN) de Miller (1995), mais ils n'évaluent pas son efficacité au sein de leur système de traduction de texte vers les pictogrammes. Imam *et al.* (2019) testent différentes méthodes de désambiguïsation lexicale – le Lesk original, un Lesk adapté, la similarité maximale, le Séparateur à Vaste Marge (SVM) – avec le WordNet anglais. Elles montrent que le système complet avec le SVM obtient les meilleurs résultats en mesurant le rappel, la précision et le F-score. Dans la version néerlandaise du système Text-to-Picto (Vandeghinste *et al.*, 2017), Sevens *et al.* (2016a) choisissaient le sens du WordNet néerlandais le plus fréquent. Ils ont ensuite utilisé un module de désambiguïsation lexicale externe basé sur un SVM et développé au sein du projet DutchSemCor (Vossen *et al.*, 2012). [1] Dans Text-to-Picto, ce module est placé après la recherche de synsets. Sevens (2018) évalue spécifiquement son apport à l'aide d'un corpus de 50 phrases contenant au moins un mot ambigu. Elle a obtenu une amélioration en précision pour Sclera (de 29/50 à 41/50) et pour Beta (de 28/50 à 42/50), démontrant l'intérêt d'intégrer une étape de désambiguïsation lexicale.

Pour le français, Vaschalde (2018), Vaschalde *et al.* (2018), Macaire *et al.* (2022) ont été les premiers à souligner l'importance d'utiliser la désambiguïsation lexicale dans un système de traduction vers les pictogrammes. Nous avons vu dans le chapitre 2 que Macaire *et al.* (2023b) ont poursuivi ces travaux avec le système Voice2Picto. Il inclut un module de désambiguisation lexicale pour les noms français, qui est basé sur le modèle neuronal – de type *transformer* (Vaswani *et al.*, 2017) – développé par Vial *et al.* (2019). [2] En ce qui concerne le langage médical, il existe des systèmes de traduction avec des pictogrammes, mais ils n'intègrent pas de désambiguïsation lexicale. C'est le cas du Text-to-Picto français (Norré *et al.*, 2021c,d, 2022). Nous visons donc à évaluer l'apport de la désambiguïsation lexicale – des noms, verbes, adjectifs et adverbes – dans le contexte d'une langue de spécialité pour la traduction automatique vers les pictogrammes, une question qui n'a pas encore été abordée dans la littérature. Nous proposons de comparer plusieurs types de modèles de langue (plongements lexicaux statiques ou contextuels | neuronaux, pré-entraînés, génératifs) incluant différents corpus d'entraînement, car aucune étude que nous venons de citer n'a évalué spécifiquement leur impact sur la désambiguïsation de mots polysémiques en pictogrammes.

1. https://github.com/rubenIzquierdo/wsd_svm
2. https://github.com/getalp/disambiguate

2 Méthode de désambiguïsation

Notre méthode s'inspire de celle d'Orkphol et Yang (2019) qui utilisent le cosinus de similarité, Word2Vec et des relations de WordNet pour désambiguïser les mots anglais. Nous présentons notre algorithme de désambiguïsation lexicale ci-dessous, au travers de l'exemple d'un mot à désambiguïser, « alcool », dans la phrase source « avez-vous bu de l'alcool ? » (Figure 5.2). Le lemme « alcool » a trois synsets différents dans WOLF (Sagot et Fišer, 2008), le WordNet français que nous utilisons par défaut dans le système Text-to-Picto (cf. chapitre 3) : {alcool, boisson alcoolisée} (07884567-n), {alcool} (14708720-n), {alcool, spiritueux} (14941230-n). Les deux possibilités de traduction vers un pictogramme Arasaac sont : l'alcool que l'on boit (id : 26626), l'alcool isopropylique (id : 2984). Le troisième synset n'est pas lié à des pictogrammes Arasaac.

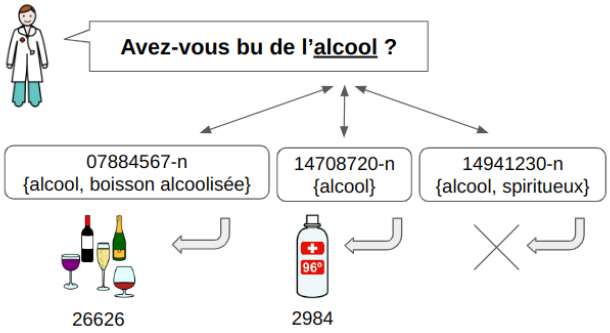


FIGURE 5.2 – Pictogrammes Arasaac : exemple du mot à désambiguïser « alcool »

Nous différencions les étapes utilisant des plongements lexicaux statiques et des plongements lexicaux contextuels en les indiquant respectivement avec (a) et (b).

1. Obtenir une représentation vectorielle de la phrase.

   (a) Récupérer dans Word2Vec (Mikolov *et al.*, 2013) ou fastText (Bojanowski *et al.*, 2017) les vecteurs des lemmes (mots de contenu) de la phrase donnée en entrée, c'est-à-dire les noms, les verbes, les adjectifs et les adverbes – étiquetés avec TreeTagger (Schmid, 1994), en incluant les auxiliaires et sans les lemmes doublons s'il y en a (par exemple, pour « avez-vous eu... », nous utilisons « avoir » et « vous »). Faire la moyenne de ces vecteurs afin d'obtenir une représentation contextuelle à partir d'une représentation statique (le vecteur de la phrase).

   (b) Récupérer dans CamemBERT (Martin *et al.*, 2020), FlauBERT (Le *et al.*, 2020), DrBERT (Labrak *et al.*, 2023a,b) ou CamemBERT-bio (Touchent *et al.*, 2023) un vecteur des lemmes (mots de contenu) de la phrase donnée en entrée afin de les utiliser comme un contexte (le vecteur de la phrase).

   Exemple (lemmes des mots de contenu) : « avoir, boire, alcool ».

2. Obtenir une représentation vectorielle des mots avec une relation sémantique.

   (a) Pour chaque synset *i* (de 1 à N) lié au lemme polysémique dans le WordNet français, récupérer tous les lemmes ayant des relations sémantiques avec le lemme : synonymes (syn), hyperonymes (hype), hyponymes (hypo), mots proches avec un POS différent (relation *eng_derivative*) (xpos). Ensuite, obtenir les représentations distribuées de tous ces mots sémantiquement liés dans Word2Vec ou fastText et faire leur moyenne afin d'obtenir une représentation contextuelle de chaque synset *i* (le vecteur de relation).

   (b) Similairement, pour chaque synset *i* (de 1 à N), récupérer la liste des mots liés sémantiquement comme en 2.a et les joindre en une chaîne unique. Ensuite, obtenir dans CamemBERT, FlauBERT, DrBERT ou CamemBERT-bio un vecteur contextuel représentant chaque synset *i* (le vecteur de relation).

   Exemple (mots des relations) :{'**synset1**' (07884567-n) : ['boisson alcoolisée' (syn), 'boisson' (hype), 'breuvage' (hype), 'drogue' (hype), 'apéritif' (hypo), 'eau-de-vie' (hypo), 'esprit' (hypo), 'gnole' (hypo), 'gnôle' (hypo), 'kava' (hypo), 'kava-kava' (hypo), 'kawa' (hypo), 'kumiz' (hypo), 'liqueur' (hypo), 'perry' (hypo), 'pulque' (hypo), 'ratafia' (hypo), 'saké' (hypo), 'spiritueux' (hypo), 'vin' (hypo), 'alcoolique' (xpos), 'alcoolisé' (xpos), 'intoxiquer' (xpos)], '**synset2**' (14708720-n) : ['liquide' (hype), 'alcool isopropylique' (hypo), 'alcool méthylique' (hypo), 'butan-1-ol' (hypo), 'butanol' (hypo), 'carbinol' (hypo), 'cyclohexanol' (hypo), "d'alcool de bois" (hypo), 'de naphte de bois' (hypo), "d'esprit de bois" (hypo), 'diol' (hypo), 'éthanol' (hypo), 'Éthanol' (hypo), 'glycérine' (hypo), 'glycérol' (hypo), 'glycol' (hypo), 'isopropanol' (hypo), 'le méthanol' (hypo), 'méthanol' (hypo), 'propanol' (hypo), 'stérol' (hypo), 'alcoolique' (xpos), 'alcoolisé' (xpos)], '**synset3**' (14941230-n) : ['spiritueux' (syn), 'liquide' (hypo)]}.

3. Calculer les similarités cosinus entre le vecteur de la phrase et le vecteur de relation de chaque synset *i*.

   Exemple (similarités cosinus) : {'**synset1**' (07884567-n) : 0.64, '**synset2**' (14708720-n) : 0.35, '**synset3**' (14941230-n) : 0.25}.

4. Utiliser les scores cosinus pour sélectionner le(s) pictogramme(s) à obtenir. Classer les synsets triés par cosinus de similarité décroissant. Commencer par récupérer le(s) pictogramme(s) du synset qui vient en premier (rang 1) ; si ce synset n'est pas lié à un pictogramme, récupérer le(s) pictogramme(s) du synset qui vient après, et ainsi de suite, jusqu'à ce qu'un pictogramme soit trouvé (rang > 1).

   Exemple (pictogrammes) : {'**synset1**' (07884567-n) : 26626, '**synset2**' (14708720-n) : 2984, '**synset3**' (14941230-n) : -}.

Pour nos modèles de langue, nous avons utilisé des modèles pré-entraînés pour le français (Table 5.1) : frWac2Vec et frWiki2Vec pour Word2Vec (Fauconnier, 2016) [3] ; Common Crawl + Wikipédia pour fastText (Grave *et al.*, 2018). [4] frWac2Vec est une collection de plongements lexicaux entrainés sur le corpus frWaC (Baroni *et al.*, 2009), qui contient 1,6 milliard de mots. Il a été construit à partir du web. Le crawl a été limité au domaine .fr, tout en utilisant les mots de moyenne fréquence du corpus Le Monde Diplomatique et des listes de vocabulaire de base du français. [5] Le corpus frWiki2Vec a été entraîné sur 600 millions de mots. Il existe 12 versions différentes de frWac2Vec et 8 de frWiki2Vec : lemmatisées ou non, étiquetées ou non, *Continuous Bag Of Words (CBOW)* ou Skip-gram, avec des vecteurs de dimensions différentes et diverses fréquences minimales de mots au sein des corpus. Nous avons repris ici des modèles de dimension 500 (1a|1b|2a), lemmatisés avec TreeTagger, mais non étiquetés.

Nous avons testé tous les modèles pré-entraînés de CamemBERT [6] (8A-F), FlauBERT [7] (9A-D), DrBERT [8] (10A-D) et CamemBERT-bio (11A), [9] avec un nombre varié de paramètres (54 à 373 M). CamemBERT et FlauBERT sont des modèles de langue généraux. CamemBERT a été entraîné sur différents corpus de taille différente (4, 135 ou 138 Go de texte) : (A-B) OSCAR (Ortiz Suárez *et al.*, 2019), (C) Wikipédia, (E-F) CCNet ; Wikipédia est le seul corpus homogène en termes de genre et de style à ne pas provenir de Common Crawl. FlauBERT a été entraîné sur un corpus hétérogène de 71 Go, constitué de 24 sous-corpus (Wikipédia, des livres, le Common Crawl, etc.). Quant aux modèles DrBERT, ils sont plus petits (4 ou 7,4 Go pour le plus grand), mais spécifiques au domaine médical, car ils ont été entraînés sur le corpus NACHOS (Labrak *et al.*, 2023a,b), qui est composé de 24 ressources médicales libres de droit. Deux modèles ont été entraînés *from scratch* (A-B), tandis que les deux autres sont le résultat d'un pré-entraînement continu de modèles existants, le PubMedBERT (Gu *et al.*, 2021) médical anglais (C) et CamemBERT (D). CamemBERT-bio est un modèle spécialisé, également issu d'un pré-entraînement continu du modèle $\text{CamemBERT}_{\text{Base}}$ sur un corpus biomédical français libre de droit de 2,7 Go.

Nous avons entraîné des modèles Word2Vec (3-4) et fastText (6-7) – CBOW (a) et Skip-gram (b), de taille 500 (sans ') et 300 (avec ') – sur le corpus CLEAR (Grabar et Cardon, 2018), [10] en utilisant les mêmes hyperparamètres de Word2Vec que Cardon (2021, p. 47). [11] Pour l'entraînement avec fastText, nous avons utilisé les hyperparamètres par

3. https://fauconnier.github.io
4. https://fasttext.cc/docs/en/crawl-vectors
5. https://wacky.sslmit.unibo.it/doku.php?id=corpora
6. https://camembert-model.fr
7. https://github.com/getalp/Flaubert
8. https://drbert.univ-avignon.fr
9. https://huggingface.co/almanach/camembert-bio-base
10. http://natalia.grabar.free.fr/resources.php#clear
11. Hyperparamètres : -window 7 -sample 1e-5 -hs 1 -negative 50 -mincount 20 -alpha 0.025.

#	Modèle	Corpus	(#) Param.	Taille	# Types | Go
1a	Word2Vec	frWac2Vec	CBOW	500	119 227
1b	Word2Vec	frWac2Vec	Skip-gram	500	119 227
2a	Word2Vec	frWiki2Vec	CBOW	500	66 819
3a	Word2Vec	CLEAR (médical + général)	CBOW	500	198 164
3a'	Word2Vec	CLEAR (médical + général)	CBOW	300	198 164
3b	Word2Vec	CLEAR (médical + général)	Skip-gram	500	198 164
3b'	Word2Vec	CLEAR (médical + général)	Skip-gram	300	198 164
4a	Word2Vec	CLEAR (médical)	CBOW	500	79 456
4a'	Word2Vec	CLEAR (médical)	CBOW	300	79 456
4b	Word2Vec	CLEAR (médical)	Skip-gram	500	79 456
4b'	Word2Vec	CLEAR (médical)	Skip-gram	300	79 456
5a	fastText	Common Crawl + Wikipédia	CBOW	300	?
6a	fastText	CLEAR (médical + général)	CBOW	500	198 164
6a'	fastText	CLEAR (médical + général)	CBOW	300	198 164
6b	fastText	CLEAR (médical + général)	Skip-gram	500	198 164
6b'	fastText	CLEAR (médical + général)	Skip-gram	300	198 164
7a	fastText	CLEAR (médical)	CBOW	500	79 456
7a'	fastText	CLEAR (médical)	CBOW	300	79 456
7b	fastText	CLEAR (médical)	Skip-gram	500	79 456
7b'	fastText	CLEAR (médical)	Skip-gram	300	79 456
8A	$\text{CamemBERT}_{\text{Base}}$	OSCAR	110 M	768	138 Go
8B	$\text{CamemBERT}_{\text{Base}}$	OSCAR (sous-ensemble)	110 M	768	4 Go
8C	$\text{CamemBERT}_{\text{Base}}$	Wikipédia	110 M	768	4 Go
8D	$\text{CamemBERT}_{\text{Base}}$	CCNet	110 M	768	135 Go
8E	$\text{CamemBERT}_{\text{Base}}$	CCNet (sous-ensemble)	110 M	768	4 Go
8F	$\text{CamemBERT}_{\text{Large}}$	CCNet	335 M	1 024	135 Go
9A	$\text{FlauBERT}_{\text{Base, uncased}}$	Wikipédia / livres / Common Crawl / etc.	137 M	768	71 Go
9B	$\text{FlauBERT}_{\text{Base, cased}}$	Wikipédia / livres / Common Crawl / etc.	138 M	768	71 Go
9C	$\text{FlauBERT}_{\text{Large, cased}}$	Wikipédia / livres / Common Crawl / etc.	373 M	1 024	71 Go
9D	$\text{FlauBERT}_{\text{Small, cased}}$	Wikipédia / livres / Common Crawl / etc.	54 M	512	71 Go
10A	$\text{DrBERT}_{\text{Base, cased}}$	NACHOS (large)	110 M	768	7,4 Go
10B	$\text{DrBERT}_{\text{Base, cased}}$	NACHOS (small)	110 M	768	4 Go
10C	$\text{DrBERT}_{\text{Base, cased}}$	NACHOS (small-PubMedBERT)	110 M	768	4 Go
10D	$\text{DrBERT}_{\text{Base, cased}}$	NACHOS (small-CamemBERT)	110 M	768	4 Go
11A	$\text{CamemBERT-bio}_{\text{Base}}$	biomed-fr	110 M	768	2,7 Go

TABLE 5.1 – Modèles de langue pour la désambiguïsation lexicale (n = 35) : Word2Vec, fastText, CamemBERT, FlauBERT, DrBERT et CamemBERT-bio

défaut. CLEAR est un corpus médical français composé de trois sous-corpus : des articles d'encyclopédies collaboratives en ligne (Wikipédia et la version Vikidia pour un public d'enfants), des notices de médicaments et des résumés de synthèses de la littérature scientifique médicale de la fondation Cochrane. Il s'agit d'un corpus comparable, avec des textes en version technique et en version simple / simplifiée. Nous avons utilisé les trois sous-corpus, une fois avec les articles d'encyclopédie médicaux (146 millions de mots au total) et une autre fois en ajoutant les articles généraux (+ 65 millions de mots). Ici, nous n'avons pas prétraité – mis en minuscule, tokenisé, lemmatisé, étiqueté – ce corpus avant l'entraînement. [12] Il convient de noter que CLEAR fait partie des corpus qui ont été utilisés pour entraîner DrBERT et CamemBERT-bio.

12. Nous avons finalement mis en minuscule, tokenisé et lemmatisé le corpus CLEAR médical avec TreeTagger avant de ré-entraîner les modèles fastText, mais nous n'avons pas obtenu de meilleurs résultats par rapport aux versions non lemmatisées (entre -2 % et +2 % de précision selon les paramètres).

Nous proposons donc d'évaluer des modèles de langue neuronaux entraînés avec des données générales et/ou médicales (Table 5.1). Nous comparons Word2Vec (1-4) et fastText (5-7) aux modèles BERT pour le français, basés sur une architecture neuronale de type *transformer* (Vaswani *et al.*, 2017) : CamemBERT (8), FlauBERT (9), DrBERT (10) et CamemBERT-bio (11). Nous testons aussi plusieurs WordNets français : WOLF et les trois versions – *coverage, f-score, precision* – du WoNeF (Pradet *et al.*, 2014).

3 Évaluation de la désambiguïsation

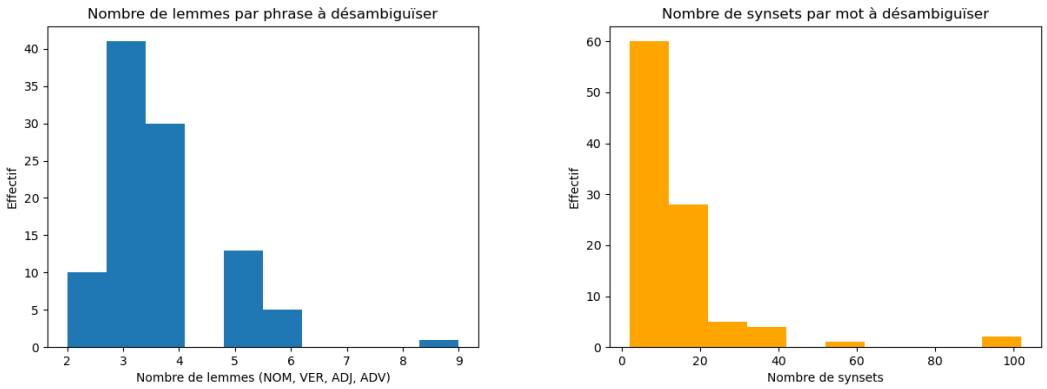
Nous décrivons d'abord comment nous avons conçu notre corpus d'évaluation pour évaluer la désambiguïsation lexicale (section 3.1). Nous présentons ensuite les résultats de notre évaluation automatique (section 3.2).

3.1 Corpus d'évaluation

Pour construire un corpus d'évaluation adapté à notre tâche, nous avons traduit automatiquement plusieurs centaines de phrases du système de traduction médicale BabelDr (Bouillon *et al.*, 2021) avec Text-to-Picto. Nous avons utilisé les modes AZ (tri des "noms" de pictogrammes selon l'ordre alphabétique) et ZA (inverse). Nous avons fait cela dans le but de détecter des mots ayant au moins deux traductions différentes en Arasaac avec la même catégorie grammaticale que le mot ambigu. Nous avons échantillonné 100 lemmes polysémiques [13] et avons extrait, pour chacun d'eux, au moins une phrase du système BabelDr – contenant au minimum deux lemmes qui soient un NOM, un VER, un ADJ ou un ADV (le nombre moyen de lemmes par phrase est de 3,67 ; la répartition est montrée dans la Figure 5.3a) –, au moins un pictogramme Arasaac avec un sens correct, un pictogramme Arasaac avec un sens incorrect et leur(s) synset(s) WOLF. Plus d'informations sur ce corpus sont disponibles dans l'annexe A.4.

Signalons encore qu'en construisant notre corpus de référence, nous avons volontairement évité les verbes pronominaux, mais aussi les expressions multi-mots qui sont utilisées comme des noms de pictogrammes par Arasaac, car nous pensons que des traitements linguistiques spécifiques afin de les traduire automatiquement par un seul pictogramme seraient nécessaires. C'est le cas de « prise de sang » traduit par deux pictogrammes « tenir » + « sang » (cf. chapitre 4, section 3.4.2). Ces expressions peuvent générer des problèmes d'ambiguïté dans le système Text-to-Picto si elles ne sont pas encodées dans un dictionnaire ou annotées avec deux synsets WordNet. En moyenne

13. Notre évaluation est basée sur celle de Sevens (2018), à savoir la *test point method* (Shiwen, 1993). « Un point de test est un problème spécifique qu'un système de traduction automatique doit résoudre. Dans la méthode des points de test, pour chaque phrase de test, la correspondance de sous-chaîne est utilisée pour déterminer si le point de test spécifique a été correctement traité » [traduction de Sevens (2018, p. 164)].



(a) Nombre de lemmes par phrase (b) Nombre de synsets par mot

FIGURE 5.3 – Statistiques du corpus de 100 mots à désambiguïser

(Figure 5.3b), les 100 mots polysémiques de notre corpus sont liés à 13,49 synsets. Le minimum est 2 synsets (pour le nom « seringue ») et le maximum 102 (pour le verbe « donner » contre 44 pour « *give* » dans le WordNet 3.1 et l'Open English WordNet). [14] Notre corpus d'évaluation est constitué de 52 noms, 38 verbes, 5 adjectifs et 5 adverbes.

3.2 Évaluation automatique

Dans un premier temps, nous avons évalué la précision des traductions Arasaac pour nos 100 mots polysémiques (Table 5.2), générées en mode AZ et ZA par le système Text-to-Picto sans module de désambiguïsation lexicale et avec des paramètres optimisés (cf. chapitre 4, Table 4.2). [15] L'ensemble des mots testés ont bien reçu une traduction, mais celle-ci n'était correcte que dans 35 % à 45 % des cas. Un mot peut avoir plusieurs traductions possibles en Arasaac, mais pas dans les autres ensembles, car il existe moins de pictogrammes en Sclera et Beta. [16] Le F-score varie entre 52 et 62 %.

	Précision ↑	**Rappel ↑**	**F-score ↑**
Arasaac			
AZ	35,35 %	**99** %	52,09 %
ZA	**45,45** %	**99** %	**62,29** %
Moyenne	40,40 %	99 %	57,19 %

TABLE 5.2 – Résultats du Text-to-Picto français, mode AZ et ZA, sur 100 mots polysémiques sans désambiguïsation lexicale pour l'ensemble de pictogrammes Arasaac avec les métriques précision, rappel et F-score : WOLF

14. https://en-word.net/lemma/give
15. Paramètres : -threshold 9 -hyperonym 15 -antonym 10 -oov 3 -direct_route 2 (sans -xpos).
16. Sur les 100 mots de notre corpus d'évaluation, 55 sont polysémiques en Sclera, 24 monosémiques et 21 ne sont pas traduits par Text-to-Picto, que ce soit en mode AZ ou ZA.

Ensuite, nous avons calculé automatiquement le rappel des principaux modèles utilisés – les CBOW de taille 500 (sauf pour le fastText général, 5a) et les $\text{BERT}_{\text{Base}}$ (A) – en nous limitant aux pictogrammes du synset avec le rang 1 (étape 4. de notre méthode). Nous constatons que beaucoup de ces synsets ne sont pas associés à un pictogramme Arasaac, Sclera ou Beta (Figure 5.4). Entre 38 et 60 % des mots associés au synset le plus similaire n'ont pas de traduction en pictogramme dans les trois ensembles. La méthode avec le rang 1 donne un faible rappel (entre 32-50 % selon le modèle de langue pour Arasaac, voire 15-29 % pour Sclera et Beta). Nous observons que la méthode avec le rang > 1 donne le même rappel que dans notre système Text-to-Picto français sans désambiguïsation lexicale (cf. Table 5.2) : entre 99-100 % pour Arasaac (Figure 5.4). Sclera et Beta ont un rappel entre 73-76 %. Cela souligne l'importance de rechercher plus d'un synset acceptable, pour tenir compte de la couverture plutôt faible des ensembles de pictogrammes.

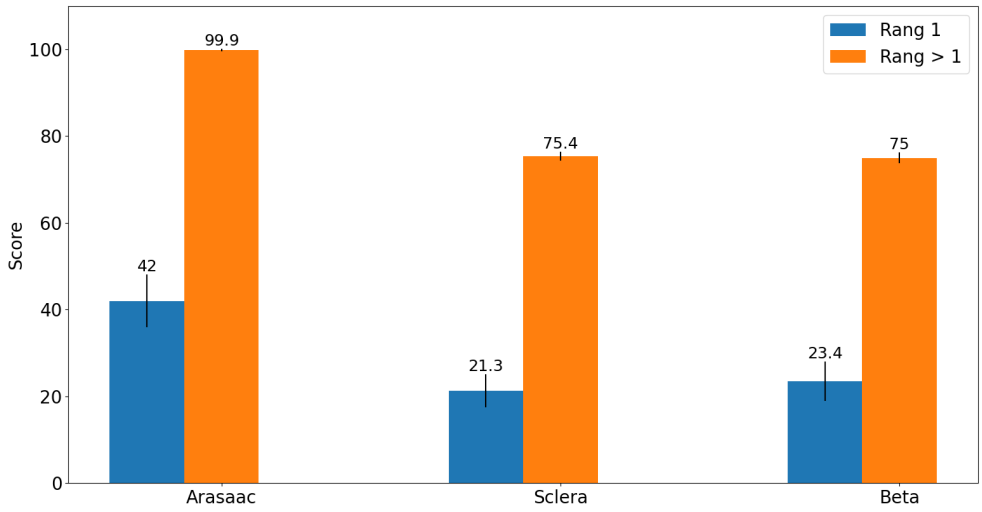


FIGURE 5.4 – Résultats de la désambiguïsation lexicale (rang 1 et rang > 1) sur 100 mots polysémiques par ensemble de pictogrammes avec la métrique rappel : WOLF

Nous avons aussi évalué automatiquement la précision de tous nos modèles de désambiguïsation lexicale sur la base des synsets corrects de chaque mot polysémique traduit en pictogramme Arasaac (Table 5.3). Pour ce faire, nous comparons chaque synset obtenu à ceux de notre corpus d'évaluation. Des pictogrammes – d'un même ensemble – liés à différents synsets ont parfois été acceptés pour un même mot, car ils étaient adaptés au contexte de la phrase comme dans l'exemple (a) de la Figure 5.5 pour la phrase « avez-vous d'autres problèmes de santé? ». Comme baseline, nous utilisons Text-to-Picto sans désambiguïsation lexicale en mode AZ – le mode par défaut dans notre système français – sur les mêmes 100 mots pour Arasaac. Le F-score est donné à titre indicatif pour permettre la comparaison avec d'autres systèmes; ce qui nous intéresse ici c'est la précision. La significativité est calculée par rapport à la baseline en utilisant le test de McNemar (1947).

#	Précision ↑	Am. rel. ↑	F-score ↑	Am. rel. ↑	#	Précision ↑	Am. rel. ↑	F-score ↑	Am. rel. ↑
AZ	35,35 %	–	52,09 %	–	ZA	45,45 %	+10,10	62,29 %	+10,20
1a	66 %	+30,65**	79,51 %	+27,42	8A	45 %	+09,65	62,06 %	+09,97
1b	**73** %	**+37,65****	**84,39** %	**+32,30**	8B	41 %	+05,65	58,15 %	+06,06
2a	58 %	+22,65**	73,41 %	+21,32	8C	46 %	+10,65	63,01 %	+10,92
3a	53 %	+17,65*	69,28 %	+17,19	8D	**48** %	**+12,65**	**64,86** %	**+12,77**
3a'	53 %	+17,65*	69,28 %	+17,19	8E	41 %	+05,65	58,15 %	+06,06
3b	58 %	+22,65**	73,41 %	+21,32	8F	44 %	+08,65	61,11 %	+09,02
3b'	61 %	+25,65**	75,77 %	+23,68	9A	44 %	+08,65	61,11 %	+09,02
4a	62,62 %	+27,27**	76,72 %	+24,63	9B	45 %	+09,65	62,06 %	+09,97
4a'	60,60 %	+25,25**	75,18 %	+23,09	9C	39 %	+03,65	56,11 %	+04,02
4b	61,61%	+26,26**	75,95 %	+23,86	9D	**49 %**	**+13,65***	**65,77** %	**+13,68**
4b'	54,54%	+19,19**	70,33 %	+18,24	10A	**45** %	**+09,65**	**62,06** %	**+09,97**
5a	**66** %	**+30,65****	**79,51** %	**+27,42**	10B	42 %	+06,65	59,15 %	+07,06
6a	56 %	+20,65**	71,79 %	+19,70	10C	**45** %	**+09,65**	**62,06** %	**+09,97**
6a'	56 %	+20,65**	71,79 %	+19,70	10D	42 %	+06,65	59,15 %	+07,06
6b	65 %	+29,65**	78,78 %	+26,69	[1-4]	60,12 %	+24,77	75,09 %	+23,00
6b'	65 %	+26,65**	78,78 %	+26,69	[5-7]	**61,22** %	**+25,87**	**75,94** %	**+23,85**
7a	60 %	+24,65**	75,00 %	+22,91	[8]	44,16 %	+08,81	61,27 %	+09,18
7a'	57 %	+21,65**	72,61 %	+20,52	[9]	44,25 %	+08,90	61,35 %	+09,26
7b	63 %	+27,65**	77,30 %	+25,21	[10]	43,50 %	+08,15	60,62 %	+08,53
7b'	63 %	+27,65**	77,30 %	+25,21	[11]|11A	46 %	+10,65	63,01 %	+10,92

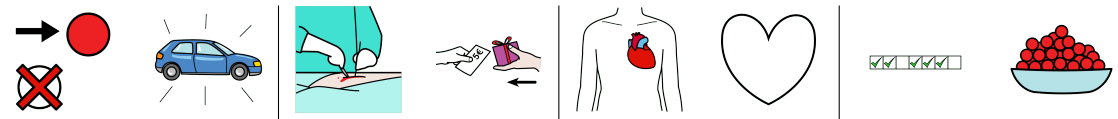
* $p < 0{,}05$; ** $p < 0{,}01$

TABLE 5.3 – Résultats de la désambiguïsation lexicale (rang > 1) sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec les métriques précision et F-score avec amélioration relative : WOLF

Le modèle qui obtient la meilleure précision – et donc le meilleur F-score – est le modèle Word2Vec Skip-gram avec le corpus frWac2Vec (1b : 73 %), suivi du même modèle en version CBOW (1a : 66 %), ainsi que du fastText pour la langue générale (5a : 66 %), puis des modèles fastText Skip-gram que nous avons entraînés sur la partie médicale et générale du corpus CLEAR (6b|6b' : 65 %). Il est important de souligner que notre meilleur modèle (1b) atteint une précision de 73 %, soit une amélioration relative de +37,65 par rapport aux performances de l'actuel système Text-to-Picto français pour Arasaac sans désambiguïsation lexicale sur des cas problématiques. La différence de tous les modèles Word2Vec et fastText est très significative, sauf deux où la différence reste significative (3a|3a'). Il faut noter que la baseline AZ et les modèles Word2Vec entraînés sur la partie médicale de CLEAR (4a|4a'|4b|4b') ont un rappel de 99 %.

Pour illustrer l'apport de la désambiguïsation, nous présentons des exemples de pictogrammes problématiques avec Text-to-Picto en mode AZ ou ZA pour 4 mots à désambiguïser (Figure 5.5). Le pictogramme de gauche représente le sens correct, celui de droite le sens incorrect. Le pictogramme le plus adapté pour l'adjectif « autre » (a) dans notre phrase « avez-vous d'autres problèmes de santé ? » a été lié à deux synsets. Nous avons donc accepté les deux (02069355-a et 02070188-a). Avec nos méthodes de désambiguïsation lexicale, le cas (a) est toujours mal traduit par le pictogramme (b) « nouveau » – associé au synset WOLF {autre, neuf, nouveau} (01640850-a) – dans nos 35 modèles (cf. Table 5.1). Cependant, le sens correct du verbe « opérer » (c) dans « il ne faut pas vous opérer » (contre le pictogramme (d) « vendre » lié au synset {opérer,

vendre, commercialiser, distribuer, échanger} dont le sens doit provenir de l'expression « opérer une transaction ») ; du nom « cœur » (e) dans « nous allons écouter votre cœur » (contre le symbole (f) et non l'organe) ; de l'adverbe « souvent » (g) dans « avez-vous souvent des aphtes ? » (contre (h) « beaucoup ») a été sélectionné dans respectivement 23, 24 et 29 modèles.



(a) 17054 (b) 4705 (c) 2715 (d) 4613 (e) 5530 (f) 6652 (g) 37029 (h) 7168

FIGURE 5.5 – Pictogrammes Arasaac : exemple des mots à désambiguïser (a|b) « autre » (adjectif), (c|d) « opérer » (verbe), (e|f) « cœur » (nom), (g|h) « souvent » (adverbe)

Comme mentionné dans la section 2, il existe plusieurs versions du corpus frWac2Vec (Table 5.4, 1-12) et frWiki2Vec (13-20). Nous avons également calculé leur rappel, précision et F-score avec leur amélioration relative. Nous constatons que les modèles frWiki2Vec ont une moins bonne précision que ceux entraînés sur frWac2Vec. Deux modèles frWiki2Vec de taille 1 000 (13|15) ont d'ailleurs une précision de 44 et 46 %, une différence qui apparût comme non significative par rapport au mode AZ en utilisant le test de McNemar. Pour tous les autres résultats, nous observons que la différence avec Text-to-Picto en mode AZ est statistiquement très significative.

#	Lem.	Phrase	Param.	Taille	Cutoff	Rappel ↑	Précision ↑	Am. rel. ↑	F-score ↑	Am. rel. ↑
1	–	–	CBOW	200	0	**100** %	69 %	+33,65**	81,65 %	+29,56
2	–	–	CBOW	200	100	**100** %	66 %	+30,65**	79,51 %	+27,42
3	–	–	Skip-gram	200	100	**100** %	66 %	+30,65**	79,51 %	+27,42
4	–	–	Skip-gram	500	100	**100** %	63 %	+27,65**	77,30 %	+25,21
5	–	–	Skip-gram	500	200	**100** %	65 %	+29,65**	78,78 %	+26,69
6|1a	v	–	CBOW	500	100	**100** %	66 %	+30,65**	79,51 %	+27,42
7|1b	v	–	Skip-gram	500	100	**100** %	**73** %	**+37,65****	**84,39** %	**+32,30**
8	v	–	Skip-gram	700	50	**100** %	65 %	+29,65**	78,78 %	+26,69
11	–	v	CBOW	500	10	**100** %	66 %	+30,65**	79,51 %	+27,42
12	–	v	CBOW	500	100	**100** %	63 %	+27,65**	77,30 %	+25,21
13	–	–	CBOW	1 000	100	**100** %	44 %	+08,65	61,11 %	+09,02
14	–	–	CBOW	1 000	200	**100** %	63 %	+27,65**	77,30 %	+25,21
15	–	–	Skip-gram	1 000	100	**100** %	46 %	+10,65	63,01 %	+10,92
16	–	–	Skip-gram	1 000	200	**100** %	**64** %	**+28,65****	**78,04** %	**+25,95**
17|2a	v	–	CBOW	500	10	**100** %	58 %	+22,65**	73,41 %	+21,32
18	v	–	CBOW	700	100	**100** %	59 %	+23,65**	74,21 %	+22,12
19	v	–	Skip-gram	1 000	100	**100** %	63 %	+27,65**	77,30 %	+25,21
20	v	–	Skip-gram	1 000	200	**100** %	**64** %	**+28,65****	**78,04** %	**+25,95**

* $p < 0{,}05$; ** $p < 0{,}01$

TABLE 5.4 – Résultats de la désambiguïsation lexicale (rang > 1) de frWac2Vec (1-12) et frWiki2Vec (13-20) sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec les métriques rappel, précision et F-score avec amélioration relative : WOLF

Nous avons évalué ces modèles sur le WOLF, mais aussi sur les trois différentes versions d'un autre WordNet français, le WoNeF (Pradet *et al.*, 2014), afin de vérifier que ce dernier n'apporte pas de meilleures performances. Nos résultats montrent que

le WOLF et les WoNeF sont très différents (Figure 5.6), le WOLF (en bleu) étant meilleur en termes de rappel, précision et F-score. Si nous comparons les WoNeF entre eux, en moyenne, la version *coverage* (en orange) obtient le meilleur rappel (49,94 %), le WoNeF *f-score* (en vert) a la meilleure précision (44,41 %), tandis que la petite version *precision* (en rouge) est extrêmement limitée pour toutes les métriques.

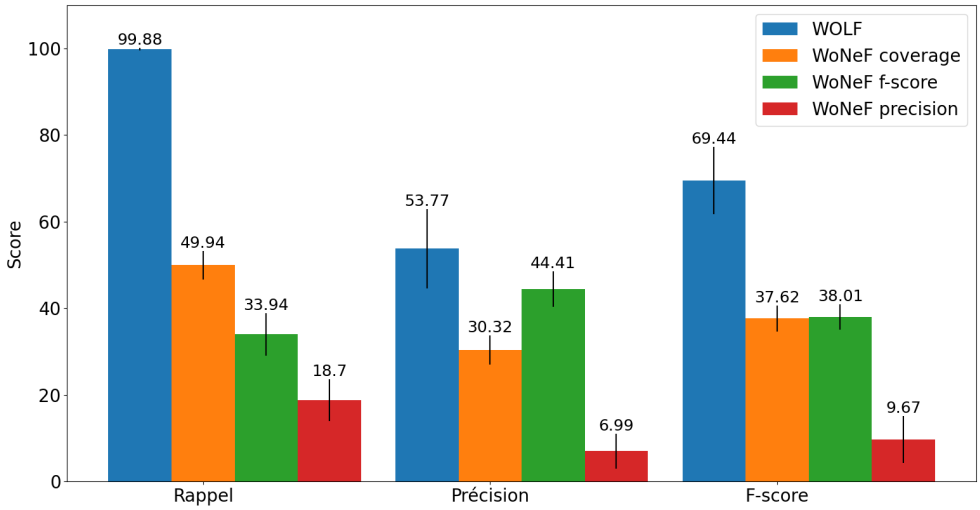


FIGURE 5.6 – Résultats de la désambiguïsation lexicale (rang > 1) sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec les métriques rappel, précision et F-score : WOLF et WoNeF *coverage*, *f-score*, *precision*

4 Discussion : analyse des variables

Nous discutons de l'impact de différentes variables sur les résultats obtenus : les modèles de langue (section 4.1), leurs corpus d'entraînement (section 4.2), mais aussi les synsets issus de WordNet et les pictogrammes associés (section 4.3).

4.1 Modèles de langue

En ce qui concerne les performances des modèles de langue (cf. Table 5.3), nous constatons que la précision moyenne des 9 modèles fastText (61,22 %) est très proche de celle des 11 modèles Word2Vec (60,12 %). Les moyennes plus faibles de CamemBERT (44,16 %), FlauBERT (44,25 %), DrBERT (43,50 %) et CamemBERT-bio (46 %) sont contre-intuitives. Nous avons donc émis l'hypothèse que la petite taille de notre contexte d'entrée pourrait être un facteur. Pour vérifier cela, nous avons réalisé des expériences tenant compte du contexte pour désambiguïser les lemmes.

Nous avons extrait les usages (<USAGE>) dans WOLF (# 48 233), c'est-à-dire des syntagmes ou des courtes phrases qui servent d'exemples d'usage. Comme ils ne sont

disponibles qu'en anglais – directement transférés du PWN –, nous avons traduit automatiquement en français les 649 usages associés aux lemmes de notre corpus d'évaluation, avec Google Translate. [17] Par exemple, un usage de WOLF traduit de l'anglais « *alcohol (or drink) ruined him* » vers le français est « l'alcool (ou la boisson) l'a ruiné » pour le synset 07884567-n du lemme nominal « alcool » (cf. Figure 5.2).

Nous avons testé plusieurs paramètres d'encodage pour les 15 modèles BERT avec WOLF (Table 5.5). [18] Il y avait deux configurations pour le vecteur de la phrase (étape 1.b de notre méthode) : A) les lemmes des mots de contenu de la phrase à désambiguïser (par exemple, « avoir boire alcool ») ; B) toute la phrase à désambiguïser (« avez-vous bu de l'alcool ? »). Pour le vecteur de relation (étape 2.b), nous avons testé six configurations : a) les mots des 4 types de relations (synonyme(s), hyperonyme(s), hyponyme(s) et synonyme(s) proche(s) avec une différente partie de discours) ; b) les mots des 4 types de relations terminés chacun par un point ; c) les usages ; d) les usages terminés par un point ; e) les usages et les synonymes terminés chacun par un point ; [19] f) les usages et les mots des 4 types de relations terminés chacun par un point.

#	Paramètre	Exemple
	Vecteur de la phrase	
A | LEMME	Lemmes des mots de contenu	avoir, boire, alcool
B | PHRASE	Phrase entière	"avez-vous bu de l'alcool?"
	Vecteur de relation	
a | 4REL	Mot(s) des 4 types de relation	'boisson alcoolisée', ... 'boisson', ..., 'vin', ..., 'alcoolique', ...
b | 4REL.	Mot(s) des 4 types de relation + [.]	'boisson alcoolisée.', ..., 'boisson.', ..., 'vin.', ..., 'alcoolique.', ...
c | USAGE	(Usage)	"l'alcool (ou la boisson) l'a ruiné"
d | USAGE.	(Usage) + ([.])	"l'alcool (ou la boisson) l'a ruiné."
e | USYN.	(Usage) + ([.]) + synonyme(s) + [.]	"l'alcool (ou la boisson) l'a ruiné.", 'boisson alcoolisée.', ...
f | U4REL.	(Usage) + ([.]) + mot(s) des 4 types de relation + [.]	"l'alcool (ou la boisson) l'a ruiné.", 'boisson alcoolisée.', 'boisson.', ...

TABLE 5.5 – Différents paramètres d'encodage pour les modèles BERT

En fonction de ces configurations d'encodage, la précision de nos modèles peut varier de -14 % à +20 % par rapport à notre méthode principale (Table 5.6), c'est-à-dire nos résultats BERT avec les paramètres A-a (LEMME-4REL). [20] Notre moins bon modèle obtient maintenant une précision de 33 % (contre 39 % pour A-a) et le meilleur 65 % (contre 49 %). En utilisant seulement les usages (A|B-c†|d†), nous obtenons un rappel de 66 % sur nos 100 mots. La configuration avec un rappel de 100 % et la précision moyenne la plus élevée sur les 15 modèles est donc la B-e (PHRASE-USYN.) avec 46,86 % (contre 44,13 %). Même si nous observons des améliorations par rapport à notre méthode principale A-a, les modèles de langue BERT restent moins précis que Word2Vec et fastText.

17. https://translate.google.be

18. Comme précisé dans la méthodologie (cf. section 2), nous donnons toujours à BERT une chaîne unique de caractères (*str*) et non une liste de mots (*list*). Nous avons en effet remarqué que ces deux types de données généraient des résultats légèrement différents.

19. Encoder une phrase suivie par une liste de mots comme entrée à des modèles BERT est une technique qui montre des résultats prometteurs pour la simplification lexicale (Wilkens *et al.*, 2022).

20. De 48 à 34 % pour 8D (B-a) et de 45 à 65 % pour 10A (A-c†).

	Précision ↑															
# Paramètre	**8A**	**8B**	**8C**	**8D**	**8E**	**8F**	**9A**	**9B**	**9C**	**9D**	**10A**	**10B**	**10C**	**10D**	**11A**	**Moyenne**
A-a | LEMME-4REL	45 %	41 %	46 %	48 %	41 %	44 %	44 %	45 %	39 %	**49** %*	45 %	42 %	45 %	42 %	46 %	44,13 %
A-b | LEMME-4REL.	41 %	40 %	38 %	44 %	39 %	41 %	36 %	45 %	45 %	**49** %	**49** %*	42 %	41 %	39 %	41 %	42,00 %
A-c | LEMME-USAGE †	51 %	53 %	53 %	50 %	59 %*	57 %*	53 %	48 %	59 %*	56 %*	**65** %**	56 %*	57 %*	46 %	48 %	54,06 %
A-d | LEMME-USAGE. †	57 %*	50 %	53 %	50 %	53 %	**60** %**	53 %	51 %	51 %	56 %*	59 %*	51 %	56 %*	46 %	48 %	52,93 %
A-e | LEMME-USYN.	43 %	46 %	45 %	**48** %	45 %	44 %	42 %	43 %	39 %	**48** %	42 %	39 %	41 %	36 %	38 %	42,60 %
A-f | LEMME-U4REL.	44 %	43 %	35 %	**48** %	42 %	46 %	41 %	43 %	44 %	45 %	40 %	43 %	39 %	38 %	41 %	42,13 %
B-a | PHRASE-4REL	44 %	38 %	43 %	34 %	42 %	49 %*	43 %	39 %	33 %	**50** %*	36 %	36 %	42 %	33 %	43 %	40,33 %
B-b | PHRASE-4REL.	41 %	41 %	36 %	43 %	43 %	45 %	46 %	42 %	38 %	**47** %	32 %	38 %	41 %	43 %	42 %	41,20 %
B-c | PHRASE-USAGE †	48 %	**57** %*	56 %*	54 %*	51 %	50 %	54 %	46 %	53 %	48 %	54 %*	46 %	51 %	45 %	51 %	50,93 %
B-d | PHRASE-USAGE. †	51 %	51 %	50 %	53 %	53 %	54 %	53 %	**59** %*	56 %*	53 %*	50 %	53 %	53 %	50 %	50 %	52,60 %
B-e | PHRASE-USYN.	44 %	48 %	48 %	**53** %**	51 %*	50 %*	41 %	46 %	47 %	47 %	**53** %**	43 %	45 %	42 %	45 %	46,86 %
B-f | PHRASE-U4REL.	44 %	43 %	35 %	**48** %	42 %	46 %	41 %	43 %	44 %	45 %	40 %	43 %	39 %	38 %	37 %	41,86 %
Moyenne / modèle	46 %	45 %	44 %	47 %	46 %	48 %	45 %	45 %	45 %	49 %	46 %	44 %	45 %	41 %	44 %	
Moyenne / famille	46,69 %						46,62 %				44,68 %				44,16 %	

* $p < 0{,}05$; ** $p < 0{,}01$ || † = 66 % Rappel

TABLE 5.6 – Résultats de la désambiguïsation lexicale (rang > 1) des différentes configurations d'encodage pour les modèles BERT sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec la métrique précision : WOLF

Les syntagmes et les phrases peuvent être utiles aux modèles contextuels BERT pour améliorer la précision (A|B-c†|d†). Nous avons cependant vu qu'encoder seulement les usages comme vecteurs des relations n'est pas efficace, car il n'y a pas assez d'usages qui sont associés à des synsets liés à des pictogrammes Arasaac (rappel à 66 % sur notre corpus d'évaluation). C'est pourquoi, il faut combiner les usages avec les synonymes afin d'avoir un rappel de 100 % (B-e). Les modèles de langue BERT appliqués aux données de WOLF avec notre méthodologie n'offrent toutefois pas une grande amélioration en précision si nous les comparons à notre système Text-to-Picto en mode ZA, qui obtient 45,45 % sur les mêmes 100 mots polysémiques (cf. Table 5.2).

Une autre piste serait d'exploiter le French SemCor (Nasiruddin *et al.*, 2015), ainsi que les données de SemEval 2013, mais elles ne sont pas adaptées aux dialogues médicaux et seuls les noms français sont annotés en sens. De plus, des étapes de prétraitements seraient nécessaires car nous utilisons le *Synset Identifier* et non les *Sense Keys*. Afin de traduire automatiquement les usages du WOLF de l'anglais vers le français, nous avons utilisé Google Translate, mais aussi Marian-NMT (Junczys-Dowmunt *et al.*, 2018), avec le modèle *transformer* pré-entraîné OPUS en-fr (Tiedemann et Thottingal, 2020). [21] Nous avons constaté de légères différences en précision avec certains de nos modèles de désambiguïsation, par exemple -5 % pour FlauBERT (9C) et +4 % pour CamemBERT (8A|8F), avec la configuration B-e. Cependant, Google Translate obtient une moyenne de 46,86 % et Marian-NMT de 46,33 % sur les 15 modèles BERT.

4.2 Corpus d'entraînement

Nous avons également évalué l'effet de différents corpus d'entraînement (langue générale, langue médicale, etc.) sur nos performances. Les modèles pré-entraînés sur des données de langue générale ont obtenu des précisions plus élevées en moyenne (66,2 % de moyenne pour les 10 modèles frWac2Vec de la Table 5.4 et 66 % pour fastText entraîné sur Common Crawl + Wikipédia) que les modèles CLEAR (médical + général : 58,37 % ; médical : 60,29 %).

Au-delà de l'effet de la taille du corpus de ces modèles pré-entraînés, une autre explication de ces résultats inattendus pourrait être que même si les mots à désambiguïser sont intégrés dans des dialogues médicaux, ils ne sont pas tous des termes médicaux : sur nos 100 mots polysémiques, 19 % se retrouvent dans le Wiktionary médical extrait par Cardon (2018), 56 % dans le lexique médical de Grabar et Hamon (2016) et 29 % dans la terminologie spécialisée de la SNOMED internationale (Côté, 1996) ; 37 % des mots de notre corpus sont dans au moins deux de ces trois ressources. [22]

21. Marian-NMT : https://huggingface.co/docs/transformers/model_doc/marian ; OPUS : https://huggingface.co/Helsinki-NLP/opus-mt-en-fr (en-fr).

22. Le Wiktionary utilisé pour la comparaison comprend 8 012 termes, le lexique médical 29 641 termes et la terminologie SNOMED 151 104 termes.

En plus des modèles Word2Vec et fastText, nous avons utilisé des modèles BERT pré-entraînés sur des données générales (CamemBERT et FlauBERT), mais aussi biomédicales (DrBERT et CamemBERT-bio). Nous avons vu que leur moyenne était très proche, que ce soit sans l'optimisation des paramètres (CamemBERT : 44,16 %; FlauBERT : 44,25 %; DrBERT : 43,50 %; CamemBERT-bio : 46 %) ou avec (CamemBERT : 46,69 %; FlauBERT : 46,62 %; DrBERT : 44,16 %; CamemBERT-bio : 43,50 %). Il faut toutefois noter que le modèle CamemBERT-bio est le plus petit et qu'à l'exception des modèles 8B, 8C et 8E, les modèles spécialisés de DrBERT sont aussi largement inférieurs en taille (cf. Table 5.1). Ce n'est pas si surprenant pour des modèles BERT, Martin *et al.* (2020) avaient montré que des modèles CamemBERT de 4 Go pouvaient fournir des résultats comparables voire meilleurs que ceux du CamemBERT standard entraîné sur 138 Go, et ce, pour différentes tâches de Traitement Automatique des Langues (TAL).

Une façon d'améliorer les résultats des modèles neuronaux déjà pré-entraînés consiste à les spécialiser sur d'autres corpus plus adaptés à la tâche en continuant leur entraînement. Yusufali *et al.* (2023) ont par exemple obtenu des meilleures performances de leur système de Communication Alternative et Améliorée (CAA) pour la prédiction de mots avec des modèles BERT et RoBERTa réglés sur un plus petit corpus spécialisé de CAA – un corpus conversationnel de 6 000 messages anglais, imaginés par des *crowds* – que sur des modèles BERT généraux. Néanmoins, il n'existe pas de corpus de phrases en pictogrammes assez grand pour pouvoir envisager cette piste.

4.3 Synsets et pictogrammes

Finalement, nous avons comparé plusieurs WordNets français (cf. Figure 5.6). Chacun d'eux a généré des pictogrammes assez différents, en raison de synsets différents. Comme nous l'avions déjà noté (Norré *et al.*, 2021d), dans WOLF et WoNeF (*coverage*, *f-score*) la moitié des synsets du WordNet anglais ont été traduits en français (cf. chapitre 3, section 2). Le WoNeF *precision* ne contient quant à lui que 15 482 lemmes. Nous avons en effet déjà montré que le WOLF permet d'obtenir un meilleur score BLEU qu'avec les trois versions de WoNeF avec Text-to-Picto pour les ensembles de pictogrammes Arasaac, Sclera et Beta.

Choisir un synset adapté n'est pas toujours suffisant pour obtenir une bonne traduction. Dans certaines situations, il faudrait en plus affiner manuellement ou automatiquement la sélection des pictogrammes au sein du synset obtenu avec la désambiguïsation lexicale. C'est le cas d'Arasaac où plusieurs pictogrammes – parfois une vingtaine – peuvent être associés à un seul synset. Il peut s'agir de pictogrammes identiques avec un personnage différent (unisexe, masculin ou féminin), mais aussi avec un sens plus ou moins différent bien qu'ils soient liés au même synset : des « crèmes » (ou « pommades ») destinées à différents usages (lésion, soleil, visage, corps, cf. Figure

5.1), « soulever » la planche des toilettes, un bébé, un objet, les volets, etc. Nous ne disposons pas d'information à propos de la méthode utilisée par Arasaac pour annoter ses pictogrammes et obtenir toutes ses métadonnées.

5 Et les grands modèles de langue génératifs ?

Pour aller plus loin, nous avons voulu savoir si les grands modèles de langue génératifs utilisés dans le chapitre précédent (cf. Table 4.10) peuvent sélectionner le synset WOLF lié au pictogramme correct d'une phrase qui serait donnée en entrée à Text-to-Picto. Nous avons donc repris notre corpus de 100 mots à désambiguïser (cf. section 3.1), ainsi que la plateforme d'évaluation des grands modèles de langue génératifs. [23] Pour chaque mot, nous avons généré automatiquement un prompt sans exemple (*zero-shot*), qui contient trois informations : la phrase entière à désambiguïser, le lemme polysémique, ainsi que tous les synsets WOLF possibles – avec Text-to-Picto – contenant les mots des 4 types de relations (synonymes, hyperonymes, hyponymes et mots proches avec un POS différent). Nous les avons entrés un par un dans la plateforme en effaçant à chaque fois l'historique de chat entre les requêtes.

Prompt : Dans la phrase "[phrase entière à désambiguïser]", quel est le sens du mot "[lemme polysémique]" le plus adapté au contexte de la phrase parmi les suivants : Sens [synset] -> [mot(s) des 4 types de relations] ; Sens [synset] -> [mot(s) des 4 types de relations](...). Donne l'identifiant du sens.

Exemple : Dans la phrase "avez-vous bu de l'alcool?", quel est le sens du mot "alcool" le plus adapté au contexte de la phrase parmi les suivants : Sens fre-30-07884567-n -> boisson alcoolisée, boisson, breuvage, drogue, apéritif, eau-de-vie, esprit, gnole, gnôle, kava, kava-kava, kawa, kumiz, liqueur, perry, pulque, ratafia, saké, spiritueux, vin, alcoolique, alcoolisé, intoxiquer ; Sens fre-30-14708720-n -> liquide, alcool isopropylique, alcool méthylique, butan-1-ol, butanol, carbinol, cyclohexanol, d'alcool de bois, de naphte de bois, d'esprit de bois, diol, éthanol, Éthanol, glycérine, glycérol, glycol, isopropanol, le méthanol, méthanol, propanol, stérol, alcoolique, alcoolisé ; Sens fre-30-14941230-n -> spiritueux, liquide. Donne l'identifiant du sens.

Pour chaque lemme, nous avons traité avec un script les sorties des grands modèles de langue génératifs afin d'extraire le synset du bruit avec une expression régulière : fre-30-\d+-(n|r|a|v). Ce traitement pourrait être directement implémenté dans Text-to-Picto. Nous avons ensuite évalué automatiquement le rappel, la précision et le F-score avec leur amélioration relative (Table 5.7). Le rappel non indiqué entre parenthèses représente le pourcentage de synsets obtenus par le grand modèle de langue

23. Paramètres : temperature = 0, max output tokens = 2 000.

génératif, tandis que le rappel entre parenthèses désigne le pourcentage de synsets obtenus par le modèle, mais qui sont liés à au moins un pictogramme Arasaac dans le système Text-to-Picto.

#	Rappel ↑	Am. rel. ↑	Précision ↑	Am. rel. ↑	F-score ↑	Am. rel. ↑	Rang
AZ	99 %	–	35,35 %	–	52,09 %	–	–
gemini-pro-dev-api	99 % (53 %)	– (-46,00)	39,39 %	+04,04	56,36 %	+04,27	4
claude-2.1	99 % (55 %)	– (-44,00)	**47,47** %	**+12,12**	**64,17** %	**+12,08**	**1**
command-r	**100** % (55 %)	**+01,00** (-44,00)	44,00 %	+08,65	61,11 %	+09,02	2
gpt-3.5-turbo-0613	97 % (47 %)	-02,00 (-52,00)	35,05 %	-00,30	51,50 %	-00,59	6
llama-2-70b-chat	98 % (52 %)	-01,00 (-47,00)	36,73 %	+01,38	53,43 %	+01,34	5
mistral-medium	**100** % (59 %)	**+01,00** (-40,00)	43,00 %	+07,65	60,14 %	+08,05	3
qwen1.5-72b-chat	**100** % (56 %)	**+01,00** (-43,00)	44,00 %	+08,65	61,11 %	+09,02	2
vicuna-33b	97 % (49 %)	-02,00 (-50,00)	28,86 %	-06,49	44,50 %	-07,59	7

* $p < 0{,}05$; ** $p < 0{,}01$

TABLE 5.7 – Résultats de la désambiguïsation lexicale avec des grands modèles de langue génératifs sur 100 mots polysémiques pour l'ensemble de pictogrammes Arasaac avec les métriques rappel, précision et F-score avec amélioration relative : WOLF

Nous n'avons pas observé de différence significative en utilisant un test de McNemar. Le rappel à 99 % pour Gemini et Claude s'explique par la présence du lemme « savoir ». Un synset possible contient des mots vulgaires liés à la sexualité. Certains grands modèles de langue génératifs possèdent bien un filtre qui rejette des mots susceptibles d'être jugés offensants pour les utilisateurs. Dans ce cas, le grand modèle de langue génératif refuse de répondre. Il s'agit néanmoins d'une limitation importante concernant l'utilisation des grands modèles de langue génératifs dans un contexte médical, en plus de la diffusion de données privées. Nous donnons ici les raisons du refus.

gemini-pro-dev-api : *I apologize if my previous response violated your content moderation guidelines. I am still under development and learning to follow the rules and regulations set by humans. I will try my best to avoid generating responses that violate your guidelines in the future.* [24]

claude-2.1 : *I do not have enough context to determine if any content violates guidelines. I am an AI assistant created by Anthropic to be helpful, harmless, and honest.* [25]

En ce qui concerne la taille de la fenêtre contextuelle, la requête pour les lemmes des verbes « donner » (102 synsets), « faire » (98 synsets) et « présenter » (61 synsets) dépassait le seuil des modèles gpt-3.5-turbo-0613 (4 096 tokens), llama-2-70b-chat (4 096 tokens) et vicuna-33b (2 048 tokens). C'est pourquoi, ils n'ont pas donné de réponse – ou alors un message d'erreur – et leur rappel est aussi un peu plus bas (respectivement, 97, 98 et 97 %).

24. Je m'excuse si ma réponse précédente a enfreint vos directives de modération de contenu. Je suis encore en développement et j'apprends à suivre les règles et réglementations fixées par les humains. Je ferai de mon mieux pour éviter de générer des réponses qui enfreignent vos directives à l'avenir.

25. Je ne dispose pas de suffisamment de contexte pour déterminer si un contenu enfreint les directives. Je suis un assistant IA créé par Anthropic pour être utile, inoffensif et honnête.

Comme les grands modèles de langue génératifs produisent du bruit et que nous avions seulement besoin d'un synset, nous avons commencé par extraire automatiquement les synsets. Cette approche a toutefois été limitée. Par exemple, Command R et Mistral donnaient parfois plusieurs synsets possibles – identiques ou différents – et/ou à éviter par mot à désambiguïser. Une expertise humaine a donc été nécessaire pour l'extraction (vérification), afin de récupérer celui qui correspondait à notre requête. Ici, le cas de Command R était moins problématique que celui de Mistral.

command-r : Le sens le plus adapté dans ce contexte est : Sens fre-30-07884567-n. C'est le sens du mot "alcool" le plus souvent associé à l'expression "avoir bu de l'alcool". L'identifiant est : fre-30-07884567-n.

mistral-medium : Le sens le plus adapté au contexte de la phrase "avez-vous appliqué une crème sur la lésion ?" est "pommade, émollient, article de toilette" (Sens fre-30-03128519-n). Dans ce contexte, "crème" fait référence à une substance médicamenteuse ou apaisante que l'on applique sur la peau pour traiter une lésion. Les autres sens proposés ne sont pas pertinents dans ce contexte. Par exemple, "laitage" et "produit laitier" (Sens fre-30-07847198-n) font référence à des aliments à base de lait, tandis que "choix, élite, choisir" (Sens fre-30-08387035-n) font référence à l'action de sélectionner ou à un groupe de personnes ou de choses sélectionnées.

Pour cette tâche, les meilleurs modèles sont : Claude (47,47 %), Command R (44 %) | Qwen (44 %) et Mistral (43 %). En termes de précision, certains grands modèles de langue génératifs auraient donc des résultats proches des modèles de type BERT ; d'autres auraient de moins bons résultats voire même moins bons que la baseline AZ de Text-to-Picto comme GPT et Vicuna. Les résultats des grands modèles de langue génératifs ne peuvent néanmoins pas être directement comparés à ceux de Text-to-Picto et des autres modèles de langue – Word2Vec, fastText, BERT – puisque nous avions évalué leur précision avec notre méthode de désambiguïsation de rang > 1 (étape 4.).

Pour pallier cette limite, nous aurions pu demander aux grands modèles de langue génératifs d'ordonner les synsets, du plus adapté au moins adapté au contexte de la phrase à désambiguïser de façon à éliminer les synsets non liés à des pictogrammes Arasaac en réduisant l'espace de recherche et ainsi améliorer la précision. Nous avons noté qu'au plus il y avait de synsets possibles, au plus les grands modèles de langue génératifs se trompaient. Au total, 12 mots ont été correctement désambiguïsés par tous les grands modèles de langue génératifs testés (9 noms et 3 verbes). Inversement, 33 mots n'ont jamais été correctement désambiguïsés.

Étant donné la durée de vie très réduite de ce type de modèles de langue – parfois quelques mois pour certaines versions –, ces premiers résultats ne seront pas longtemps reproductibles. Néanmoins, nous avons montré que les modèles de langue génératifs actuels ne semblent pas être la meilleure solution pour la désambiguïsation

lexicale des traductions automatiques en pictogrammes. À notre connaissance, ceci n'avait pas été testé. Des améliorations doivent encore être réalisées. La qualité des modèles et des données (dont celle des synsets de WordNet) en fait également partie.

6 Conclusion : des perspectives pour la désambiguïsation

Dans ce chapitre, nous avons réalisé des expériences concernant la désambiguïsation lexicale avec différents modèles de langue, statiques (Word2Vec et fastText), ou contextuels (CamemBERT, FlauBERT, DrBERT et CamemBERT-bio), en français médical. Nous avons montré que la désambiguïsation lexicale améliore la précision de la traduction automatique vers les pictogrammes sur des cas problématiques confirmant ainsi les résultats de Sevens (2018), qui portaient néanmoins sur la langue générale. Nous avons observé que la méthode la plus performante consiste à utiliser Word2Vec ou fastText (cf. Tables 5.3 et 5.4). Selon nos expériences, l'efficacité des modèles de langue contextuels est assez limitée comparée aux représentations vectorielles statiques pour cette tâche. L'avantage de cette méthode est qu'elle est facilement applicable pour d'autres langues qui possèdent un WordNet et des corpus de taille moyenne, qui peuvent être utilisés pour entraîner Word2Vec ou fastText. De plus, elle pourrait remplacer la route sémantique et l'algorithme de recherche A* de Text-to-Picto. Nous avons aussi constitué et rendu disponible un premier corpus d'évaluation pour la désambiguïsation lexicale de phrases médicales françaises en pictogrammes Arasaac.

Notre méthode peut néanmoins être encore adaptée. Par exemple, nous pourrions tester d'autres opérations que la moyenne afin de produire une représentation contextuelle de vecteurs statiques ou remplacer le cosinus de similarité. Il serait également possible d'utiliser d'autres relations dans WOLF, au-delà des synonymes, hyperonymes, hyponymes et mots proches avec un POS différent. Le WOLF et les trois WoNeF proposent 18 relations exploitables.[26] Pour l'anglais, Orkphol et Yang (2019) avaient par exemple utilisé une méthode de désambiguïsation combinant Word2Vec, le calcul de cosinus de similarité, les définitions et les relations de WordNet suivantes : « *hypernym, hyponym, meronym, holonym, entailment* ». Une autre perspective pour améliorer le système serait de réaliser la désambiguïsation lexicale sur la base d'autres métadonnées de l'API d'Arasaac comme les différents noms des pictogrammes ou encore avec la ressource de synonymes ReSyf (cf. chapitre 4). Nous pourrions aussi utiliser la ressource UMLS pour restreindre le choix des pictogrammes du synset obtenu. Enfin, nous aurions également pu tester le modèle de désambiguïsation lexicale pour le français de Vial *et al.* (2019). Des adaptations auraient toutefois été nécessaires.

26. Les relations sont : *has_hyperonym, eng_derivative, also_see, be_in_state, category_domain, causes, derived, holo_member, holo_part, holo_portion, instance_hypernym, near_antonym, participle, region_domain, similar_to, subevent, usage_domain* et *verb_group.*

Chapitre 6

Études utilisateurs

Introduction

Après avoir travaillé sur le développement de systèmes de traduction automatique vers les pictogrammes, nous avons voulu savoir si ces outils produisent une sortie utile et compréhensible pour des personnes avec une Déficience Intellectuelle (DI), qui ne connaissent pas forcément ces images. Nous avons en effet vu dans le chapitre 1 qu'il existe beaucoup d'ensembles de pictogrammes différents qui peuvent être utilisés dans diverses situations de communication. Il est donc fort probable que tous les patients avec une DI n'utiliseront pas l'ensemble d'images employé au sein de la structure de soins qui les accueillera. Ce chapitre décrit deux séries d'évaluations effectuées auprès d'adultes avec une DI. D'une part, nous détaillons notre première étude (Norré *et al.*, 2024), qui vise à évaluer des traductions médicales en pictogrammes Arasaac générées par les systèmes de traduction automatique Text-to-Picto et PictoDr (section 1). Comme nous l'avons vu, il s'agit des deux seuls systèmes de traduction vers les pictogrammes pour le français qui peuvent être utilisés pour le domaine médical. La compréhension des pictogrammes isolés et la compréhension de phrases indépendantes en pictogrammes est mesurée ; différentes variables susceptibles d'influencer la compréhensibilité sont analysées (par exemple, les caractéristiques des traductions). D'autre part, nous décrivons notre deuxième étude, qui évalue la compréhension de phrases traduites en pictogrammes avec le système qui aura généré les traductions les plus compréhensibles, PictoDr (section 2). Il s'agit de questions posées lors d'une anamnèse du COVID-19. Les deux études présentent la même structure : l'objectif et les hypothèses de recherche (1.1 et 2.1), les participants (1.2 et 2.2), le design expérimental (1.3 et 2.3), l'annotation et les données collectées (1.4 et 2.4), les modèles mixtes utilisés pour les analyses statistiques (1.5 et 2.5), les résultats quantitatifs et qualitatifs (1.6 et 2.6), la discussion (1.7 et 2.7). Enfin, nous présentons les limites méthodologiques de ces deux études (section 3), avant de résumer les principales conclusions (section 4).

1 Évaluation de deux traductions en pictogrammes

Dans cette section, nous présentons une première évaluation de phrases de BabelDr traduites automatiquement en pictogrammes Arasaac par Text-to-Picto (version 2 sans désambiguïsation lexicale) et PictoDr (version avec CamemBERT, début 2023). Elle a été menée en mai-juin 2023, auprès de personnes avec une DI en Suisse, avec la collaboration du bureau Langage simplifié de Pro Infirmis Fribourg. [1]

1.1 Objectif et hypothèses de recherche

Nous présentons ici une étude, à la fois quantitative et qualitative, dont l'**objectif** est d'évaluer la compréhension de phrases médicales traduites en pictogrammes via Text-to-Picto et PictoDr auprès d'adultes avec une DI. Le but n'est pas de déterminer quel est le meilleur système de traduction automatique (c'est-à-dire celui qui traduit le plus et obtient le moins d'erreur de traduction en termes de rappel et de précision) en situation réelle, mais de voir si des personnes peuvent comprendre les questions et instructions médicales traduites en pictogrammes et si certaines différences de traduction contribuent ou non à améliorer cette compréhension. Nous avons choisi d'effectuer des entretiens – ici semi-structurés – car il s'agit de la méthode de collecte de données la plus souvent utilisée auprès de personnes avec une DI (Beail et Williams, 2014). Nos hypothèses peuvent être formulées comme suit :

- **H1 :** La compréhension des pictogrammes isolés est meilleure que la compréhension des phrases en pictogrammes.
- **H2 :** La compréhension des pictogrammes et la compréhension des phrases en pictogrammes est meilleure pour les adultes avec un niveau de langue plus élevé.
- **H3 :** La compréhension des pictogrammes et la compréhension des phrases en pictogrammes est influencée par les différences de traduction entre les systèmes Text-to-Picto et PictoDr : le nombre de pictogrammes par phrase, leur ordre, l'ajout ou non d'un pictogramme pour exprimer le temps, les (semi-)auxiliaires, le choix des pronoms, les expressions multi-mots et les autres mots.
- **H4 :** La compréhension des instructions est meilleure que la compréhension des questions en pictogrammes.

1. Pro Infirmis est la plus grande organisation d'aide aux personnes en situation de handicap en Suisse, qui travaille pour leur autodétermination et leur inclusion. Le bureau Langage simplifié propose un service de traduction de textes en langage simplifié pour le français, l'allemand et l'italien, c'est-à-dire trois langues nationales, en plus de l'anglais (https://www.langage-simplifie.ch). Il respecte les règles européennes du Facile A Lire et à Comprendre (FALC), déjà abordées dans le chapitre 2.

- **H5 :** Certaines catégories morphosyntaxiques (POS) – les noms, verbes, etc. – des mots traduits en pictogrammes sont plus compréhensibles que d'autres.

En ce qui concerne les différences de compréhension entre Text-to-Picto et PictoDr (H3), nous nous sommes intéressé à l'impact de plusieurs variables, c'est-à-dire les paradigmes de traduction qui distinguent ces deux systèmes. De plus, nous avons cherché quels étaient les facteurs qui influencent le plus la compréhension des pictogrammes et la compréhension des phrases ; c'est pourquoi nous proposerons deux modèles prédictifs. Enfin, nous voulions savoir si les personnes évaluées rencontraient habituellement des difficultés à communiquer avec un médecin et si les pictogrammes pouvaient être utiles.

1.2 Participants

Pour cette étude, nous avons collaboré avec une facilitatrice suisse de Pro Infirmis, [2] qui était chargée de questionner les participants avec une DI. Elle a l'habitude de travailler avec des personnes en situation de handicap afin de s'assurer que les textes qui leur sont soumis sont faciles à lire et à comprendre. Elle a également déjà participé à une étude de réception similaire portant sur des textes traduits humainement ou automatiquement en langue facile, *Easy Language* (post-édités ou non) auprès de ce public (Rodríguez Vázquez et Bouillon, 2021).

Le recrutement des participants a été pris en charge par la facilitatrice de l'étude. Il s'agissait d'un recrutement en interne au sein des collaborateurs (relecteurs) réguliers de Pro Infirmis. Le public cible était constitué de 9 adultes, 5 hommes (H) et 4 femmes (F), avec une DI (Table 6.1). Nous n'avons pas pu accéder à plus d'informations sur le type de handicap. Tout comme pour le focus group sur Text-to-Picto de Sevens (2018), le Quotient Intellectuel (QI) n'était pas un critère de sélection des participants. [3] L'anonymat des participants a été garanti tout au long du processus expérimental grâce à l'utilisation d'un identifiant personnel (P1 à P9). Ils sont tous des locuteurs natifs du français et nous avons 3 participants appartenant à un groupe de relecture FALC de chaque niveau de langue : A1, A2 et B1. [4] Tous nos participants savent lire.

En ce qui concerne l'âge, 5 participants sont dans la tranche des [26–35] ans et 4 dans celle des [36–45] ans. Même si certains participants ont dit avoir déjà vu ou « utilisé » des pictogrammes, aucun ne les utilisent pour communiquer et n'ont déjà vu ces images-ci. Bien que la facilitatrice ne l'ait pas toujours indiqué dans la grille des participants, nous avions noté lors des entretiens que plusieurs participants semblaient

2. Pour cette collaboration avec Pro Infirmis (et la suivante), nous avons convenu avec la facilitatrice d'une rémunération de 45 CHF (francs suisses) par participant, qui était payé comme lors de leur relecture de textes en FALC (dans le cadre de leur travail avec le bureau Langage simplifié).

3. Le QI est un mauvais indicateur pour évaluer une DI (Hessels et Hessels-Schlatter, 2010).

4. Le Cadre Européen Commun de Référence pour les langues (CECR) n'est pas non plus un indicateur de la DI. Il est toutefois utilisé par Pro Infirmis pour catégoriser ses groupes de relecteurs FALC.

P	Genre	Tranche d'âge	Niveau de français	Handicap	Commentaires (handicap, autres façons de communiquer, etc.)	Déjà utilisé les pictogrammes ?	Temps Min : sec.
P1	H	[36–45]	A2	Intellectuel	Pas de spécificité reconnue	Non	22 : 40
P2	F	[26–35]	B1	Intellectuel	Pas de spécificité reconnue	Non	18 : 00
P3	F	[36–45]	B1	Intellectuel	Pas de spécificité reconnue	Non	24 : 02
P4	H	[36–45]	B1	Intellectuel	Pas de spécificité reconnue	Oui	25 : 37
P5	H	[26–35]	A2	Intellectuel	Pas de spécificité reconnue	Non	18 : 06
P6	F	[26–35]	A1	Intellectuel	Pas de spécificité reconnue	Oui	19 : 55
P7	H	[26–35]	A1	Intellectuel	Trisomie	Oui	19 : 36
P8	H	[36–45]	A1	Intellectuel	Trisomie	Oui	23 : 39
P9	F	[26–35]	A2	Intellectuel	Pas de spécificité reconnue	Non	19 : 31

H = Homme | F = Femme || Min. : sec. = Minute : seconde

TABLE 6.1 – Description des participants de l'évaluation de deux traductions

avoir une trisomie. Certains avaient aussi des problèmes de vue et préféraient être plus proches de l'écran. Un participant a demandé si on pouvait agrandir les pictogrammes.

Avant chaque test, nous leur avons demandé leur consentement oral et ils ont chacun signé un formulaire de consentement écrit adapté en langage simplifié (en annexe). Les deux premiers tests ont eu lieu dans les locaux de Pro Infirmis Fribourg où se déroulent habituellement les relectures de textes en FALC. Nous avons assisté aux tests 3 à 9 dans l'atelier de travail adapté des participants à Fribourg. Les 7 tests ont eu lieu le même jour les uns après les autres. Nous avons observé et pris des notes supplémentaires, mais nous ne sommes pas intervenu. La facilitatrice avait un stagiaire qui assistait aussi aux tests et qui a interrogé un participant avec son aide.

Nous avons indiqué le temps de passation de chaque participant. Il s'agit de la durée de chaque enregistrement audio, qui est mesurée après la phrase exemple (pour les tests 3 à 9). Le temps comprend les 20 phrases de test en pictogramme, celle qu'on a ajouté avec la nouvelle – au moment de l'étude – façon de présenter les pictogrammes dans PictoDr et les trois questions de satisfaction. La durée moyenne était de 21 minutes et 14 secondes. La durée totale prévue pour chaque participant comprenant l'installation, la contextualisation de l'étude et les instructions était de 30 minutes.

1.3 Design expérimental

Pour cette étude, nous avons sélectionné 20 phrases sources différentes, extraites de l'échelle de tri des urgences hospitalières et provenant du système de traduction BabelDr (Bouillon *et al.*, 2021). Nous n'avons retenu que des phrases dont les traductions automatiques en pictogrammes Arasaac nous semblaient correctes, afin d'éviter la post-édition. Nous avons utilisé plusieurs autres critères de sélection en fonction de nos variables d'intérêt (Table 6.2). Tout d'abord, la moitié des phrases traduites expriment des questions et l'autre moitié, des instructions. Ensuite, chaque phrase est déclinée en deux versions (Figure 6.3) : la traduction automatique de Text-to-Picto et celle de PictoDr. Les phrases devaient avoir une traduction différente dans les deux

versions. Le nombre de pictogrammes[5] de la traduction de Text-to-Picto devait être strictement supérieur à celui de la traduction de PictoDr, ce qui est d'ailleurs souvent le cas pour diverses raisons (sous l'effet des variables 2, 3, 6 et 7, cf. infra).

#	Variable	Text-to-Picto	PictoDr
1	Nombre de pictogrammes	> PictoDr	< Text-to-Picto
2	Ordre des pictogrammes	sujet verbe objet	objet | verbe → objet | verbe
3	Temps – passé | futur	passé | futur	[non traduit]
4	(Semi-)auxiliaires – être | devoir | pouvoir	pictogramme Arasaac	[non traduit]
5	Pronoms – je | nous | vous	pictogramme Arasaac	pictogramme Arasaac adapté avec médecin
6	Expressions multi-mots	2-3 pictogrammes	1 pictogramme
7	Autres mots (1)	pictogramme différent	pictogramme différent | [non traduit]
8	Autres mots (2)	pictogramme identique	pictogramme identique

TABLE 6.2 – Différences de traduction entre Text-to-Picto et PictoDr : variables

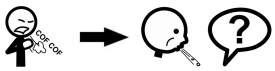
# Variable	Text-to-Picto	PictoDr
1 | 3 | 5 | 6	NOUS ALLONS FAIRE DES VACCINS **[futur]** (3) | nous (5) | faire (6) | vaccin (6)	nous (5) | **(faire) vaccin** (6)
1 | 2 | 4 | 5 | 7 | 8	DEVEZ-VOUS CRACHER QUAND VOUS TOUSSEZ ? vous (5) | **devoir** (4) | cracher (8) | quand (7) | vous (5) | tousser (7) |?	 **tousser** (7) | causer (2) | cracher (8) |?

TABLE 6.3 – Différences de traduction entre Text-to-Picto et PictoDr : exemples

Au-delà du nombre différent de pictogrammes par phrase (variable 1), les traductions de Text-to-Picto et PictoDr présentaient d'autres distinctions liées à l'architecture de ces systèmes et aux choix de leurs concepteurs. Des règles ont par exemple été implémentées dans PictoDr pour restructurer certaines phrases sujet + verbe + objet en objet ou verbe + concept « causer » (pictogramme →) + objet ou verbe (variable 2).[6] Dans Text-to-Picto, nous utilisons un pictogramme temporel (variable 3) pour le passé « avoir » (déjà) et le futur proche exprimé par le semi-auxiliaire d'aspect « aller » (cf. chapitre 4, Figure 4.5). L'auxiliaire « être » ainsi que les semi-auxiliaires modaux « devoir » et « pouvoir » sont aussi traduits par un pictogramme alors qu'ils ne le sont pas dans PictoDr (variable 4). Concernant les pronoms (variable 5), Text-to-Picto exploite les pronoms Arasaac, tandis que de nouveaux pictogrammes ont été créés par la Faculté de Traduction et d'Interprétation de Genève avec les pictogrammes « médecin » et les pronoms d'Arasaac. Dans nos traductions, les expressions multi-mots sont traduits

5. Le minimum est 2 pictogrammes par phrase et le maximum 8. Le nombre moyen de pictogrammes par phrase – avec le pictogramme « ? » – pour Text-to-Picto est 5 (questionnaire 1 : 5,4 ; questionnaire 2 : 4,6) ; celui pour PictoDr est de 3,5 (questionnaire 1 : 3,9 ; questionnaire 2 : 3,1). Au moins 68 pictogrammes différents ont été présentés dans cette étude.

6. Les deux phrases évaluées avec la variable 2 sont : « êtes-vous allergique à des médicaments ? » traduite par « médicament causer allergie » avec PictoDr ; « devez-vous cracher quand vous toussez » traduite par « tousser causer cracher » (cf. Figure 6.3).

par plusieurs pictogrammes avec Text-to-Picto et un seul avec PictoDr (variable 6). Par exemple, « faire des vaccins ». Quant aux autres mots de la phrase source, soit ils sont non traduits dans PictoDr ou traduits par un pictogramme différent dans les deux versions (variable 7), soit ils sont traduits par un pictogramme identique (variable 8).

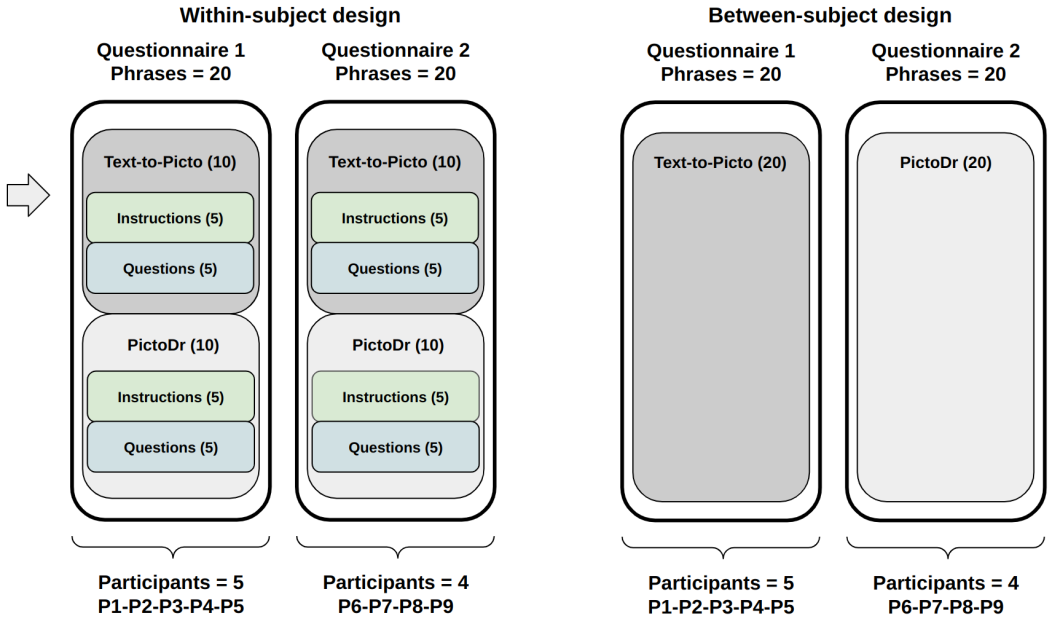


FIGURE 6.1 – Design expérimental de l'évaluation de deux traductions

La facilitatrice et son stagiaire ont fait passer 9 entretiens individuels. Chaque participant était enregistré avec un smartphone (audio seulement). Les phrases en pictogrammes étaient présentées individuellement au fur et à mesure sur un ordinateur portable au centre de l'écran sans texte. Tous les pictogrammes avaient la même taille quelle que soit la longueur de la phrase. Au total, nous avons réalisé deux questionnaires constitués de 20 questions chacun, soit 10 phrases de chacun des deux systèmes – Text-to-Picto et PictoDr (cf. annexe B.1.1) –, 5 instructions et 5 questions différentes (Figure 6.1). Les 5 premiers participants ont répondu oralement au premier questionnaire et les 4 derniers au deuxième questionnaire.[7] Contrairement à Rodríguez Vázquez et Bouillon (2021), nous avons choisi d'utiliser un *within-subject design* (et non un *between-subject design*).[8] Chaque participant a testé toutes les conditions d'une même variable (le système et le type de phrase), les phrases traduites par Text-to-Picto et les phrases traduites par PictoDr, les instructions et les questions. Nous avons fait ce choix car nous avions peu de participants et en raison de la potentielle forte variabilité interindividuelle liée au handicap. De plus, la conception intra-sujets nécessite

7. Une même phrase a donc été vue 5 ou 4 fois par un participant différent.

8. Dans le cas d'un *between-subject design*, plus de phrases d'un système auraient été évaluées par rapport à l'autre avec 9 participants. Pour 10 participants, Rodríguez Vázquez et Bouillon (2021) avaient utilisé trois versions : la première avec la traduction automatique (4 participants) ; la deuxième avec la traduction humaine (3 participants) ; la troisième avec la post-édition (3 participants).

souvent moins de participants. Nous avons veillé à ce que les phrases des deux questionnaires restent assez proches (cf. Table 6.8).

Pour éviter l'effet d'amorçage (*priming*), qui peut être lié à la répétition des conditions propre au *within-subject design*, l'ordre des phrases a bien été randomisé pour chaque participant (Saldanha et O'Brien, 2014, pp. 114, 117), même si nous avons fait en sorte que les deux versions – Text-to-Picto et PictoDr – d'une même phrase source ne se suivent jamais directement. Les questionnaires débutaient toujours avec la phrase « avez-vous mal au ventre quand vous marchez ? », traduite avec PictoDr, qui ne comptait pas dans les résultats et qui servait d'exemple pour expliquer les tâches aux participants (la version séquentielle, Figure 6.2a). Pour les tests 3 à 9, nous avons ajouté une nouvelle phrase, présentée sur papier et non sur ordinateur. Il s'agissait d'une variante de la phrase source utilisée comme exemple, mais avec une version plus récente de PictoDr et une nouvelle façon de présenter les pictogrammes, qui s'inspire des bandes dessinées (la version contextuelle, Figure 6.2b).[9] Nous voulions savoir si les participants comprenaient mieux qui étaient les personnes représentées.

(a) Phrase 0 : exemple de PictoDr (b) Phrase 21 : exemple de PictoDr

FIGURE 6.2 – Pictogrammes Arasaac : questions supplémentaires « avez-vous mal (au ventre) quand vous marchez ? » avec la version séquentielle et la version contextuelle

Afin de préparer l'étude, nous avons organisé des réunions à distance avec nos promoteurs, dont deux avec la facilitatrice préalablement contactée par email. Pour les évaluations, nous lui avons également fourni différents documents, dont plusieurs étaient des adaptations à notre étude de ceux de Rodríguez Vázquez et Bouillon (2021) :

- les 9 listes de questions numérotées de A à I pour chaque participant (P1 : A, ..., P9 : I), qui contenaient chacune les 20 phrases sources – numérotées et randomisées – écrites en français (le sens « correct »), le sens de chaque pictogramme isolé séparé par des '|', ainsi que le pictogramme à évaluer obligatoirement pour chaque phrase, indiqué en gras (cf. annexe B.1.2) ;
- les 9 questionnaires, qui contenaient une phrase en pictogrammes Arasaac au centre de chaque page (en format paysage) ; le numéro de la phrase et le numéro du pictogramme obligatoire à évaluer étaient affichés en petit et en couleur grise dans le coin inférieur droit pour aider la facilitatrice (cf. annexe B.1.3) ;

9. La version contextuelle avait déjà été conçue afin de palier à certains problèmes de compréhensibilité rencontrés par un public allophone. Elle vise à améliorer la reconnaissance entre les questions et les assertions (instructions, prescriptions, etc.), leur portée ainsi que l'identification des (inter)locuteurs.

- les 9 documents numérotés de A à I, appelés listes de contrôle, à remplir sur la base des réponses des participants (cf. annexe B.1.4) ;
- la grille à remplir avec les informations (niveau de français, tranche d'âge, handicap, etc.) de chaque participant (cf. annexe B.2) ;
- un petit guide reprenant le déroulement complet des séances (cf. annexe B.3) ;
- les instructions détaillées pour les questions (cf. annexe B.4) ;
- le formulaire de consentement à faire signer par chacun des participants et que nous avons réalisé nous-même en langage simplifié (cf. annexe B.5).

Pour chaque phrase traduite en pictogrammes Arasaac, les 9 participants avaient pour tâches : 1) d'expliquer son sens (« que veut dire cette phrase » ou « quelle(s) information(s) comprenez-vous ») ; 2) d'expliquer le sens d'un pictogramme en particulier (« qu'est-ce que ça signifie » ou « que signifie | veut dire cette image-ci ») ; 3) d'expliquer le sens des pictogrammes avec des personnes (les pronoms « je », « nous », « vous » représentant le médecin et/ou lui/elle en tant que patient.e). Au total, la facilitatrice a fait évaluer 180 phrases et 180 pictogrammes (un échantillon que nous avions prédéterminé). [10] En fonction de la réponse au premier point, elle pouvait poser des questions supplémentaires sur les pictogrammes de son choix. Elle devaient noter si les réponses lui semblaient correctes ou non sur la base des réponses attendues données dans nos documents. Les paraphrases, reformulations et synonymes étaient acceptés. Des questions sur la communication avec le médecin en général, l'utilité et les améliorations possibles de ces phrases en pictogrammes étaient posées à la fin de l'entretien.

1. Est-ce que vous rencontrez habituellement des difficultés à exprimer ou comprendre des informations de votre médecin ?
2. Aimez-vous utiliser ces (traductions en) pictogrammes ? Ou aimeriez-vous continuer à utiliser ces pictogrammes lors d'une consultation médicale avec le médecin ou à l'hôpital ? Pourquoi ? Pensez-vous que c'est utile ?
3. Et que faut-il améliorer ?

10. Nous avions choisi les pictogrammes qui avaient le moins de chance d'être traduits à l'oral et qui nous semblaient importants (par exemple, les pictogrammes temporels de Text-to-Picto, ceux représentant des expressions multi-mots, les semi-auxiliaires, etc.). Par extension, il ne s'agissait pas des pictogrammes supposés être les plus facilement interprétables.

1.4 Annotation et données collectées

Deux chercheurs en linguistique (dont l'auteur) ont retranscrit manuellement une partie des entretiens (questions-réponses), en incluant les spécificités langagières des personnes avec une DI et les nombreuses disfluences verbales (répétitions, hésitations et phrases syntaxiquement incorrectes) : « uh huh », « euh », etc. Le système de reconnaissance vocale Whisper d'OpenAI (Radford *et al.*, 2023) [11] a aussi été employé pour obtenir les retranscriptions automatiques avec un grand modèle de langue pour le français. Soit nous les avons utilisées pour corriger nos retranscriptions manuelles, soit nous en avons directement post-éditées – par exemple, pour les mots mal reconnus (« pourquoi » reconnu comme « pour toi », cf. Figures B.1a et B.1c en annexe B.6). [12]

Métadonnées	**PictoDr**	**Text-to-Picto**
Participant	P9	P9
Niveau de langue	A2	A2
Questionnaire	Q2	Q2
Numéro de la phrase	3/20 (Ph17)	18/20 (Ph17)
Type de phrase	question	question
Phrase évaluée	avez-vous les yeux rouges?	avez-vous les yeux rouges?
Pictogrammes	vous | œil/yeux rouge(s) |?	vous | œil/yeux | rouge |?
Pictogramme évalué	œil/yeux rouge(s)	rouge
Variables – phrase	1, 5, 6	1, 5, 6
Variable – pictogramme	6	6
POS – pictogramme	n	adj
Réponse(s) – phrase	« pourquoi t'as les yeux rouges? »	« comment il s'est mis la peinture dans les yeux? »
A1-Phrase	1	0
A2-Phrase	1	0
Réponse(s) – pictogramme	« des problèmes de yeux mais on sait pas quoi », « yeux rouges »	« rouges », « peinture »
A1-Picto	1	0
A2-Picto	1	0

TABLE 6.4 – Annotation de la compréhension de deux pictogrammes et de la compréhension d'une phrase selon deux annotateurs (A1-A2) : exemple de la phrase « avez-vous les yeux rouges? », ainsi que des pictogrammes « yeux rouges » et « rouge » (P9)

En nous basant sur nos retranscriptions (cf. annexe B.6), nous avons déterminé pour les 180 phrases et les 683 pictogrammes si le sens donné par le participant était correct (codé en « 1 ») ou incorrect (« 0 »). Nous nous sommes référé autant que possible aux mots de chaque participant ou aux reformulations de la facilitatrice approuvées par le participant (Table 6.4). Il pouvait aussi arriver que le participant donne des réponses plus ou moins différentes pour le sens d'un pictogramme et/ou d'une phrase. À noter que nous avions préalablement annoté chaque phrase avec nos variables (cf. Table 6.2) et chaque pictogramme avec le POS du mot de la phrase source : « n » (nom), « adj » (adjectif), etc. Nous avons créé et ajouté un tag « tp » (temps) pour des pictogrammes de Text-to-Picto (variable 3).

Nous avons donc évalué la compréhensibilité à deux niveaux : pour chaque pictogramme et pour chaque phrase. L'évaluation a été réalisée par deux annotateurs :

11. https://github.com/openai/whisper

12. Whisper ne distingue pas le changement de locuteurs, supprime les répétitions, hésitations et disfluences tout en corrigeant les locuteurs (syntaxe, ponctuation, etc.).

nous, à partir des retranscriptions (A1) ; la facilitatrice, pendant l'expérience (A2). A1 a utilisé deux catégories : « Correct » et « Incorrect » ; A2 a en plus utilisé « Correct, mais pas certain », « Incorrect, mais pas certain » et « Ne sait pas ». Les pictogrammes « ? » des questions n'ont pas été annotés par A1 et A2. Contrairement à A2, nous avons annoté tous les pictogrammes de chaque questionnaire, c'est-à-dire 683 (Table 6.5). A2 n'avait pas d'expérience particulière avec les pictogrammes.

	Questionnaire 1 | P1 à P5	**Questionnaire 2 | P6 à P9**	**Total**	**A1**	**A2**
Arasaac					
# Phrases	20 | 100	20 | 80	**180**	**180**	**180**
→ Text-to-Picto (ou PictoDr)	10 | 50	10 | 40	90	90	90
→ Instructions (ou Questions)	10 | 50	10 | 40	90	90	90
# Pictogrammes (sans « ? »)	83 | 415	67 | 268	**683**	**683**	**180**
→ Text-to-Picto	49 | 245	41 | 164	409	409	90
→ PictoDr	34 | 170	26 | 104	274	274	90
# Pictogrammes (avec « ? »)	93 | 465	77 | 308	773		
→ Text-to-Picto	54 | 270	46 | 184	454		
→ PictoDr	39 | 195	31 | 124	319		

TABLE 6.5 – Nombre de phrases et nombre de pictogrammes Arasaac par questionnaire, évalués ou non par les deux annotateurs (A1-A2)

Des accords inter-annotateurs ont été calculés avec nos annotations et celles de la facilitatrice par pictogramme (ceux correspondant aux 180 annotations de A2) et par phrase. Nous avons adapté son échelle d'annotation à la nôtre en regroupant les catégories « (In)Correct » et « (In)Correct, mais pas certain » en raison du degré de subjectivité ressenti par les annotateurs concernant le caractère (in)certain des réponses lié à la difficulté de la tâche des participants, mais aussi des annotateurs. Nous proposons deux façons de calculer les scores inter-annotateurs selon les catégories d'A2 : (a) et (b).

Exactitude des pictogrammes		**A2**		
		Correct	**Incorrect**	**Total**
A1	**Correct**	36 20,0 %	26 14,4 %	**62** **34,4 %**
	Incorrect	9 5,0 %	109 60,6 %	**118** **65,6 %**
	Total	**45** **25,0 %**	**135** **75,0 %**	**180** **100 %**

Exactitude des phrases		**A2**		
		Correct	**Incorrect**	**Total**
A1	**Correct**	36 20,0 %	10 5,5 %	**46** **25,5 %**
	Incorrect	9 5,0 %	125 69,5 %	**134** **74,5 %**
	Total	**45** **25,0 %**	**135** **75,0 %**	**180** **100 %**

FIGURE 6.3 – Matrices de confusion avec les catégories « Correct » et « Incorrect / Ne sait pas » (A1), ainsi que les catégories « Correct » et « Incorrect / Ne sait pas » (A2) d'exactitude au niveau des pictogrammes (n = 180) et des phrases (n = 180)

(a) Si nous fusionnons aussi les catégories « Incorrect » et « Ne sait pas » (Figure 6.3), le coefficient κ (Cohen, 1960) est de 0,54 pour les pictogrammes (n = 180) ; 0,72 pour les phrases (n = 180), soit respectivement un accord modéré [0,41–0,60] et un accord substantiel [0,61-0,80] selon l'échelle de Landis et Koch (1977). Comme cette

échelle fait débat dans la communauté scientifique (Bregeon *et al.*, 2019), nous avons à chaque fois calculé le coefficient α de Krippendorff (2004). Ils étaient identiques aux coefficients κ pour les pictogrammes et les phrases, mais ceux-ci sont respectivement considérés comme non fiables [< 0,67] et pauvres [0,67–0,80].

(b) Si nous ne tenons pas compte des réponses appartenant à la catégorie « Ne sait pas » de A2 – en la traitant comme des 'NA' – (Figure 6.4), les coefficients κ et α sont alors de 0,63 pour les pictogrammes (n = 159) ; 0,75 pour les phrases (n = 160), soit deux accords substantiels [0,61–0,80] selon Landis et Koch (1977) ou un accord non fiable [< 0,67] et un accord pauvre [0,67–0,80] selon Krippendorff (2004). Sur la vingtaine de données non prises en compte (ce qui aurait représenté un peu plus de 11 %), la moitié concernait des pictogrammes jugés d'après A1 comme corrects (10/21, soit 5,6 %) ou incorrects (11/21, soit 6,1 %) ; tandis que la grande majorité concernait des phrases incorrectes selon A1 (17/20, soit 9,4 %).

Exactitude des pictogrammes		A2			
		Correct	Incorrect	Total	(NA)
A1	Correct	36 22,6 %	16 10,0 %	**52** **32,6 %**	(10) (5,6 %)
	Incorrect	9 5,7 %	98 61,7 %	**107** **67,4 %**	(11) (6,1 %)
	Total	**45** **28,3 %**	**114** **71,7 %**	**159** **100 %**	(21) (11,7%)
	(NA)	(0) (0 %)	(0) (0 %)	(0) (0 %)	–

Exactitude des phrases		A2			
		Correct	Incorrect	Total	(NA)
A1	Correct	36 22,5 %	7 4,4 %	**43** **26,9 %**	(3) (1,7 %)
	Incorrect	9 5,6 %	108 67,5 %	**117** **73,1 %**	(17) (9,4 %)
	Total	**45** **28,1 %**	**115** **71,9 %**	**160** **100 %**	(20) (11,1 %)
	(NA)	(0) (0 %)	(0) (0 %)	(0) (0 %)	–

FIGURE 6.4 – Matrices de confusion avec les catégories « Correct » et « Incorrect » (A1), ainsi que les catégories « Correct » et « Incorrect » (A2) d'exactitude au niveau des pictogrammes (n = 159) et des phrases (n = 160)

Les taux de pictogrammes jugés corrects par A1 et incorrects par A2 (14,4 et 10 %) sont presque deux à trois fois supérieurs aux taux de pictogrammes jugés incorrects par A1 et corrects par A2 (5 et 5,7 %). Ces taux sont toutefois assez équivalents au niveau des phrases (respectivement 5,5 et 4,4 % contre 5 et 5,6 %). A1 surestime un peu plus la compréhension des pictogrammes corrects (au total 34,4 et 32,6 % contre 25 et 28,3 % pour A2). Les matrices de confusion indiquent aussi que les deux annotateurs sont assez d'accord sur les pictogrammes et les phrases incorrectes (entre 60-70 % des données). De plus, les scores inter-annotateurs montrent qu'il est plus difficile de trouver un accord sur la compréhension ou non de nos participants au niveau des pictogrammes que sur la compréhension des phrases traduites en pictogrammes.

En raison des mauvais accords inter-annotateurs pour les pictogrammes et les phrases, nous utiliserons les annotations d'un seul annotateur. À noter que cette première étude utilisateur est en quelque sorte une étude pilote.

1.5 Modèles linéaires généralisés à effets mixtes

Toutes nos analyses statistiques ont été réalisées avec la version 4.2.1 du logiciel R. Pour vérifier nos hypothèses (cf. section 1.1), nous avons choisi d'utiliser uniquement les annotations de A1. D'une part, elles sont plus complètes pour les pictogrammes (cf. Table 6.5). D'autre part, A1 connaissait mieux l'ensemble d'images utilisées que A2. Nous utilisons des modèles linéaires généralisés – des régressions logistiques avec ici une distribution binomiale de la réponse $Y \sim B(n, p)$ –, des modèles à effets mixtes construits avec la version 1.1.8 du package R *glmmTMB* (Brooks *et al.*, 2017). [13] Pour l'analyse des régressions, la vérification des hypothèses sur les résidus ou la détection de la multicolinéarité, nous avons respectivement employé la version 3.1-2 de *car* (Fox et Weisberg, 2019) / 6.0-94 de *caret* (Kuhn, 2008), 0.4.6 de *DHARMa* (Hartig, 2022) et 0.10.8 de *performance* (Lüdecke *et al.*, 2021). [14]

1.5.1 Définition

Pour débuter par une définition simple, reprenons celle de Tack (2021, p. 249) :

> « Mixed-effects models (Bates *et al.*, 2015) [...] are statistical models that extend (generalized) linear models by combining fixed and random factors. The models fixed effects correspond to the independent variables usually included in any linear model, whereas the models' random effects characterize a variance in the linear model's intercept and slope(s). Because such random effects represent a variability between groups, mixed models often have a hierarchical structure with observations nested into one or more higher-order levels. »

Tack (2021) avait par exemple utilisé ces modèles pour prédire les difficultés lexicales de lecture d'apprenants du français langue étrangère dans sa thèse de doctorat. Les modèles à effets mixtes semblent également adaptés pour expliquer et prédire la compréhension de – phrases en – pictogrammes.

Dans notre cas, les observations représentent des pictogrammes ou des phrases en pictogrammes (niveau 1) compris par un adulte avec une DI (un participant) particulier (niveau 2), dont la compréhension peut être expliquée/prédite à partir de facteurs au niveau des pictogrammes/phrases et des différences individuelles des adultes

13. Le package R *Generalized Linear Mixed Models using Template Model Builder* (https://github.com/glmmTMB/glmmTMB). Il s'agit d'une alternative au package *lme4* (Bates *et al.*, 2015).

14. Cette recherche a bénéficié de conseils statistiques de la part de la plateforme technologique de Support en Méthodologie et Calcul Statistique (SMCS/LIDAM) de l'Université catholique de Louvain.

avec une DI. Comme nous avons utilisé un *within-subject design*, nous avions bien des mesures répétées pour les mêmes sujets (les participants).

La Figure 6.5 montre des modèles linéaires généralisés à effets mixtes, construits avec nos données, qui prédisent automatiquement la compréhension de six pictogrammes [15] et de deux phrases auprès d'un adulte avec une DI selon un annotateur (A1). [16] Il s'agit ici de P1, qui a un niveau de langue A2. Les pictogrammes et les phrases ont diverses caractéristiques (les métadonnées annotées). Nous employons la fonction de lien par défaut dans le package R *glmmTMB* pour les modèles de régression logistique avec une distribution binomiale (Brooks *et al.*, 2017), la fonction logit. La fonction sigmoïde est ensuite utilisée pour estimer la probabilité d'appartenance à une classe (cf. section 1.5.3) : 0 (« Incorrect / Non compris ») et 1 (« Correct / Compris »).

Pour illustrer la prédiction, nous donnons un exemple. Le modèle prédit correctement – par rapport à nos annotations de référence – que P1 ne comprend pas (0) la phrase « vous devez refaire le bandage » traduite par PictoDr (Ph5, prédiction = 0,494) et la phrase « êtes-vous allergique à des médicaments ? » traduite par Text-to-Picto (Ph9, 0,198), ainsi que les pictogrammes « vous » (Ph5, 0,467), « être » (Ph9, 0,080) et « allergique » (Ph9, 0,124). Il prédit correctement que le pictogramme « (re)faire le bandage » (Ph5, 0,587) est compris (1), mais il ne prédit pas correctement la bonne compréhension (1) des pictogrammes « vous » (Ph9, 0,466) et « médicament » (Ph9, 0,124).

Bien que P1 avait compris la moitié des pictogrammes évalués dans chacune des deux phrases (Ph5 : 1/2 ; Ph9 : 2/4), A1 avait jugé que le sens des deux phrases n'était pas compris (0). Pour la Ph5, P1 avait compris le sens du deuxième pictogramme (« là, il met un bandage au... bras, au poignet »), mais quand la facilitatrice lui avait demandé qui faisait le bandage, il avait répondu : « l'infirmière, le le, la dame, le médecin ». Il s'agit donc d'un contresens par rapport à la phrase source. Pour la Ph9, P1 avait bien dit qu'il voyait un « sirop » et puis un « médicament », mais que la personne (allergique) devait avoir « mal à la tête ou ou ou... en en voyant qu'elle est rouge [...] elle est malade ou... ». Nous allons donc voir comment construire et évaluer ces modèles prédictifs.

15. Nous ne cherchons pas à prédire la compréhension du pictogramme « ? », même si le choix de celui-ci était différent dans les deux types de traduction comme montré dans la Figure 6.4.

16. La formule de ce modèle est la suivante : A1-Picto ou A1-Phrase ~ niveau + traduction + V2 + V3 + V4 + V5 + V6 + V7 + V8 + type_phrase + pos + (1|participant). Les variables sont décrites plus loin.

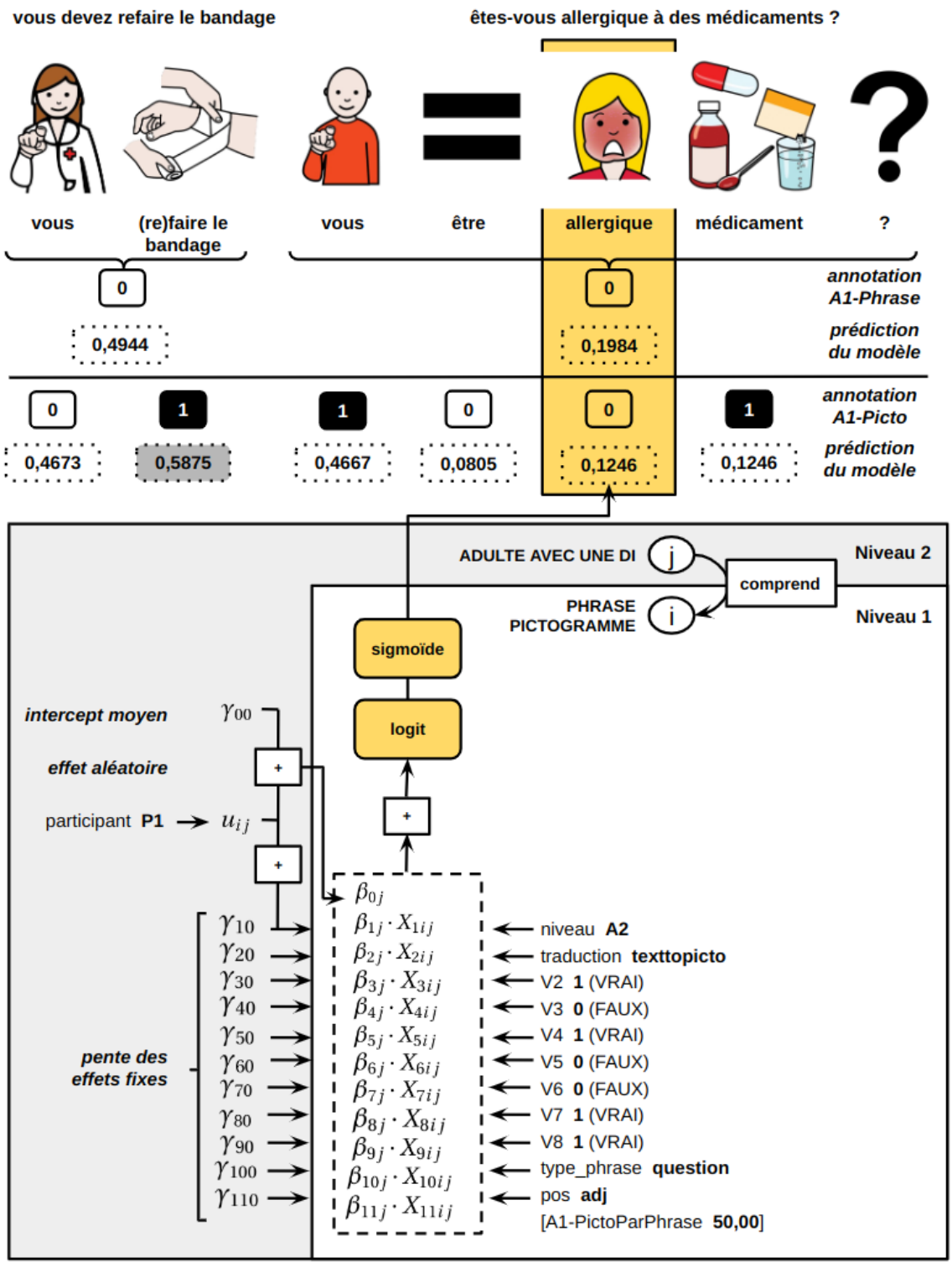


FIGURE 6.5 – Modèles linéaires généralisés à effets mixtes qui prédisent la compréhension des pictogrammes et la compréhension des phrases à partir des annotations (A1-Phrase|A1-Picto) : exemple des phrases « vous devez refaire le bandage » et « êtes-vous allergique à des médicaments ? », ainsi que du pictogramme « allergique » (P1)

Pour identifier les facteurs significatifs qui expliquent le mieux nos données, nous avons employé au moins un modèle linéaire généralisé à effets mixtes par hypothèse – pour la première, nous avons aussi utilisé le test de Student pairé (Howell *et al.*, 2008). Toutes nos variables sont qualitatives (nominales ou ordinales), à l'exception de « A1-PictoParPhrase » et « V1 ». Les variables dépendantes Y de nos modèles sont dichotomiques, ce sont les annotations binaires (0 ou 1) de A1 pour les 683 pictogrammes (A1-Picto) et/ou les 180 phrases (A1-Phrase). Toutes les autres sont les variables indépendantes X. Nous avons d'abord vérifié s'il existait des différences significatives entre les participants. La formule $Y \sim X$ de ce modèle non mixte est la suivante :

(1) A1-Picto|A1-Phrase ~ participant

où les variables sont (la première modalité de la variable explicative X est la référence) :

A1-Picto|A1-Phrase = les annotations de A1, incorrect ou correct : {0, 1}
participant = les participants avec une DI : {P1, P2, P3, P4, P5, P6, P7, P8, P9}

Dans nos modèles mixtes, nous avons toujours inclus un effet aléatoire sur les participants afin d'en tenir compte (1|participant). Nous avons mis un effet fixe sur la variable de l'hypothèse étudiée et sur le niveau de langue. Bien que nous ayons utilisé un test de Student pairé pour vérifier **H1**, nous avons également testé un modèle à effets mixtes destiné à prédire si chaque phrase (A1-Phrase) pouvait être comprise comme la somme de ses pictogrammes :

(2) A1-Phrase ~ A1-PictoParPhrase + niveau + (1|participant)

A1-PictoParPhrase = le pourcentage de pictogrammes corrects par phrase selon A1
niveau = le niveau de langue du participant : {A1, A2, B1}
participant = la constante aléatoire : {P1, P2, P3, P4, P5, P6, P7, P8, P9}

Pour **H2**, nous avons donc remplacé « A1-PictoParPhrase » par le « niveau » de langue dans la formule (2) :

(3) A1-Picto|A1-Phrase ~ niveau + (1|participant)

Pour **H3**, nous avons testé trois adaptations du modèle (3). Dans la première – formule (4) – , nous avons ajouté le système, la « traduction » ; dans la deuxième – formule (5) –, nous avons inclus les 7 variables, qui représentent les distinctions de traduction entre les deux systèmes (cf. section 1.3, Table 6.2) ; dans le troisième modèle – formule (6) –, nous avons évalué la variable quantitative « V1 » seulement au niveau des phrases, puisqu'elle ne fait pas sens au niveau des pictogrammes :

(4) A1-Picto|A1-Phrase ~ traduction + niveau + (1|participant)

(5) A1-Picto|A1-Phrase ~ V2 + V3 + V4 + V5 + V6 + V7 + V8 + niveau + (1|participant)

(6) A1-Phrase ~ V1 + niveau + (1|participant)

traduction = le type de traduction : {Text-to-Picto, PictoDr}
V1 = le nombre de pictogrammes par phrase : {2, 3, 4, 5, 6, 7, 8}
V2 = variable binaire pour l'ordre des pictogrammes : {0, 1}
V3 = variable binaire pour le temps : {0, 1}
V4 = variable binaire pour les (semi-)auxiliaires : {0, 1}
V5 = variable binaire pour les pronoms : {0, 1}
V6 = variable binaire pour les expressions multi-mots : {0, 1}
V7 = variable binaire pour les autres mots différents ou non traduits : {0, 1}
V8 = variable binaire pour les autres mots identiques : {0, 1}

Pour **H4**, les variables ont été remplacées par « type_phrase » dans la formule (5) :

(7) A1-Picto|A1-Phrase ~ type_phrase + niveau + (1|participant)

type_phrase = le type de phrase : {instruction, question}

Pour **H5**, nous n'avons pas calculé la compréhension au niveau des phrases et nous avons remplacé « type_phrase » par « pos » dans la formule (7) :

(8) A1-Picto ~ pos + niveau + (1|participant)

pos = le POS de chaque pictogramme : {adj, adv, interj, n, pro, verb, tp}[17]

1.5.2 Sélection

Pour chacun de nos modèles linéaires généralisés à effets mixtes, nous donnerons : les coefficients de régression et les Odds Ratio (OR) des modalités significatives, la valeur du test du chi-carré (χ^2) si celle-ci est significative, avec le degré de liberté (ddl) et le nombre d'observations (n). L'OR permet de quantifier la taille des corrélations significatives, il désigne l'exponentielle des coefficients de régression : entre 0 et 1, l'effet est négatif ; égal à 1, il n'y a pas d'effet ; au-dessus de 1, l'effet est positif. Nous avons employé la fonction *Anova* du package R *car*, qui effectue par défaut le test du χ^2 de Wald de type II pour les effets fixes d'un modèle linéaire généralisé à effets mixtes. Dans ce contexte-ci, le ddl est le nombre de modalités d'une variable moins 1.

Le coefficient de détermination (r^2 ou R^2) permet de mesurer la qualité globale d'un modèle de régression linéaire. Il n'est cependant pas utilisé pour les modèles mixtes, son utilité étant discutable pour ce type de modèles. Pour mesurer l'adéquation de ces modèles aux données, il existe toutefois différents critères de sélection en fonction de si les modèles sont hiérarchisés ou non. Si on compare des modèles qui

17. Dans l'ordre : adjectif, adverbe, interjection, nom, pronom, verbe, temps (qui n'est pas un POS).

sont imbriqués entre eux, on peut utiliser : 1) le test du rapport de vraisemblance, 2) les critères d'information (AIC, BIC, etc.) – qui fonctionnent également avec des modèles non hiérarchisés –, 3) le bootstrap paramétrique.

Les critères d'information sont différents et ne mènent pas nécessairement aux mêmes conclusions. Afin d'éviter des biais dans le choix du modèle, un seul critère doit être choisi et calculé pour tous les modèles avant de regarder les résultats. Le meilleur sera celui qui obtiendra la plus petite valeur. Nous avons choisi d'utiliser l'*Akaike Information Criterion (AIC)* d'Akaike (1973), dont la définition est proche du *Bayesian Information Criterion (BIC)* de Schwarz (1978). Le BIC comprend un poids supplémentaire sur le nombre de dimensions puisqu'il prend en compte le nombre d'observations.

$$\text{AIC} = -2\ln(\mathcal{L}) + 2d$$
$$\text{BIC} = -2\ln(\mathcal{L}) + d\ln(n)$$

$\mathcal{L}$ = la valeur de la fonction de vraisemblance
d = le nombre de dimensions du modèle ou le nombre total de paramètres estimés
n = le nombre d'observations

Enfin, il convient de noter qu'il y a plusieurs paramètres à vérifier pour tester la validité d'un modèle. Les résidus doivent être analysés de la même façon qu'une régression linéaire simple. Ils sont utilisés pour s'assurer que les hypothèses de linéarité de la régression et d'homogénéité de la variance sont vraies. De plus, il est aussi important de vérifier la multicolinéarité au sein des variables explicatives car il peut y avoir un risque d'obtenir des estimations imprécises des valeurs des coefficients, un surapprentissage ainsi qu'une certaine instabilité lorsque l'on ajoute/retire une variable ou modifie légèrement une observation. On considère qu'il n'y a pas de multicolinéarité si la *Variance Inflation Factor (VIF)* est inférieure à 10. Le VIF est calculé pour chaque effet fixe du modèle de régression.

1.5.3 Prédiction

Une fois que nous aurons comparé les modèles entre eux et sélectionné les deux meilleurs selon l'AIC, nous utiliserons différents outils et mesures pour évaluer leur capacité prédictive. La matrice de confusion est une des façons de visualiser les résultats d'un problème de classification et de prédiction binaire (Table 6.6). Pour effectuer ces calculs, nous emploierons respectivement les fonctions *predict* et *confusionMatrix* des packages R *glmmTMB* (Brooks *et al.*, 2017) et *caret* (Kuhn, 2008).

Vrais Positifs (VP) = le nombre d'items de la classe 0 correctement prédits
Vrais Négatifs (VN) = le nombre d'items de la classe 1 correctement prédits
Faux Négatifs (FN) = le nombre d'items de la classe 0 incorrectement prédits
Faux Positifs (FP) = le nombre d'items de la classe 1 incorrectement prédits

		Référence		**(Total)**
		0	1	
Prédictions	0	VP	FP	VP + FP
	1	FN	VN	FN + VN
(Total)		VP + FN	FP + VN	n

TABLE 6.6 – Matrice de confusion pour la prédiction binaire : exemple

Les éléments de cette matrice de confusion permettent d'analyser la répartition des prédictions d'un modèle par rapport à une référence, ici les annotations de A1 par pictogramme (A1-Picto) et par phrase (A1-Phrase). Le seuil d'appartenance à une classe 1 est fixé à > 0,5 par prédiction variant entre 0 (« Non compris ») et 1 (« Compris »). Il existe plusieurs métriques basées sur les différentes catégories de la matrice de confusion, dont l'exactitude, la spécificité, le rappel (ou la sensibilité), la précision, le F-score (F1) et le coefficient kappa ci-dessous. Le κ de Cohen (1960) permet d'évaluer le niveau d'accord entre les valeurs réelles et les prédictions. Comme son nom l'indique, l'Intervalle de Confiance (IC) à 95 % désigne l'intervalle dans lequel nous pouvons être sûr à 95 % que se situe la vraie valeur – de l'exactitude dans ce cas-ci. À noter que l'IC à 95 % des coefficients de régression donne une indication sur leur significativité ; les variables pour lesquelles 0 appartient à l'IC à 95 % n'ont pas d'effet.

$$\text{Exactitude} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}}$$

$$\text{Spécificité} = \frac{\text{VN}}{\text{VN} + \text{FP}}$$

$$\text{Rappel} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

$$\text{Précision} = \frac{\text{VP}}{\text{VP} + \text{FP}}$$

$$\text{F-score (F1)} = 2 \cdot \frac{\text{Rappel} \cdot \text{Précision}}{\text{Rappel} + \text{Précision}} = \frac{2 \cdot \text{VP}}{2 \cdot \text{VP} + \text{FN} + \text{FP}}$$

$$\kappa = \frac{2 \cdot (\text{VP} \cdot \text{VN} - \text{FN} \cdot \text{FP})}{(\text{VP} + \text{FP}) \cdot (\text{FP} + \text{VN}) + (\text{VP} + \text{FN}) \cdot (\text{FN} + \text{VN})}$$

Notre but est de concevoir un modèle pour prédire la compréhension des pictogrammes – et un autre pour la compréhension des phrases –, qui soit le plus proche de nos données à expliquer (démarche explicative). Nous ferons d'abord les prédictions sur les données qui ont servi à construire ces modèles. Néanmoins, il faut rester prudent quant à la capacité de généralisation de nos modèles car ils pourraient donner des résultats différents sur de nouvelles données. C'est pourquoi, nous testerons également une validation croisée [18] à 10 plis (k = 10, seed = 1234 pour la reproductibilité), avec la version 0.3.3 du package R *cvTools* (Alfons, 2024).

18. Comme l'a décrit François (2011, p. 148), « il s'agit d'une technique d'évaluation de la capacité de généralisation d'un modèle. Elle consiste à appliquer celui-ci sur des données différentes de celles ayant servi à l'entraîner, ce qui permet d'obtenir des résultats sans biais ».

1.6 Résultats

Tout d'abord, nous détaillons les résultats quantitatifs sur la compréhension des pictogrammes et la compréhension des phrases par hypothèse (section 1.6.1). Ensuite, les meilleurs modèles seront utilisés pour la prédiction de la compréhension (section 1.6.2). Enfin, les résultats qualitatifs des dernières questions – 1) la communication, 2) l'utilité des pictogrammes, 3) les améliorations – seront présentés (section 1.6.3).

1.6.1 Compréhension des pictogrammes et des phrases

En ce qui concerne notre **objectif** et **H1**, la Table 6.7 présente les pourcentages de réponses correctes au niveau des pictogrammes et des phrases, par participant (P) et par annotateur (A1 et A2). Chez A1, la moyenne pour les 683 pictogrammes (42,92 %) est plus élevée que celle des échantillons (34,44 % et 25 %), c'est-à-dire l'ensemble de pictogrammes à évaluer obligatoirement, un obligatoire par phrase (#180). Le pourcentage moyen de réponses correctes entre A1 et A2 pour les 180 pictogrammes (29,72 %) est plus élevé que pour les 180 phrases (25,27 %). P8 n'a rien compris.

P	A1-Picto (#683)	A1-Picto (#180)	A2-Picto (#180)	A1-Phrase (#180)	A2-Phrase (#180)
P1	39,75 %	40 %	**40** %	35 %	40 %
P2	54,21 %	**60** %	25 %	**45** %	**55** %
P3	**60,24** %	45 %	35 %	35 %	35 %
P4	51,80 %	50 %	**40** %	15 %	20 %
P5	32,53 %	30 %	05 %	20 %	05 %
P6	40,29 %	25 %	15 %	25 %	25 %
P7	49,25 %	30 %	30 %	25 %	25 %
P8	10,44 %	05 %	00 %	00 %	00 %
P9	47,76 %	25 %	35 %	30 %	20 %
Moy. (A1 ou A2)	**42,92** %	**34,44** %	25 %	**25,55** %	25 %
Moy. (A1 et A2)		29,72 %		25,27 %	

TABLE 6.7 – H1 : Compréhension des pictogrammes et compréhension des phrases par participant selon deux annotateurs (A1-A2)

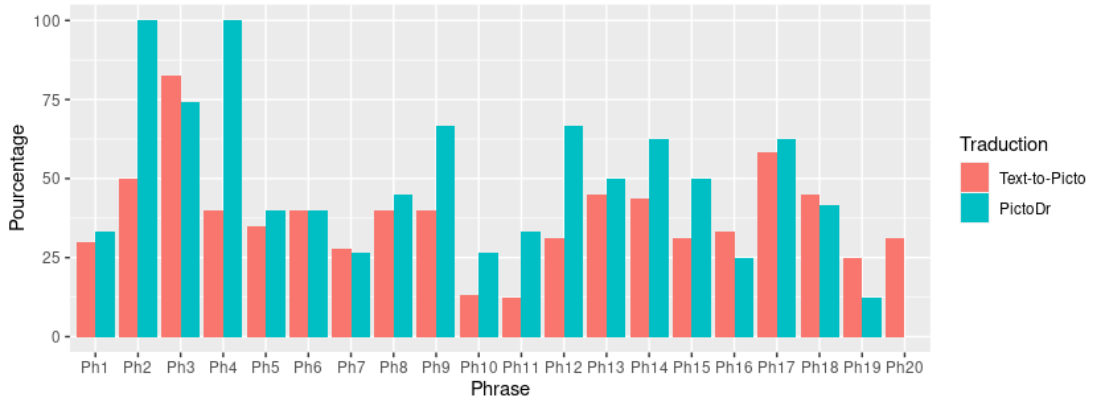
La différence entre les pourcentages A1|A2-Picto (#180) et A1|A2-Phrase (#180) – 29,72 % contre 25,27 % – n'est pas significative en utilisant un test de Student pairé. Avec les 683 pictogrammes annotés par A1 (42,92 %), la différence entre les pictogrammes et les phrases est très significative ($t = 5{,}0326$; $p = 0{,}001$). Selon notre modèle (1), pour les pictogrammes (cf. section 1.5.1), P3 a un impact positif très significatif sur nos résultats (coefficient de régression de P3 de 0,83** ; OR = 2,29 ; χ^2 = (ddl = 8, $n = 683$) = 42,987***). [19] P8 a un impact négatif extrêmement significatif (coefficient de P8 de -1,73*** ; OR = 0,17). Le modèle (2) confirme le test *t*, la variable représentant le pourcentage de pictogrammes corrects par phrase est hautement significative au niveau des phrases (coefficient de 0,07*** ; OR =1,07 ; χ^2 = (ddl = 1, $n = 180$) = 39,003***).

19. Les taux de significativité sont indiqués ainsi : * $p < 0{,}05$; ** $p < 0{,}01$; *** $p < 0{,}001$.

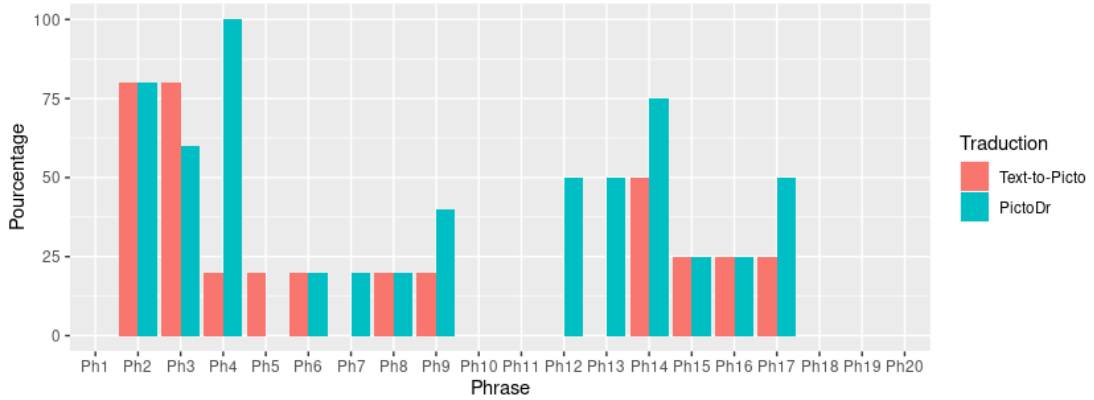
La Table 6.8 présente les 20 phrases (Ph), évaluées dans les deux questionnaires et dont les traductions en pictogrammes sont disponibles dans l'annexe B.1.1.

Ph	Questionnaire 1 | Phrases sources | P1 à P5	Ph	Questionnaire 2 | Phrases sources | P6 à P9
Ph1	bonjour je suis le médecin	Ph11	bonjour je suis l'infirmière
Ph2	nous allons faire des vaccins	Ph12	nous allons écouter votre cœur
Ph3	vous ne pouvez pas boire, pas manger, ni fumer	Ph13	il ne faut pas vous opérer
Ph4	je vais vous prescrire des gouttes	Ph14	je vais examiner votre gorge
Ph5	vous devez refaire le bandage	Ph15	vous allez être hospitalisé
Ph6	mangez-vous beaucoup de viande rouge?	Ph16	êtes-vous confus?
Ph7	avez-vous déjà fait une crise cardiaque?	Ph17	avez-vous les yeux rouges?
Ph8	avez-vous de la fièvre depuis deux jours?	Ph18	avez-vous déjà fumé du tabac?
Ph9	êtes-vous allergique à des médicaments?	Ph19	avez-vous mal depuis deux semaines?
Ph10	devez-vous cracher quand vous toussez?	Ph20	comprenez-vous ce que je dis?

TABLE 6.8 – Corpus de test de l'évaluation de deux traductions



(a) Compréhension des pictogrammes par phrase et par type de traduction



(b) Compréhension des phrases par phrase et par type de traduction

FIGURE 6.6 – Compréhension selon un annotateur (A1)

Les graphiques montrent que la Ph2, la Ph3 et la Ph4 sont les mieux comprises au niveau des pictogrammes isolés (Figure 6.6a) et au niveau des phrases (Figure 6.6b). Le sens global de certaines phrases n'est pas compris du tout (c'est-à-dire par aucun participant), que ce soit avec seulement Text-to-Picto (Ph7, Ph12, Ph13), PictoDr (Ph5), à la fois Text-to-Picto et PictoDr (Ph1, Ph10, Ph11, Ph18, Ph19, Ph20). Les Tables 6.9 et 6.10 présentent les résultats de compréhension par pictogramme avec leur identifiant et leur nom Arasaac, ainsi que leur sens dans nos traductions entre parenthèses.

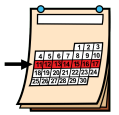
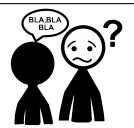
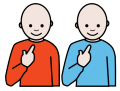
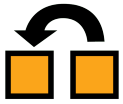
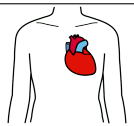
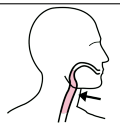
											
9839 passé (avoir déjà) 0/9 0 %	9829 futur (aller) 0/22 0 %	3423 égal (être) 0/22 0 %	8592 goutte (gouttes) 0/5 0 %	24208 atteindre (re/faire) 0/5 0 %	37029 souvent (beaucoup) 0/10 0 %	37752 première fois (déjà) 0/9 0 %	31662 allergie (allergies) 0/5 0 %	32874 quand 0/5 0 %	7244 semaine 0/8 0 %	27363 je n'ai pas compris 0/4 0 %	6531 avoir une idée 0/4 0 %
											
15525 obliger (devoir) 1/14 7,14 %	– causer |pictodr| 1/10 10 %	7185 nous 1/9 11,11 %	37162 encore (déjà) 1/9 11,11 %	7077 depuis 1/9 11,11 %	16693 cracher 2/10 20 %	2808 rouge rouge(s) 2/9 22,22 %	2352 confus 1/4 25 %	35575 (confus)ion mentale 1/4 25 %	30620 avoir mal 1/4 25 %	32982 aidante (infirmière) 1/4 25 %	– vous |pictodr| 14/53 26,41 %
											
6522 bonjour 6/22 27,27 %	6632 je 5/18 27,77 %	37371 jour (jours) 3/10 30 %	6561 médecin 2/5 40 %	24807 viande (rouge) 2/5 40 %	5490 crise cardiaque 4/10 40 %	26508 tousser 2/5 40 %	2628 deux 8/18 44,44 %	6044 marcher (faire) 5/10 50 %	2715 cœur 4/8 50 %	3332 gorge 2/4 50 %	28789 (avoir mal) aux fesses 2/4 50 %

TABLE 6.9 – Compréhension de tous les pictogrammes, les moins compris selon un annotateur (A1)

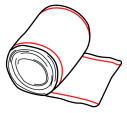
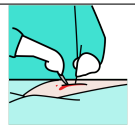
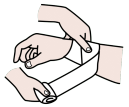
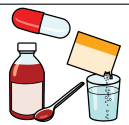
											
9693 dire 2/4 50 %	31390 infirmière 2/4 50 %	2608 vous 38/72 52,77 %	34641 correct (pouvoir) 3/5 60 %	6456 manger 6/10 60 %	3220 plus (ni) 6/10 60 %	6243 bandage 3/5 60 %	35585 tousser 3/5 60 %	32530 fièvre 6/10 60 %	33910 hospitaliser 5/8 62,5 %	– je/nous |pictodr| 18/27 66,66 %	5526 non (ne pas) 19/28 67,85 %
											
30692 ausculter (écouter) 3/4 75 %	5530 opérer 3/4 75 %	27925 opération (opérer) 3/4 75 %	27719 examiner 6/8 75 %	6573 œil (yeux) 3/4 75 %	25010 conjonctivite (yeux rouges) 3/4 75 %	5101 tabac 3/4 75 %	6572 écouter 3/4 75 %	2467 médecin 4/5 80 %	7295 bander 4/5 80 %	2316 viande 4/5 80 %	– allergie |pictodr| 4/5 80 %
											
28413 manger 9/10 90 %	7112 fumer 13/14 92,85 %	30073 vaccin (vaccins) 10/10 **100** %	6061 boire 10/10 **100** %	8163 médicaments 10/10 **100** %	7111 fumer 4/4 **100** %	26366 prescrire 10/10 **100** %	26757 gouttes 5/5 **100** %	3418 ? – –	7217 ? – –		

TABLE 6.10 – Compréhension de tous les pictogrammes, les mieux compris selon un annotateur (A1)

Niveau de langue | Participant	A1-Picto (#683)	A1-Phrase (#180)
A1 | P6-P7-P8	33,33 %	16,66 %
A2 | P1-P5-P9	39,48 %	28,33 %
B1 | P2-P3-P4	**55,42** %	**31,66** %

TABLE 6.11 – H2 : Compréhension des pictogrammes et compréhension des phrases par niveau de langue selon un annotateur (A1)

Pour **H2**, la Table 6.11 montre les pourcentages de réponses correctes – selon A1 – des pictogrammes et des phrases par niveau de langue (A1, A2, B1). Ceux du niveau B1 sont les plus élevés pour les pictogrammes (55,42 %) et pour les phrases (31,66 %). D'après le modèle (3), par rapport à A1, le niveau B1 a un impact positif significatif sur les réponses annotées pour les pictogrammes (coefficient de régression de B1 de 0,99*; OR = 2,68; χ^2 = (ddl = 2, n = 683) = 6,3*), mais pas significatif au niveau des phrases.

Type de traduction	A1-Picto (#683)	A1-Phrase (#180)
Text-to-Picto	39,36 %	20,00 %
PictoDr	**49,63** %	**31,11** %

TABLE 6.12 – H3 : Compréhension des pictogrammes et compréhension des phrases par type de traduction pour tous les participants selon un annotateur (A1)

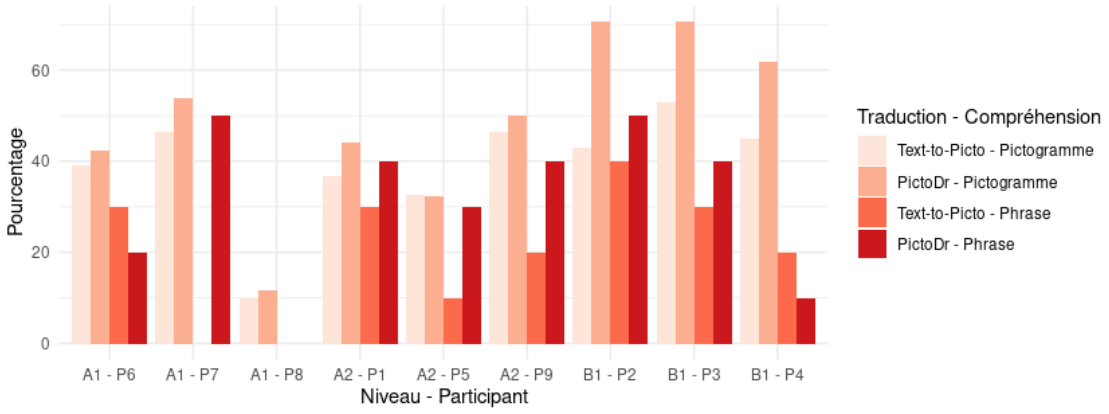


FIGURE 6.7 – Compréhension des pictogrammes et compréhension des phrases par type de traduction et par participant avec le niveau de langue selon un annotateur (A1)

Pour **H3**, la Table 6.12 présente les pourcentages de réponses correctes – selon A1 – des pictogrammes et des phrases par type de traduction (Text-to-Picto, PictoDr) pour tous les participants et la Figure 6.7 les présente par participant avec le niveau de langue. PictoDr a obtenu un score plus élevé pour les pictogrammes (49,63 %) et pour les phrases (31,11 %). Le modèle (4) indique que la variable PictoDr a un impact positif très significatif sur nos résultats, par rapport à Text-to-Picto, pour les pictogrammes (coefficient de régression de PictoDr de 0,44**; OR = 1,54; χ^2 = (ddl = 1, n = 683) = 7,14**), mais pas pour les phrases. Nous avons voulu savoir plus précisément quelles

différences de traduction entre Text-to-Picto et PictoDr étaient significatives. Selon le modèle (5), plusieurs des 7 variables binaires ont un impact significatif sur la compréhension des pictogrammes (coefficients de V2 de -1,76***, de V3 de 0,45* et de V5 de -1,65**; OR = 0,17, 1,58 et 0,19; χ^2 = (ddl = 1, n = 683) = 17,24***, 4,51* et 9,89**) et des phrases (coefficient de V7 de -1,48**; OR = 0,26; χ^2 = (ddl = 1, n = 180) = 6,77**). Le modèle (6) indique que le nombre de pictogrammes par phrase n'est pas significatif.

Type de phrase	A1-Picto (#683)	A1-Phrase (#180)
Instructions	**52,49** %	**36,66** %
Questions	34,50 %	14,44 %

TABLE 6.13 – H4 : Compréhension des pictogrammes et compréhension des phrases par type de phrase selon un annotateur (A1)

Pour **H4**, la Table 6.13 montre les pourcentages de réponses correctes – selon A1 – des pictogrammes et des phrases par type de phrase source (instructions, questions) pour tous les participants. Les instructions ont un pourcentage plus élevé pour les pictogrammes (52,49 %) et pour les phrases (36,66 %). Le modèle (7) confirme que les questions ont un impact négatif extrêmement significatif, par rapport aux instructions, sur la compréhension des pictogrammes (coefficient de régression des questions de -0,80***; OR = 0,44; χ^2 = (ddl = 1, n = 683) = 24,06***) et des phrases (coefficient de -1,28***; OR = 0,27; χ^2 = (ddl = 1, n = 180) = 11,33***).

Catégorie	A1-Picto (#683)	Text-to-Picto (#409)	PictoDr (#274)
Adjectifs [adj] (n = 35)	34,28 %	36,36 % †	30,76 %
Adverbes [adv] (n = 80)	33,75 %	31,91 %	36,36 % †
Interjections [interj] (n = 18)	27,77 %	33,33 % †	22,22 %
Noms [n] (n = 130)	**53,84** %	**51,35** %	**57,14** % †
Pronoms [pro] (n = 183)	42,07 %	43,68 % †	40,00 %
Verbes [verb] (n = 206)	51,45 %	42,27 %	65,06 % †
(Temps [tp] (n = 31))	00,00 %	00,00 %	–

† = Meilleur résultat entre Text-to-Picto et PictoDr

TABLE 6.14 – H5 : Compréhension des pictogrammes par POS selon un annotateur (A1)

Pour **H5**, la Table 6.14 présente les pourcentages de réponses correctes – selon A1 – des pictogrammes par POS pour tous les participants. Nous constatons que les noms sont les mieux compris (53,84 %), suivis par les verbes (51,45 %) et les pronoms (42,07 %). L'ordre des catégories les mieux reconnues est différent entre les traductions des deux systèmes. En top 3, il y a les noms, les pronoms et les verbes pour Text-to-Picto; les verbes, les noms et les pronoms pour PictoDr. Les pictogrammes temporels de Text-to-Picto ne sont pas du tout compris. D'après le modèle (8), la significativité d'un POS spécifique par rapport aux autres dépend du POS de référence. La variable catégorie est significative au niveau des pictogrammes (χ^2 = (ddl = 6, n = 683) = 15,79*).

1.6.2 Prédire automatiquement la compréhension

Afin de déterminer les modèles qui expliquent le mieux la compréhension des pictogrammes et des phrases, nous avons testé plusieurs modèles linéaires généralisés à effets mixtes avec les variables que nous avons décrites dans la section 1.5.1. Tous nos modèles sont montrés dans la Table 6.15. La syntaxe « (niveau|participant) » (dans les formules 18 à 23) représente une pente aléatoire du niveau au sein des participants avec l'intercept corrélé.

Nous avons également testé un encodage one-hot pour toutes nos variables catégorielles (étape de prétraitement), puis construit des modèles linéaires généralisés à effets mixtes comprenant uniquement chacune des variables binaires obtenues. Les modalités significatives ont ainsi été sélectionnées afin de concevoir des modèles adaptés à nos données au niveau des phrases (la formule 24) et des pictogrammes (25).

	Variables dépendantes (à prédire) →	**A1-Picto (#683)**	**A1-Phrase (#180)**
#	**Variables indépendantes (explicatives) – effets fixes et effets aléatoires ↓**	**AIC ↓**	
1	~ participant	896,1	203,7
2	~ A1-PictoParPhrase + niveau + (1|participant)	–	**119,2**
3	~ niveau + (1|participant)	908,8	208,3
4	~ traduction + niveau + (1|participant)	903,6	207,3
5	~ V2 + V3 + V4 + V5 + V6 + V7 + V8 + niveau + (1|participant)	891,3	198,4
6	~ V1 + niveau + (1|participant)	–	207,7
7	~ type_phrase + niveau + (1|participant)	886,1	197,9
8	~ pos + niveau + (1|participant)	867,9	–
9	~ A1-PictoParPhrase + (1|participant)	–	**116,3**
10	~ traduction + (1|participant)	904,3	206,2
11	~ V2 + V3 + V4 + V5 + V6 + V7 + V8 + (1|participant)	893,9	200,0
12	~ type_phrase + (1|participant)	887,0	196,8
13	~ pos + (1|participant)	868,1	–
14	~ A1-PictoParPhrase + traduction + (1|participant)	–	**117,3**
15	~ A1-PictoParPhrase + traduction + niveau + (1|participant)	–	120,2
16	~ traduction + V2 + V3 + V4 + V5 + V6 + V7 + V8 + type_phrase + niveau + (1|participant)	879,4	192,9
17	~ traduction + V2 + V3 + V4 + V5 + V6 + V7 + V8 + type_phrase + pos + niveau + (1|participant)	**828,3**	–
18	~ niveau + (niveau|participant)	NA	NA
19	~ A1-PictoParPhrase + niveau (niveau|participant)	–	NA
20	~ traduction + niveau (niveau|participant)	NA	NA
21	~ V2 + V3 + V4 + V5 + V6 + V7 + V8 + niveau + (niveau|participant)	NA	NA
22	~ type_phrase + niveau + (niveau|participant)	NA	NA
23	~ pos + niveau + (niveau|participant)	–	NA
24	~ V3 + V7 + type_phrase{instruction} + type_phrase{question} + (1|participant)	–	194,9
25	~ niveau[A1] + niveau[B1] + traduction['Text-to-Picto] + traduction[PictoDr] + V2 + type_phrase[instruction] + type_phrase[question] + pos[n] + pos[verb] + (1|participant)	860,6	–

TABLE 6.15 – Comparaison de l'AIC des modèles linéaires généralisés à effets mixtes pour la compréhension des pictogrammes et la compréhension des phrases

Pour tous les modèles, nous avons calculé l'AIC (cf. section 1.5.2). Bien que les formules comprenant seulement les modalités significatives aient un AIC faible par rapport à beaucoup de modèles testés, nous constatons que les meilleurs – ceux qui obtiennent l'AIC le plus bas – sont la formule 17 au niveau des pictogrammes, ainsi que les formules 2, 9 et 14 au niveau des phrases. Nous utiliserons le modèle qui a la meilleure exactitude parmi les trois, le 2. Il est intéressant de noter que le pourcentage de pictogrammes corrects par phrase – A1-PictoParPhrase – est la variable retenue dans tous les modèles avec un faible AIC pour la prédiction de phrases (< 121 contre > 190 pour les autres). Le meilleur modèle pour les pictogrammes comprend toutes nos

variables explicatives. Nous n'avons pas calculé l'AIC des modèles incluant une variable indépendante qui nous semblait moins adaptée à la variable réponse. La valeur 'NA' est renvoyée lorsque le modèle rencontre un problème de convergence. [20]

Pour les modèles linéaires généralisés à effets mixtes 2 (A1-Phrase) et 17 (A1-Picto), nous donnons pour chaque variable explicative les estimations des coefficients de régression (non-)significatifs, les erreurs types, les IC à 95 % des coefficients, l'OR, le ddl, le χ^2 (non-)significatif et le VIF (Table 6.16). Les résidus ont été vérifiés et sont bons (Figure 6.8). Peu de variables sont significatives dans le modèle qui sera utilisé pour la prédiction des pictogrammes, nous voyons que 0 appartient souvent à l'IC des coefficients. Nous ne détectons pas de problème de multicolinéarité des variables (VIF < 10).

Variable dépendante – A1-Picto (#683)							
Effets fixes	**Coefficient**	**Erreur type**	**IC à 95 %**	**OR**	**ddl**	χ^2	**VIF**
intercept	1,64	0,92	[-0,17–3,45]	5,16	–	–	–
traduction	–	–	–	–	1	2,79	1,01
→ traduction[PictoDr]	0,29	0,17	[-0,05–0,63]	1,34	–	–	–
V2	-0,95	0,56	[-2,06–0,15]	0,38	1	2,85	3,84
V3	0,14	0,31	[-0,47–0,76]	1,15	1	0,21	2,58
V4	-0,52	0,37	[-1,25–0,20]	0,59	1	1,96	4,57
V5	-1,59**	0,55	[-2,67–-0,50]	0,20	1	8,29**	2,25
V6	-0,22	0,36	[-0,94–0,50]	0,80	1	0,36	4,17
V7	0,21	0,33	[-0,45–0,87]	1,23	1	0,38	2,44
V8	-0,31	0,38	[-1,07–0,44]	0,73	1	0,64	4,13
type_phrase	–	–	–	–	1	9,90**	4,64
→ type_phrase[question]	-1,17**	0,37	[-1,91–-0,44]	0,30	–	–	–
pos	–	–	–	–	6	15,52	1,61
→ pos[adv]	-0,62	0,48	[-1,57–0,32]	0,53	–	–	–
→ pos[interj]	-1,19	0,70	[-2,57–0,18]	0,30	–	–	–
→ pos[n]	0,22	0,44	[-0,64–1,09]	1,24	–	–	–
→ pos[pro]	-0,19	0,43	[-1,05–0,66]	0,82	–	–	–
→ pos[tp]	-20,89	4871,87	[-9569,60–9527,80]	0,00	–	–	–
→ pos[verb]	0,29	0,44	[-0,58–1,16]	1,33	–	–	–
niveau	–	–	–	–	2	7,12	1,06
→ niveau[A2]	0,44	0,46	[-0,46–1,34]	1,55	–	–	–
→ niveau[B1]	1,22**	0,46	[0,30–2,14]	3,41	–	–	–
Effets aléatoires	**Variance**	**Écart-type**	**IC à 95 %**	**OR**	***n***		
participant (intercept)	0,23	0,48	[0,25–0,94]	1,62	9		

Variable dépendante – A1-Phrase (#180)							
Effets fixes	**Coefficient**	**Erreur type**	**IC à 95 %**	**OR**	**ddl**	χ^2	**VIF**
intercept	-5,52***	0,92	[-7,33–3,70]	0,01	–	–	–
A1-PictoParPhrase	0,07***	0,01	[0,05–0,09]	1,07	1	39,00***	1,01
niveau	–	–	–	–	2	1,15	1,01
→ niveau[A2]	0,67	0,70	[-0,71–2,05]	1,95	–	–	–
→ niveau[B1]	0,13	0,71	[-1,25–1,53]	1,14	–	–	–
Effets aléatoires	**Variance**	**Écart-type**	**IC à 95 %**	**OR**	***n***		
participant (intercept)	0,09	0,30	[0,01–8,28]	1,35	9		

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

TABLE 6.16 – Meilleurs modèles linéaires généralisés à effets mixtes pour prédire la compréhension des pictogrammes et la compréhension des phrases

20. https://cran.r-project.org/web/packages/glmmTMB/vignettes/troubleshooting.html

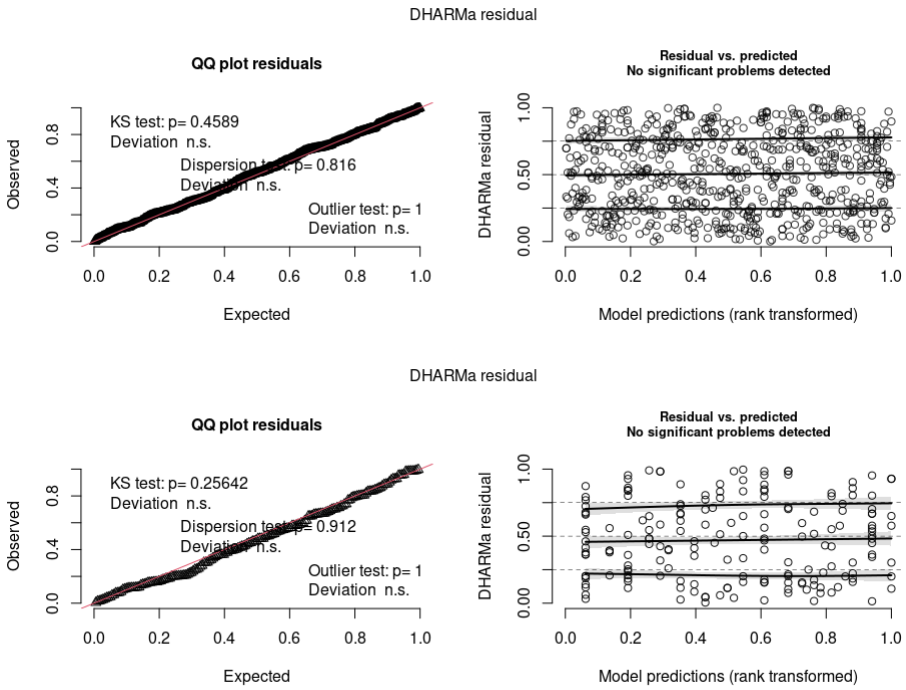


FIGURE 6.8 – Vérification des résidus des meilleurs modèles pour la compréhension des pictogrammes (en haut) et la compréhension des phrases (en bas)

Nous avons utilisé les deux modèles présentés ci-dessus pour prédire automatiquement la compréhension des pictogrammes et la compréhension des phrases (Figure 6.9).[21] Pour les 683 pictogrammes, le modèle 17 a une exactitude de 0,69 ($\kappa = 0{,}36$), l'IC à 95 % (exactitude) de [0,65–0,72], une spécificité de 0,60, un rappel (ou sensibilité) de 0,75, une précision de 0,71 et un F-score de 0,73 ; pour les 180 phrases, le modèle 2 a l'exactitude la plus élevée – parmi les trois meilleurs modèles en termes d'AIC – c'est-à-dire une exactitude de 0,90 ($\kappa = 0{,}71$), l'IC à 95 % de [0,84–0,93], une spécificité de 0,67, un rappel de 0,97, une précision de 0,89 et un F-score de 0,93. Les modèles prédisent que le pourcentage de pictogrammes corrects (39,8 %) est plus élevé que celui pour la compréhension des phrases (18,9 %).

Prédiction de la compréhension des pictogrammes		Référence		
		0	1	Total
Prédictions	0	293 42,9 %	118 17,3 %	**411** **60,2 %**
	1	93 13,6 %	179 26,2 %	**272** **39,8 %**
	Total	**386** **56,5 %**	**297** **43,5 %**	**683** **100 %**

Prédiction de la compréhension des phrases		Référence		
		0	1	Total
Prédictions	0	131 72,8 %	15 8,3 %	**146** **81,1 %**
	1	3 1,7 %	31 17,2 %	**34** **18,9 %**
	Total	**134** **74,5 %**	**46** **25,5 %**	**180** **100 %**

FIGURE 6.9 – Matrices de confusion pour la prédiction de la compréhension des pictogrammes ($n = 683$) et la compréhension des phrases ($n = 180$), sans validation croisée

21. Le code source des analyses et des modèles de prédiction est disponible dans l'annexe C.

Nous avons également effectué une validation croisée à 10 plis comme mentionné dans la section 1.5.3. [22] Nous constatons que les matrices de confusion sans et en validation croisée sont très proches (Figures 6.9 et 6.10). Pour les 683 pictogrammes, le modèle 17 a une exactitude de 0,66 (κ = 0,32), l'IC à 95 % (exactitude) de [0,63–0,70], une spécificité de 0,57, un rappel (ou sensibilité) de 0,73, une précision de 0,69 et un F-score de 0,71 ; pour les 180 phrases, le modèle 2 a une exactitude de 0,87 (κ = 0,64), l'IC à 95 % de [0,81–0,91], une spécificité de 0,67, un rappel de 0,94, une précision de 0,89 et un F-score de 0,91. Les résultats de toutes ces métriques restent donc également très proches. En validation croisée, les modèles prédisent aussi que le pourcentage de pictogrammes corrects (39,9 %) est plus élevé que celui pour la compréhension des phrases (21,7 %).

Prédiction de la compréhension des pictogrammes		Référence		
		0	1	Total
Prédictions	0	285 41,8 %	125 18,3 %	**410** **60,1 %**
	1	101 14,8 %	172 25,1 %	**273** **39,9 %**
	Total	**386** **56,6 %**	**297** **43,4 %**	**683** **100 %**

Prédiction de la compréhension des phrases		Référence		
		0	1	Total
Prédictions	0	126 70,0 %	15 8,3 %	**141** **78,3 %**
	1	8 4,5 %	31 17,2 %	**39** **21,7 %**
	Total	**134** **74,5 %**	**46** **25,5 %**	**180** **100 %**

FIGURE 6.10 – Matrices de confusion pour la prédiction de la compréhension des pictogrammes (n = 683) et la compréhension des phrases (n = 180), en validation croisée

Sans ou avec une procédure de validation croisée, il est plus facile de prédire automatiquement la compréhension (ou non) des phrases en pictogrammes que celle des pictogrammes isolés. Une raison pourrait être que nous avions une classe majoritaire (0) plus importante dans la référence au niveau des phrases (74,5 %), qu'au niveau des pictogrammes (56,5 %). Nous avons calculé le coefficient de corrélation de Pearson entre les valeurs prédites par chaque modèle et les valeurs réelles de compréhension (les annotations 0 et 1 selon A1). Bien que les coefficients soient tous les deux significatifs, la variance expliquée par le modèle pour les pictogrammes est moins élevée (13 %, $r = 0{,}36$; $p < 0{,}01$; $r^2 = 0{,}13$) que pour les phrases (51 %, $r = 0{,}72$; $p < 0{,}01$; $r^2 = 0{,}52$).

Comme l'exactitude des phrases est élevée et que la variable A1-PictoParPhrase revenait dans tous les modèles avec un AIC plus bas, nous avons calculé la corrélation de Pearson entre le pourcentage de pictogrammes corrects par phrase et la compréhension des phrases selon un annotateur. Elle confirme que la taille de l'effet de cette variable est importante ; elle explique 44 % de la variance ($r = 0{,}66$; $p < 0{,}01$). Une façon de le visualiser consiste à utiliser une boîte à moustaches (Figure 6.11). Plus le nombre

22. Pour les pictogrammes, les données d'entraînement de chaque pli (*k-fold*) sont composées de 615 observations et les données de test de 68 observations. Pour les phrases, nous avons respectivement 162 et 18 observations pour les données d'entraînement et les données de test de chaque pli.

de pictogrammes corrects par phrase est élevé, plus il y a de chance que la phrase soit comprise. Cependant, des phrases avec la moitié des pictogrammes corrects (50 %) n'ont pas toujours été jugées comme comprises. C'est par exemple le cas de la phrase « vous devez refaire le bandage » de la Figure 6.5, qui illustrait un contresens.

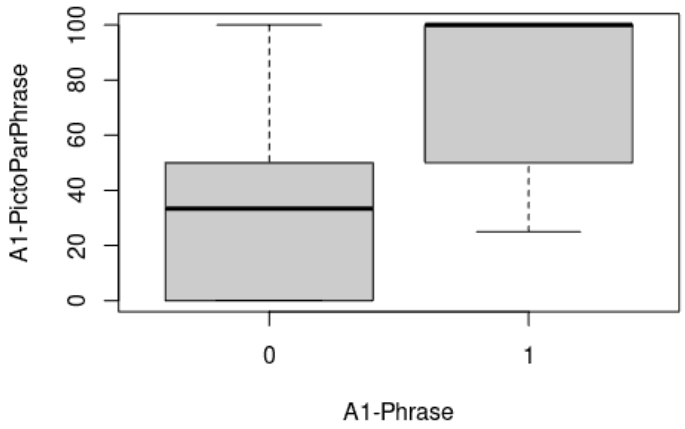


FIGURE 6.11 – Boîte à moustaches entre le pourcentage de pictogrammes corrects par phrase (A1-PictoParPhrase) et la compréhension des phrases (A1-Phrase)

1.6.3 Communication, utilité des pictogrammes et amélioration

Nous détaillons ici les réponses aux trois dernières questions de ces évaluations. La **première** portait sur les difficultés liées à la communication avec le médecin en général. Au total, 7 participants sur 9 ont rapporté qu'ils avaient parfois du mal à comprendre ce que le médecin leur dit (P1 : « des fois oui, des fois non » ; P3 : « ça dépend, ça dépend pourquoi » ; P5 : « pffff oui et non » ; P6 : « des fois oui, des fois c'est pas très compréhensible » ; P7 : « parfois c'est compliqué » ; P8 : « ça dépend, des fois oui des fois non » ; P9 : « oui c'est pour ça que je vais avec ma maman »). Seulement 2 personnes ont dit ne pas avoir de difficultés ou que cela allait (P2 : « non [...] ça va » ; P4 : « non »). En ce qui concerne les difficultés d'expression (et non de compréhension), nous avons que trois réponses : un participant en rencontrait parfois (P1 : « des fois »), mais pas les deux autres (P4 : « non » ; P8 : « pour moi c'est ok »).

La **deuxième** question concernait le ressenti des participants quant à l'utilité des pictogrammes. Des avis étaient positifs (P1 : « oui » ; P3 : « oui parce que des fois on a besoin d'aide en allant chez le médecin » ; P6 : « moi je dirais oui, ça sera utile » ; P8 : « peut être utile, oui »). Cependant, les pictogrammes n'ont pas toujours été jugés nécessaires, des participants ont émis quelques conditions en fonction du public cible et des – phrases en – pictogrammes (P2 : « ça dépend pour les personnes [...] ça pourrait être utile mais s'ils sont plus clairs [c'est-à-dire quand] il y avait moins d'images » ; P4 : « pas forcément » ; P5 : « peut-être pas forcément des pictogrammes... des phrases plus

simples expliquées » ; P7 : « pas toutes [ces phrases] » ; P9 : « oui certaines phrases oui, mais certaines, je, je comprenais pas trop »).

Enfin, la **troisième** question abordait les éventuelles pistes d'amélioration. Certains participants ont parlé des limites des pictogrammes et ce qu'il faudrait améliorer (P2 : « il faudrait qu'ils soient plus explicites, plus clairs dans les pictogrammes [...] on ne comprend pas tout », P3 : « peut-être écrire qu'est-ce que c'est », P4 : « peut-être plus indiquer qu'est-ce qu'il voulait dire [...] un peu clarifier, oui », P5 : « expliquer avec des mots plus simples », P6 : « puis que les médecins ils expliquent mieux ; qu'ils expliquent mieux aussi ce qu'on a et tout ça quand on va chez contrôler tout ça », P7 : « s'il y a trop [de pictogrammes] euh... j'ai de j'ai de la peine » ; P9 : « pour moi ce serait plus facile si on écrit [...] ou que on met sur chaque..., la flèche pour indiquer ce que c'est »). Cette dernière question n'a pas toujours été bien comprise, cela a été le cas pour deux participants – de niveaux de langue A1 et A2 – (P8 et P1). P8 a toutefois reconnu que pour lui c'était plus important d'avoir quelqu'un qui vienne avec lui et qui puisse lui expliquer après plutôt qu'avoir des pictogrammes comme cela.

Nous constatons donc un réel besoin que les médecins expliquent mieux, quelle que soit la façon (oral, écrit, plus simple, avec ou sans images).

1.7 Discussion

Dans cette partie, nous discutons de l'objectif et des hypothèses de recherche, de la compréhension de quelques pictogrammes isolés, du problème de la compositionnalité, de certaines limites qui seront prises en compte dans la deuxième étude et enfin des éléments que nous aurions pu explorer davantage dans cette évaluation-ci.

Objectif et hypothèses de recherche

L'**objectif** principal de cette première étude était d'évaluer la compréhension des pictogrammes et la compréhension des phrases en pictogrammes générés par Text-to-Picto et PictoDr, deux systèmes de traduction automatique, auprès d'adultes avec une DI. Nous avons déterminé plusieurs facteurs qui influencent la compréhensibilité.

Hypothèses | Variables	**A1-Picto (#683)**	**A1-Phrase (#180)**
H1 | Niveau de compréhension : Picto > Phrase (A1-PictoParPhrase)	NA	***
H2 | Niveau de langue : B1 > A1 (niveau)	*	-
H3 | Type de traduction : PictoDr > Text-to-Picto (traduction)	**	-
H4 | Type de phrase : Questions < Instructions (type_phrase)	***	***
H5 | Catégorie : Adj ≠ Adv ≠ Interj ≠ N ≠ Pro ≠ Verb ≠ Tp (pos)	*	NA

* $p < 0{,}05$; ** $p < 0{,}01$; *** $p < 0{,}001$

TABLE 6.17 – Niveaux de significativité des résultats de compréhension des pictogrammes et de compréhension des phrases par hypothèse de l'évaluation de deux traductions selon un annotateur (A1)

Les niveaux de significativité des résultats par hypothèse sont résumés dans la Table 6.17 (H1 à H5). D'abord, nous avons vu qu'il y avait bien une différence significative entre la compréhension de tous les pictogrammes et la compréhension des phrases (**H1**). De plus, nous avons montré que la proportion de pictogrammes correctement compris par phrase est le facteur qui permet de mieux prédire – autrement dit, qui influence le plus – la compréhension des phrases en pictogrammes (cf. Table 6.15). Plus le niveau de langue était élevé, plus les scores de compréhension des pictogrammes étaient élevés (**H2**). Les pictogrammes de PictoDr étaient mieux compris que ceux de Text-to-Picto (**H3**). Certaines différences de traduction étaient significatives. Les participants avaient aussi plus facile à comprendre les instructions en pictogrammes que les questions en pictogrammes (**H4**). Enfin, nous avons constaté des différences significatives de compréhension des pictogrammes par POS représenté (**H5**). Il faut noter que les variables au niveau des pictogrammes étaient plus souvent significatives que les variables au niveau des phrases. Cela pourrait s'expliquer par le nombre élevé de pictogrammes et le nombre de phrases plus faible (683 contre 180). Dans la deuxième étude, nous avons donc fait le choix de ne plus évaluer des pictogrammes isolés pour pouvoir nous concentrer sur plus de phrases comportant moins de pictogrammes. Nous proposons de revenir sur des points liés à chacune de nos hypothèses.

Tout d'abord (**H1**), nous pensons que le fait que les phrases soient plus difficiles à comprendre que les pictogrammes isolés est dû à une sollicitation de la mémoire de travail et une charge cognitive plus élevées dans le cas des phrases. Nous avons aussi vu que le sens d'un pictogramme peut être mal interprété (par exemple, le pictogramme « rouge » qui a été pris pour de la « peinture » ou le pictogramme « vous » représentant le patient et compris comme le « médecin »). La mauvaise interprétation d'un seul pictogramme peut avoir un impact sur la compréhension globale de la phrase. De plus, une phrase étant par définition composée ici de plusieurs pictogrammes, le nombre d'interprétations possibles est plus élevé. Nous avons observé que nos participants avec une DI n'étaient pas toujours certains de leur réponse. Ils en donnaient parfois plusieurs et il arrivait qu'ils répondent un peu au hasard quand la facilitatrice reformulait en leur donnant plusieurs possibilités. Même s'il était parfois difficile de déterminer si les participants comprenaient, les scores inter-annotateurs ont montré que cela restait plus facile de juger de la compréhension des phrases que de celle des pictogrammes.

Bien que le niveau de langue ne caractérise pas la DI (**H2**), nous avons un peu analysé les réponses des participants de niveaux de langue A1 et A2. Par exemple, P6 (A1) et P7 (A1) avaient tendance à moins interpréter ce qu'ils voyaient et à rester au stade de la description. P8 (A1) était extrêmement descriptif. Pour chaque image, il relevait des détails qui n'étaient pas toujours importants pour le sens des pictogrammes et le sens des phrases. Il décrivait chaque personnage (« le premier en rouge » à la place de « je/moi » ou le « patient », « il est habillé en noir, il fait comme ça avec la main »,

« il est pas content », etc.). Il donnait beaucoup d'interprétations assez surprenantes ; des éléments qui n'étaient pas toujours représentés (« elle travaille dans un magasin, une boutique, je sais pas où », « c'est comme un champignon », « un œil ouvert sur le monde »). Au contraire, P1 (A2) donnait des réponses très concises ou répondaient simplement « oui » aux reformulations de sens de la facilitatrice qui essayait d'obtenir des réponses – plus précises – en posant plus de questions. À noter que de temps en temps, elle encourageait les participants à dire quand ils ne savaient pas.

Ensuite, nous avons fait évaluer des traductions de Text-to-Picto et PictoDr, qui se distinguent par différents choix de conception (**H3**). Text-to-Picto est basé sur une ressource générale, alors que PictoDr a été conçu avec une ontologie médicale et des pictogrammes qui ont été validés par plusieurs personnes – les concepteurs et des étudiants en médecine –, voire adaptés pour cet outil, ce qui explique que les pictogrammes de PictoDr ont été mieux compris que Text-to-Picto. En ce qui concerne les différences de compréhension entre les traductions générées par les deux systèmes, toutes les variables n'étaient pas significatives. Cela peut s'expliquer par le fait que nous n'avions pas assez de phrases par type de variables. Bien que le nombre de pictogrammes ne soit pas significatif, des participants ont préféré les phrases avec moins de pictogrammes. Plusieurs ont montré leur découragement à la vue de longues suites de pictogrammes à interpréter. L'ordre des pictogrammes était significatif. Les phrases de PictoDr avec une inversion du sujet n'ont pas été compris, car le sens du pictogramme « causer » représenté par une flèche a été mal interprété dans 9 cas sur 10. L'ordre des pictogrammes n'était pas significatif dans l'étude de Shiojiri *et al.* (2013), mais le public et la langue traités étaient différents. Ils ont obtenu des pourcentages proches avec trois ordres de phrases en pictogrammes différents (sujet + objet + verbe, sujet + objet + verbe, aléatoire). Ils ont expliqué que les japonais avaient plus l'habitude de comprendre des phrases écrites avec des mots ordonnés différemment.

Quant à la différence entre la compréhension des instructions et des questions en pictogrammes (**H4**), cela pourrait s'expliquer par l'originalité de notre approche. À notre connaissance, des phrases complètes en pictogrammes ne sont pas utilisées pour poser des questions médicales. [23] Nous pensons également que la mauvaise compréhension de nos questions en pictogrammes provient d'un manque de contextualisation de notre étude. Les participants avaient des difficultés à s'identifier à un patient et ne voyaient pas la facilitatrice comme un médecin (jeu de rôle). Il y a plusieurs raisons à cela, par exemple, la formulation de la tâche (« que veut dire cette phrase »), [24] voire la mauvaise interprétation des pictogrammes représentant les (inter)locuteurs.

23. Dans les applications mobiles MediPicto AP-HP et SantéBD, le patient sélectionne des phrases représentées par une seule image. Dans le système Glyph (Hill *et al.*, 2016; Zeng-Treitler *et al.*, 2014; Bui *et al.*, 2012), des phrases médicales sont traduites en séquences de plusieurs pictogrammes, mais il s'agit d'instructions de sortie, qui ne sont pas destinées à la communication des personnes avec une DI.

24. La facilitatrice les aidait en rappelant souvent le contexte (phrase du médecin au patient).

Quand nos participants devaient expliquer le sens de nos phrases, ils donnaient rarement leur réponse sous la forme d'une question. Les annotateurs ont décidé que les phrases étaient correctes si le sens donné l'était aussi (indépendamment de la syntaxe).

Enfin, nous avons examiné les POS de tous les pictogrammes évalués après les avoir annotés manuellement (**H5**). [25] Il s'agit d'une approche d'analyse plus proche de celle utilisée dans Text-to-Picto que PictoDr. En effet, même si Text-to-Picto ne génère pas une traduction littérale, il fait appel à un étiqueteur morphosyntaxique contrairement à PictoDr qui a été entraîné sur une glose sémantique, définie indépendamment de la morphosyntaxe. Nous avons fait le choix des POS, car l'annotation nous paraissait plus objective qu'utiliser des catégories sémantiques telles que « entité », « action », « qualificatif », etc. Les catégories les mieux comprises étaient les noms et les verbes (cf. Table 6.14). Ce sont celles qui représentent le sens principal de nos phrases. Ces catégories nous semblent plus facilement représentables que les autres. En évaluant la *translucency* de pictogrammes représentant différents POS – issus de cinq ensembles pour la Communication Alternative et Améliorée (CAA) –, Bloomberg *et al.* (1990) ont montré que les noms étaient plus translucides que les verbes qui étaient plus translucides que les modificateurs, la différence entre les noms et les verbes étant significative. Fuller et Lloyd (1991) avaient aussi mentionné que les POS sont une variable qui peut influencer l'apprentissage de pictogrammes utilisés en CAA (cf. chapitre 1, section 2.3).

Pictogramme, compositionnalité, apprentissage et multimodalité

En ce qui concerne la reconnaissance des pictogrammes isolés (cf. Tables 6.9 et 6.10), nous constatons que les pictogrammes les moins compris (entre 0 et 30 %) sont relativement abstraits. Il s'agit souvent de ceux avec des formes géométriques comme des carrés, des rectangles, des flèches et/ou des ronds qui représentent des verbes (« égal/être », « refaire », « devoir », « causer ») ou la temporalité (les pictogrammes « souvent/beaucoup », « première fois/déjà », « encore/déjà », « depuis »), les calendriers (« semaine » et « jour » interprétés comme la date du « prochain rendez-vous »). Il y a certaines expressions (« je n'ai pas compris », « avoir une idée »), mais aussi des sentiments ou des états (« confus », « avoir mal »). Les pictogrammes les mieux compris (entre 90 et 100 %) représentent des actions fréquentes (« manger », « fumer », « boire »), même s'il y a des concepts médicaux (« vaccin », « médicaments », « prescrire », « gouttes »). Aucun participant n'a été impressionné par des pictogrammes, alors que cela aurait pu être le cas de certains (« opérer », « yeux rouges », etc.).

Cette analyse a montré que représenter le temps (ici « passé » et « futur ») par

25. Cette annotation a quelques limites : un mot peut avoir plusieurs POS en fonction du contexte (« deux » est un adjectif ou un nom), on peut accepter plusieurs appellations par pictogramme (avec différents POS, par exemple, (avoir|faire) mal = souffrir), l'annotation était un peu plus subjective pour les pictogrammes représentant les expressions multi-mots (quel est le POS principal).

des horloges – comme plusieurs auteurs l'avaient proposé (Sevens, 2018; Imam *et al.*, 2019; Norré *et al.*, 2022) – perturbe plus qu'autre chose les personnes qui n'ont jamais appris à utiliser des pictogrammes. Par exemple, les réponses par participant pour le pictogramme « passé » étaient presque toutes descriptives : « horloge midi » (P1), « on doit attendre en salle d'attente, le temps d'attendre, l'horloge tourne » (P2), « prends un rendez-vous » (P3), « une horloge qui indique [...] midi, minuit, il montre que c'est le sens contraire de l'horloge » (P4), « horloge, un réveil, une montre au bras, une horloge accrochée » (P5), « une horloge » (P6), « je sais pas » (P7), « réveil » (P8), « rendez-vous, l'horloge, pause, l'heure, l'heure tourne » (P9). Quand la facilitatrice leur demandait d'expliquer le sens de la phrase avec leurs mots, ils pensaient à l'horloge ou l'agenda en tant qu'objet et non en tant que valeur temporelle ou propriété grammaticale.

L'apprentissage des images et la possibilité de les enlever dans ces technologies, autrement dit, leur adaptation en fonction des capacités et besoins individuels reste donc essentielle. En cherchant quel était le meilleur modèle pour prédire la compréhension (cf. section 1.6.2), nous avons vu qu'il y avait beaucoup de facteurs qui ont une influence sur la compréhensibilité des pictogrammes. Le meilleur facteur au niveau de la phrase reste la compréhension des pictogrammes qui la composent même s'il ne faut pas négliger les contresens qui peuvent être dus à la mauvaise compréhension d'un seul pictogramme. La représentation d'un pictogramme – ou d'une séquence de pictogrammes – peut être ambiguë, c'est-à-dire suggérer un sens et son contraire. Dans le chapitre 1 (cf. section 2.3, Figure 1.9), nous avions d'ailleurs vu que certains pictogrammes se comprennent – et donc s'apprennent – les uns par rapport aux autres. Ici, nous avons montré qu'un peu moins d'un pictogramme sur deux a été compris et une phrase sur quatre (cf. Table 6.7), respectivement 42,92 % et 25,27 % pour les types de traduction des deux systèmes (Text-to-Picto et PictoDr). Cela peut s'expliquer par le fait que nos participants avec une DI ne connaissaient pas ces pictogrammes.

Bien que nos participants étaient tous verbaux et n'utilisaient pas d'images pour communiquer, nous avons vu qu'ils avaient parfois des difficultés à comprendre ce que le médecin leur dit. Certains ont mentionné que les pictogrammes pouvaient être utiles pour d'autres personnes et ont évoqué la piste de la multimodalité (image + texte ou oral) pour améliorer la compréhension. L'explicitation et la clarté des images doivent aussi être améliorées. Dans la prochaine section, nous verrons d'ailleurs que des changements ont été effectués au sein du système PictoDr depuis cette étude. Ces modifications prennent en compte nos observations et nous les évaluerons. Il convient de noter que nous avions déjà présenté à certains participants (P3 à P9) une première version contextuelle des traductions de PictoDr (cf. Figure 6.2). Avec nos données, nous n'avons pas pu analyser si les participants comprenaient mieux la version contextuelle (Figure 6.2b) par rapport à la version séquentielle (Figure 6.2a).

Limites et perspectives

Les limites méthodologiques communes entre cette étude-ci et la suivante seront développées dans la discussion générale (cf. section 3). Nous pouvons toutefois déjà mentionner ici certaines pistes d'amélioration ou des éléments qui auraient pu être évalués. Par exemple, le corpus constitué des retranscriptions post-éditées aurait pu être davantage exploité en vue d'une (méta-)analyse plus linguistique : le nombre de disfluences, les types d'erreurs, de phrases ou de mots utilisés par les adultes avec une DI (antonymes, hyperonymes, hyponymes, etc.), s'ils ont reçu de l'aide ou le nombre de questions supplémentaires de la facilitatrice pour obtenir la bonne réponse (par pictogramme et/ou par phrase). En ce qui concerne les modèles explicatifs de la compréhension, d'autres variables liées aux pictogrammes auraient pu être analysées. Par exemple, le type de représentation selon la taxonomie de Zeng-Treitler *et al.* (2014), présentée dans le chapitre 1 (cf. section 2.1) ; la compositionnalité des pictogrammes médicaux, la compréhension des pictogrammes dans un domaine particulier et s'il existe un lien avec les caractéristiques sociodémographiques des participants comme l'avait testé Cantini (2023) avec des concepts relatifs à la santé de la femme (en gynécologie, chirurgie obstétrique, etc.). Plus d'annotations et d'annotateurs auraient été nécessaires, en plus d'un guide d'annotation pour chaque nouvelle métadonnée.

Il convient aussi de noter qu'il s'agissait d'un test de compréhension de phrases indépendantes, sans lien les unes avec les autres et qui ne constituaient donc pas un véritable dialogue. Or, les systèmes de traduction vers les pictogrammes tels que nous les envisageons ici sont conçus pour traduire une anamnèse dans une situation de communication réelle : un patient adulte avec une DI, accompagné ou non, qui vient consulter un médecin pour des symptômes spécifiques. Nous pensons qu'un contexte médical précis pourrait améliorer les scores de compréhension. C'est pourquoi, nous proposons une deuxième évaluation basée sur un contexte plus réaliste. Le choix d'un *within-subject design* a pu aussi amener de la confusion chez nos participants. Par exemple, le fait d'alterner aléatoirement différents pictogrammes pour représenter les changements des (inter)locuteurs à chaque phrase. Les mêmes phrases sources étant également répétées, cela nous éloignait d'autant plus d'une situation réelle. Dans un ou deux cas, des participants se sont souvenus avoir vu des pictogrammes légèrement différents pour les mêmes concepts et ont essayé de comprendre le sens des phrases les unes par rapport aux autres. Par exemple, P9 a dit « là, je viens de trouver pourquoi [...] il avait des yeux rouges, c'est de la peinture [...] comment il s'est mis la peinture dans les yeux » (cf. Annexe B.6, Figure B.1b). Le design expérimental a pu avoir un impact – non désiré – sur certaines réponses. C'est pour ces raisons que nous adapterons un peu notre méthodologie.

2 Évaluation d'une anamnèse en pictogrammes

Nous proposons une deuxième étude quantitative et qualitative avec une méthodologie d'évaluation très proche de la précédente et un contexte plus authentique. Nous utilisons ici les phrases en pictogrammes générées par un seul système de traduction automatique : PictoDr. Il ne s'agit plus d'évaluer des phrases indépendantes, mais une anamnèse sur un cas précis. Il aurait été plus difficile de sélectionner plusieurs dizaines de phrases sur une même pathologie, dont les traductions automatiques en pictogrammes seraient toutes correctes, en plus d'être différentes à la fois entre les deux systèmes et entre les deux études. De plus, nous avons montré que les pictogrammes des traductions médicales de PictoDr étaient mieux compris par des adultes avec une DI que ceux de Text-to-Picto (cf. Table 6.12). À noter que les phrases de PictoDr étaient aussi mieux comprises, mais la différence n'était pas significative.

2.1 Objectif et hypothèses de recherche

Cette étude ne vise pas non plus à mesurer la précision et le rappel d'un système de traduction vers les pictogrammes comme c'est le cas des évaluations automatiques. Il ne s'agit pas non plus d'une évaluation en situation réelle. L'**objectif** est d'évaluer la compréhension de phrases traduites en pictogrammes sans texte auprès d'adultes avec une DI dans un contexte précis. En effet, il s'agit uniquement de questions posées lors d'une anamnèse du COVID-19. Nous avons vu qu'une limite de l'évaluation des deux traductions avait été le manque de contexte. Nous avons choisi la maladie du COVID-19, car la crise sanitaire qui a débuté en février 2020 était mondiale. Tout le monde a été impacté d'une façon ou d'une autre – même ceux qui n'ont pas eu le virus. Les personnes en situation de handicap en ont particulièrement subi les conséquences (Unia, 2020). De plus, nous avons déjà montré, via d'autres tests utilisateurs, qu'employer des pictogrammes pour répondre aux questions ouvertes du médecin pouvait être une solution pour clarifier un problème médical – rencontré par des patients allophones –, tel que le COVID-19 (Norré *et al.*, 2021a; Bouillon *et al.*, 2023). Nous proposons donc de nouvelles hypothèses de recherche adaptées à ce cas-ci.

- **H1 :** La compréhension des phrases (questions) en pictogrammes est influencée par la catégorie de question liée au COVID-19 (symptômes, paramètres vitaux, critères de vulnérabilité, habitudes de vie, exposition).
- **H2 :** Les phrases (questions) avec moins de pictogrammes sont mieux comprises que les phrases (questions) avec plus de pictogrammes.

- **H3 :** Les adultes avec une DI qui n'utilisent pas les pictogrammes pour communiquer perçoivent ces pictogrammes comme utiles pour une consultation médicale avec le médecin ou à l'hôpital.

Pour répondre à notre objectif, nous discuterons d'abord du score de compréhension global au niveau des phrases en pictogrammes, c'est-à-dire de la moyenne de tous nos participants avec une DI. Ensuite, nous analyserons les données qui peuvent expliquer les hypothèses que nous avons formulées pour les raisons ci-dessous.

H1 Lors de la crise sanitaire, nous avons noté que les médecins ont dû poser diverses questions en lien avec la maladie du COVID-19, afin de la dépister. C'est pourquoi, nous pensons qu'il pourrait y avoir des différences de compréhension en fonction du type d'information demandé. Par exemple, les questions plus spécifiques liées aux paramètres vitaux ou à l'exposition (épidémiologique) pourraient être plus difficilement compréhensibles en pictogrammes que celles plus générales sur les habitudes de vie.

H2 Nous avons précédemment évalué l'influence du nombre de pictogrammes par phrase (V1), mais nous n'avons pas obtenu de résultat significatif. Or, nos participants avec une DI avaient préféré les phrases avec moins de pictogrammes. Nous voulons donc savoir si la différence est significative avec plus de phrases évaluées.

H3 Dans la première étude, nous avons également vu que bien que nos participants étaient tous verbaux, plusieurs avaient trouvé les pictogrammes utiles pour eux ou pour d'autres personnes (cf. section 1.6.3). Il nous semble donc intéressant de déterminer – toujours de façon qualitative – si c'est aussi le cas après un contexte plus réaliste.

2.2 Participants

Il s'agit d'une nouvelle collaboration avec le bureau Langage simplifié de Pro Infirmis Fribourg et la facilitatrice de l'étude décrite dans la section 1.2. Les tests ont été réalisés en mai-juin 2024, soit un an plus tard. Elle a eu le même rôle, à savoir : recruter les participants, les questionner et évaluer leur compréhension de chaque phrase. Le public cible était constitué de 10 adultes avec une DI (Table 6.18, P1 à P10), 4 hommes (H) et 6 femmes (F), dont 7 personnes ayant déjà participé à la première évaluation. Nous avons 2 participants de niveau A1 et 1 de niveau A2, car beaucoup de participants sont B1 (les 3 nouveaux, 3 anciens et P3 qui a progressé de niveau depuis notre première étude). Tous nos participants savent lire. Nous n'avons pas plus d'informations sur les types de handicap. L'âge des participants variait entre 28 et 47 ans (moyenne : 35,6 ; écart-type : 7,35), 5 participants sont dans la tranche des [26–35] ans et 5 dans celle des [36+]. Avant chaque test, nous leur avons demandé leur consentement oral et ils ont chacun signé notre formulaire de consentement écrit adapté en langage simplifié (cf. annexe B.5).

P	Genre	Âge	Niveau de français	Handicap	Commentaires (handicap, autres façons de communiquer, etc.)	Déjà utilisé les pictogrammes? Si oui, fréquence?	Déjà participé en 2023?
P1	H	44	A2	Intellectuel	Pas de spécificité reconnue	Oui, pas souvent	Oui
P2	F	28	A1	Intellectuel	Pas de spécificité reconnue	Non	Oui
P3	F	29	B1	Intellectuel	Pas de spécificité reconnue	Oui, pas souvent	Oui
P4	F	40	B1	Intellectuel	Pas de spécificité reconnue	Non	Oui
P5	H	47	A1	Intellectuel	Trisomie	Oui, de temps en temps	Oui
P6	H	38	B1	Intellectuel	Pas de spécificité reconnue	Non	Oui
P7	F	29	B1	Intellectuel	Pas de spécificité reconnue	Non	Non
P8	H	30	B1	Intellectuel	Pas de spécificité reconnue	Oui, pas beaucoup	Non
P9	F	29	B1	Intellectuel	Pas de spécificité reconnue	Oui, des fois	Oui
P10	F	42	B1	Intellectuel	Pas de spécificité reconnue	Non	Non

H = Homme | F = Femme

TABLE 6.18 – Description des participants de l'évaluation d'une anamnèse

2.3 Design expérimental

La facilitatrice a fait passer 10 entretiens individuels. Chaque participant était enregistré avec un smartphone (audio seulement). Les phrases en pictogrammes étaient présentées individuellement au fur et à mesure sur un ordinateur portable au centre de l'écran sans texte. Tous les pictogrammes avaient la même taille quelle que soit la longueur de la phrase. En ce qui concerne les phrases évaluées dans cette étude, nous avons adapté une partie du formulaire COVID-19 des Hôpitaux Universitaires de Genève (HUG), créé le 27 février 2020 (cf. annexe B.7). Nous l'avions déjà utilisé de la même façon pour des tests utilisateurs avec des pictogrammes Arasaac (Norré *et al.*, 2021a; Bouillon *et al.*, 2023) : pour chaque point à vérifier chez un patient suspect au COVID-19 (SARS-CoV-2), nous avons gardé la formulation proposée par les médecins pour le domaine COVID-19 du système de traduction BabelDr (Bouillon *et al.*, 2021).

Au total, nous avons un questionnaire de 30 questions (Table 2.3) : 18 liées aux symptômes (section A du formulaire des HUG), 3 concernant les paramètres vitaux et signes cliniques (section B), 3 sur les critères de vulnérabilité ou facteurs de risque (section C), 2 au sujet des habitudes de vie (section D), 4 à propos de l'exposition (section E). [26] Nous n'avons retenu que des phrases dont les traductions automatiques en pictogrammes nous semblaient correctes. Pour une ou deux questions, les résultats ont été post-édités; des pictogrammes ont dû être créés ou adaptés pour certains concepts, par exemple, « *Tachypnea* » (19B), ainsi que « *Recent* » (27E). Il s'agit principalement de questions fermées, dont les réponses sont « oui » ou « non ».

26. Seule la phrase de test n'était pas une question (« bonjour je suis le médecin »). Elle était montrée au début, ne comptait pas dans les résultats et servait à expliquer la tâche aux participants.

Ph	Questionnaire Phrases sources P1 à P10	Ph	Questionnaire Phrases sources P1 à P10	Ph	Questionnaire Phrases sources P1 à P10
1A	avez-vous le nez qui coule ?	11A	avez-vous la diarrhée ?	21B	à combien était votre température la dernière fois que vous l'avez mesurée ?
2A	avez-vous mal à la gorge ?	12A	avez-vous mal au ventre ?	22C	avez-vous eu un accident vasculaire cérébral ?
3A	avez-vous des douleurs musculaires ?	13A	êtes-vous fatigué ?	23C	avez-vous eu une chimiothérapie ?
4A	avez-vous des frissons ?	14A	avez-vous des difficultés à respirer ?	24C	fumez-vous ?
5A	avez-vous une toux grasse ?	15A	avez-vous mal à la poitrine ?	25D	avez-vous trop de tension artérielle ?
6A	est-ce sanglant quand vous toussez ?	16A	avez-vous mal à la tête ?	26D	avez-vous été vacciné contre le covid ?
7A	devez-vous cracher quand vous toussez ?	17A	avez-vous des allergies (connues) ?	27E	avez-vous voyagé récemment ?
8A	avez-vous de la fièvre ?	18A	à quelle date ont commencé les symptômes ?	28E	avez-vous été en contact avec une personne confirmée positive au coronavirus dans les quatorze derniers jours ?
9A	avez-vous des troubles de l'odorat ?	19B	respirez-vous trop rapidement ?	29E	à quelle date avez-vous été en contact avec quelqu'un de positif au coronavirus ?
10A	avez-vous des nausées ?	20B	avez-vous pris votre température ?	30E	étiez-vous en quarantaine en tant que contact proche d'un cas confirmé de coronavirus ?

A = Symptômes | B = Param. vitaux | C = Crit. de vulnérabilité | D = Habitudes de vie | E = Exposition

TABLE 6.19 – Corpus de test de l'évaluation de l'anamnèse du COVID-19

Nous avons fait traduire automatiquement ces questions en pictogrammes avec le système PictoDr (cf. annexe B.8.1), une version de début 2024 (Gerlach *et al.*, 2024). Des changements ont été réalisés depuis l'étude précédente. Le système ne propose plus que la version contextuelle. Le pictogramme « vous » de PictoDr (la combinaison des pictogrammes Arasaac « médecin » féminin et « vous ») est toujours affiché à gauche avec une bulle de bande dessinée à droite où se trouve la phrase traduite en pictogrammes (Figure 6.12). L'intérieur de la bulle étant vert, on distingue maintenant la délimitation individuelle de chaque pictogramme avec leur fond blanc respectif. Dans le système en ligne quand plusieurs pictogrammes sont possibles pour un même concept, l'utilisateur peut sélectionner le pictogramme qu'il préfère via un bouton spécifique en-dessous. Les concepts de l'*Unified Medical Language System (UMLS)* sont affichés en haut, en anglais, mais nous ne les afficherons pas dans ces tests.

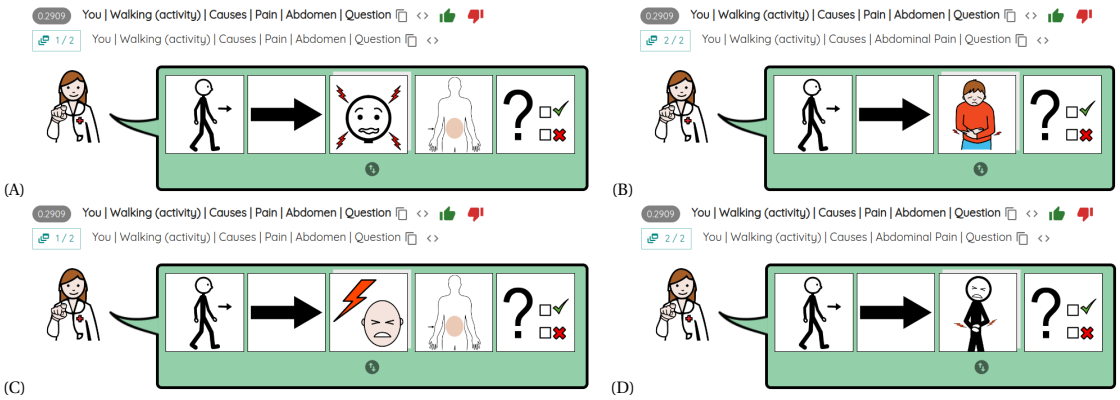


FIGURE 6.12 – Traductions de PictoDr : exemple de la phrase « avez-vous mal au ventre quand vous marchez? »

Pour certaines phrases, dont « avez-vous mal au ventre quand vous marchez? » ci-dessus (Figure 6.12), un bouton (1/2 ou 2/2) à côté de la glose sémantique – la séquence de concepts séparés par des '|' – permet de choisir entre une version compositionnelle (par défaut, ici il y a deux pictogrammes pour traduire « mal au ventre » dans A et C) ou non-compositionnelle (un seul dans B et D). De nouveaux pictogrammes continuent d'être créés manuellement par la Faculté de Traduction et d'Interprétation de l'Université de Genève, c'est le cas pour les concepts UMLS « *Abdomen* » (dans A et C), « *Pain* » (« douleur » dans C), mais aussi « *Abdominal Pain* » (« douleur abdominale » dans D) représenté par un personnage noir unisexe. Nous avons toujours sélectionné la version non-compositionnelle avec des nouveaux pictogrammes pour les phrases qui existaient dans les deux versions, par exemple, pour les expressions « avoir mal » + « partie du corps humain » (« gorge », « ventre », « poitrine », « tête »), c'est-à-dire les concepts UMLS « *Pain* » + ceux dont le type sémantique est « *Body Location or Region* » ou « *Body Part, Organ, or Organ Component* ».[27] Nous avons aussi privilégié les picto-

27. Nous avions envisagé de comparer la compréhension des versions compositionnelles et non-

grammes unisexes quand c'était possible afin que tous les participants puissent plus facilement s'auto-identifier à la situation de communication.

Le fait d'évaluer un système et pas deux – et ce, par un même participant – devrait les rendre moins confus puisque les différentes stratégies de traductions (par exemple, le choix des pictogrammes pour les pronoms) ont pu les déstabiliser et ainsi avoir un impact sur nos résultats de compréhension. Le nombre de pictogrammes par phrase est moins élevé. Le minimum est 1 pictogramme par phrase (dans la bulle verte) et le maximum 4. Le nombre moyen de pictogrammes par phrase avec le pictogramme « ? » (des questions oui/non) – est 2,36 ; celui sans ce pictogramme est de 1,46. Au moins 40 pictogrammes différents ont été présentés dans cette étude. Comme il s'agissait d'une anamnèse reproduite à partir d'un formulaire électronique, nous avons décidé de ne pas randomiser l'ordre des phrases pour chaque participant. Les documents destinés à la facilitatrice ont tous été légèrement adaptés et simplifiés, par exemple, le questionnaire (cf. annexe B.8.2), le guide (cf. annexe B.9) et les instructions (cf. annexe B.10).

Pour améliorer la contextualisation, nous avons aussi introduit la multimodalité (des gestes). Pour chaque phrase traduite en pictogrammes : 1) la facilitatrice devait dire « je suis le médecin » en se montrant du doigt, puis en montrant le pictogramme représentant le médecin, dire « vous êtes le patient », le désigner et montrer la bulle verte ; 2) les 10 participants avaient pour tâches d'expliquer le sens de la phrase (« qu'est-ce que je veux savoir en tant que médecin » à la place de « que veut dire cette phrase » pour éviter le lien avec une troisième personne). Si la réponse au deuxième point était incorrecte, qu'il ne répondait pas ou qu'il ne savait pas : 3) elle devait donner la bonne réponse et poser deux questions supplémentaires (« est-ce que le(s) pictogramme(s) aide(nt) à comprendre le mot ou la phrase » et « comment comprenez-vous le mot ou la phrase avec le(s) pictogramme(s) »), afin de vérifier l'origine de l'erreur. Au total, la facilitatrice a fait évaluer 300 phrases, soit 120 de plus que dans l'autre étude.

Enfin, des questions sur la connaissance de la maladie du COVID-19, l'utilité et les améliorations possibles de ces phrases en pictogrammes étaient également posées, la première au début de l'entretien, la deuxième et la troisième à la fin. Les deux dernières questions sont identiques à celles de la précédente étude (cf. section 1.3).

1. Savez-vous ce que c'est la maladie du COVID-19, pouvez-vous la décrire ?

2. Aimez-vous utiliser ces (traductions en) pictogrammes ? Ou aimeriez-vous continuer à utiliser ces pictogrammes lors d'une consultation médicale avec le médecin ou à l'hôpital ? Pourquoi ? Pensez-vous que c'est utile ?

3. Et que faut-il améliorer ? Qu'est-ce qui était (plus) difficile à comprendre ?

compositionnelles comme l'avait fait Cantini (2023). Néanmoins, nous avons constaté que seules 7 phrases sur les 30 sélectionnées pour notre anamnèse existaient en version compositionnelle dans PictoDr. La probabilité d'obtenir des résultats significatifs étant faible, nous n'évaluerons pas cette variable.

2.4 Annotation et données collectées

Contrairement à l'autre étude, nous n'avons pas retranscrit manuellement les entretiens. Le système de reconnaissance vocale Whisper (Radford *et al.*, 2023) a été réutilisé. Nous avons post-édité les retranscriptions automatiques (ajout des locuteurs, répétitions, disfluences, etc.). En nous basant sur ces retranscriptions post-éditées, nous avons déterminé pour les 300 phrases si le sens donné par le participant était correct (codé en « 1 ») ou incorrect (« 0 »). Nous nous sommes référé autant que possible aux mots de chaque participant ou aux reformulations de la facilitatrice approuvées par le participant. Pour chaque phrase, nous avons aussi annoté si le participant avait reçu de l'aide de la facilitatrice et combien de questions supplémentaires elle avait dû poser pour essayer d'arriver à la réponse finale (« 0 », « 1 », « 2 » ou « 3++ »). Ces informations ont eu un impact sur notre annotation afin de la rendre la plus objective possible.

La compréhensibilité a été évaluée au seul niveau des phrases. L'évaluation a été réalisée par deux annotateurs : nous, à partir des retranscriptions (A1) ; la facilitatrice, durant l'expérience (A2). Toutes les phrases ont été évaluées par les deux annotateurs (Table 6.20). Nous n'avons pas annoté les pictogrammes, car les phrases étant très courtes (20 sont composées de 2 pictogrammes, dont celui représentant toujours le «? »), les résultats auraient souvent été identiques pour les phrases. De plus, nos variables indépendantes n'étaient pas directement liées aux pictogrammes. A2 avait un peu plus d'expérience avec les phrases en pictogrammes puisque qu'elle avait participé à la première étude qui était plus complexe que celle-ci.

	Questionnaire | P1 à P10 - Total	**A1**	**A2**
Arasaac			
# Phrases (PictoDr & Questions)	30 | 300	**300**	**300**
# Pictogrammes (bulle sans «? »)	44 | 440		
# Pictogrammes (bulle avec «? »)	71 | 710		
# Pictogrammes (médecin + bulle avec «? »)	101 | 1 010		

TABLE 6.20 – Nombre de phrases et nombre de pictogrammes Arasaac du questionnaire lié au COVID-19, évalués ou non par les deux annotateurs (A1-A2)

Des accords inter-annotateurs ont été calculés avec nos annotations et celles de la facilitatrice par phrase et pour toutes les phrases (Figure 6.13). Nous avons utilisé la même échelle d'annotation : « Correct » et « Incorrect ». Si le participant ne savait pas répondre, l'annotateur devait indiquer « Incorrect ». Les coefficients κ et α sont de 0,82, soit un accord presque parfait [0,81–1,00] selon l'échelle de Landis et Koch (1977) et un bon accord [0,81–1,00] selon Krippendorff (2004). Les coefficients sont donc plus élevés que ceux dans la première étude (cf. section 1.4). A1 a annoté un peu plus sévèrement la compréhension qu'A2 : 8 % de phrases incorrectes selon A1 et correctes selon A2 contre seulement 1 % de phrases correctes selon A1 et incorrectes selon A2.

Exactitude des phrases		A2		
		Correct	Incorrect	Total
A1	Correct	135 45,0 %	3 1,0 %	**138** **46,0 %**
	Incorrect	24 8,0 %	138 46,0 %	**162** **54,0 %**
	Total	**159** **53,0 %**	**141** **47,0 %**	**300** **100 %**

FIGURE 6.13 – Matrice de confusion avec les catégories « Correct » et « Incorrect » (A1), ainsi que les catégories « Correct » et « Incorrect » (A2) d'exactitude au niveau des phrases (n = 300)

2.5 Modèles linéaires généralisés à effets mixtes

Pour répondre à nos deux premières hypothèses, nous réutilisons la version 1.1.8 du package R *glmmTMB* (Brooks *et al.*, 2017), afin de concevoir des modèles linéaires généralisés à effets mixtes (cf. section 1.5). Pour rappel, nous évaluons la compréhension des phrases en pictogrammes selon un annotateur (A1) – Y la variable dépendante dichotomique {0, 1} –, afin de vérifier l'impact de différents facteurs – X les variables indépendantes (effets fixes et effets aléatoires) –, avec la formule suivante : $Y \sim X$. Le modèle non mixte (9) vérifie s'il existe des différences significatives entre les participants.

(9) A1-Phrase ~ participant

où les variables sont (la première modalité de la variable explicative X est la référence) :

participant = les participants avec une DI : {P1, P2, P3, P4, P5, P6, P7, P8, P9, P10}

Pour **H1**, nous analysons s'il y a des différences significatives entre les catégories de questions liées au COVID-19 avec le modèle (10) ci-dessous. Pour **H2**, le modèle (11) est utilisé en vue d'évaluer l'influence du nombre de pictogrammes ; le modèle (12) combine toutes ces variables (**H1**, **H2**).

(10) A1-Phrase ~ section + (1|participant)

(11) A1-Phrase ~ V1 + (1|participant)

(12) A1-Phrase ~ section + V1 + (1|participant)

participant = la constante aléatoire : {P1, P2, P3, P4, P5, P6, P7, P8, P9, P10}
section = les sections du formulaire COVID-19 des HUG : {A, B, C, D, E}
V1 = le nombre de pictogrammes par phrase dans la bulle sans « ? » : {1, 2, 3}

2.6 Résultats

Tout d'abord, nous détaillons les résultats quantitatifs sur la compréhension des phrases pour H1 à H2 (section 2.6.1). Les résultats qualitatifs des trois questions – 1) les connaissances préalables des participants sur la maladie du COVID-19, 2) l'utilité des pictogrammes (H3), 3) les améliorations – seront présentés par après (section 2.6.2).

2.6.1 Compréhension des questions

En ce qui concerne notre **objectif**, la Table 6.21 présente les pourcentages de réponses correctes au niveau des phrases, par participant (P) et par annotateur (A1 et A2). D'après A1 et A2, le meilleur est P9 (71,66 % en moyenne) et le moins bon est P5 (16,66 %). La moyenne est de 49,5 %, soit presque 1 phrase sur 2 est comprise. Dans la première étude, nous avons obtenu 25,55 % pour les 180 phrases (questions et instructions) de Text-to-Picto et PictoDr selon A1 (cf. Table 6.7) ; 31,11 % pour les 90 phrases de PictoDr (Table 6.12), mais sur d'autres phrases.

Selon notre modèle (9), P9 a un impact positif significatif sur nos résultats (coefficient de régression de P9 de 1,25* ; OR = 3,50 ; χ^2 = (ddl = 9, n = 300) = 26,86**). [28] P5 a un impact négatif significatif (coefficient de régression de P5 de -1,79* ; OR = 0,16). [29] Ces résultats ne sont pas surprenants : P5 (ex-P8) avait déjà obtenu les pires scores au niveau des pictogrammes et des phrases dans l'autre étude (cf. Table 6.7) ; P9 (ex-P2) avait obtenu les meilleurs scores au niveau des phrases. [30] Le plus grand écart entre les scores des annotateurs concerne P1, qui a un niveau de langue A2. Cela pourrait s'expliquer par le niveau plus bas ; P2 et P5 de niveau A1 ont aussi un écart-type élevé.

Participant	A1-Phrase (#300)	A2-Phrase (#300)	Moyenne	Écart-type	Rang
P1	40,00 %	56,66 %	48,33 %	11,78	7
P2	43,33 %	56,66 %	50,00 %	09,42	6
P3	56,66 %	60,00 %	58,33 %	**02,35**	4
P4	53,33 %	66,66 %	60,00 %	09,42	3
P5	10,00 %	23,33 %	16,66 %	09,42	10
P6	53,33 %	56,66 %	55,00 %	**02,35**	5
P7	63,33 %	60,00 %	61,66 %	**02,35**	2
P8	36,66 %	40,00 %	38,33 %	**02,35**	8
P9	**70,00** %	**73,33** %	**71,66** %	**02,35**	**1**
P10	33,33 %	36,66 %	35,00 %	**02,35**	9
Moyenne	46,00 %	53,00 %	49,50 %	05,42	

TABLE 6.21 – Compréhension des phrases de PictoDr par participant selon deux annotateurs (A1-A2)

28. Les taux de significativité sont indiqués ainsi : * $p < 0{,}05$; ** $p < 0{,}01$; *** $p < 0{,}001$.

29. À noter que nous avons calculé le coefficient de corrélation intra-classe du modèle vide, il est d'environ 0,10 ; ce qui signifie que 10 % de la variabilité s'explique par la structure hiérarchique.

30. P1 (ex-P1, cf. Table 6.7), P2 (ex-P6), P3 (ex-P9), P4 (ex-P3), P5 (ex-P8), P6 (ex-P4) et P9 (ex-P2).

Pour aller plus loin, nous proposons de comparer les scores de nos deux évaluations (2024 et 2023). La Figure 6.14 présente les pourcentages de compréhension des phrases (questions) de PictoDr par participant (pourcentage sur 30 phrases), en plus de leurs résultats à l'étude précédente (les questions et instructions traduites par PictoDr sur 10 phrases ou par PictoDr et Text-to-Picto sur 20 phrases). Les 7 personnes qui ont participé à la première évaluation ont un score plus élevé (ou équivalent) à cette étude-ci au niveau des phrases (P5, P2, P1, P6, P4, P3, P9). P7 – qui n'avait pas participé – a le deuxième score le plus élevé, tandis que P10 et P8 sont antépénultième et pénultième (33,33 et 36,66 %). Il faut aussi noter que le fait d'avoir participé ou non à la première étude ne semble pas avoir d'impact significatif sur les résultats, d'après un modèle mixte qui modélise A1-Phrase ~ étude2023 + (1|participant), où « étude2023 » est une variable binaire pour la participation (1) ou non (0) à notre étude de 2023.

Bien que nous ayons que deux participants de niveau A1 et un de niveau A2, nous avons ré-utilisé le modèle (3) A1-Phrase (cf. section 1.5.1), afin de vérifier le niveau de langue pour les phrases. Par rapport à A1, le niveau B1 a un impact positif significatif sur les réponses annotées (coefficient de régression de B1 de 1,17*; OR = 1,51; χ^2 = (ddl = 2, n = 300) = 6,31*). Dans la première étude, la différence entre le niveau A1 et B1 était significative au niveau des pictogrammes, mais pas au niveau des phrases.

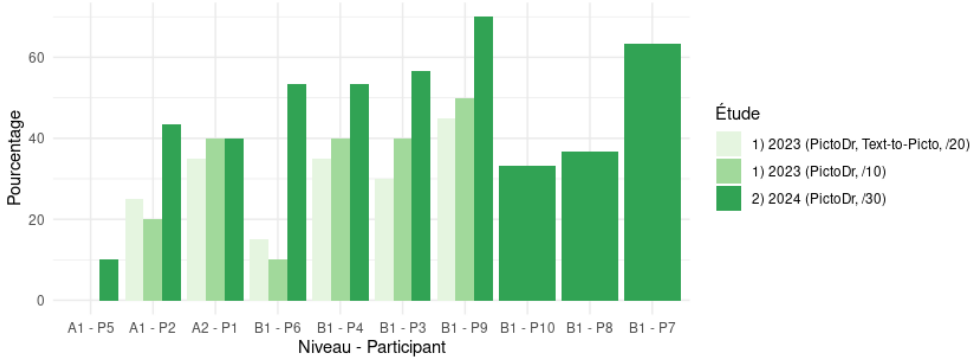


FIGURE 6.14 – Compréhension des phrases par participant avec le niveau de langue selon un annotateur (A1), résultats des deux études utilisateurs (2023 : instructions et questions de PictoDr et Text-to-Picto ou de PictoDr; 2024 : questions de PictoDr)

Pour **H1** (cf. section 2.1), la Table 6.22 montre les résultats par section du formulaire COVID-19 des HUG. Les phrases de la section D (Habitudes de vie) sont les mieux comprises, suivies par celles des sections A (Symptômes) et C (Critères de vulnérabilité). Les moins comprises sont les sections E (Exposition), puis B (Paramètres vitaux). Pour A1, le modèle (10) montre que par rapport à la section A de référence, les sections B et E ont un impact négatif extrêmement significatif (coefficient de régression de B de -2,26***; OR = 0,10; χ^2 = (ddl = 4, n = 300) = 34,751*** et coefficient de E de -2,12***; OR = 0,11).

La Figure 6.16 présente les pourcentages de compréhension par phrase (1-30), incluant la section (A-E) et par annotateur (A1 ou A2). Le graphique montre que les

Section COVID-19	A1-Phrase (#300)	A2-Phrase (#300)	Moyenne	Écart-type	Rang
A | Symptômes ($n = 180$)	55,00 %	63,33 %	59,16 %	5,89	2
B | Paramètres vitaux ($n = 30$)	13,33 %	10,00 %	11,66 %	2,35	5
C | Critères de vulnérabilité ($n = 30$)	50,00 %	50,00 %	50,00 %	**0,00**	3
D | Habitudes de vie ($n = 20$)	**70,00** %	**80,00** %	**75,00** %	7,07	**1**
E | Exposition ($n = 40$)	15,00 %	27,50 %	21,25 %	8,83	4

TABLE 6.22 – H1 : Compréhension des phrases de PictoDr par section COVID-19 pour tous les participants selon deux annotateurs (A1-A2)

phrases 2A et 24C sont les mieux comprises (« avez-vous mal à la gorge ? » et « fumez-vous ? »), deux phrases avec un pictogramme. Le sens global de certaines phrases n'est pas compris du tout (c'est-à-dire par aucun participant), que ce soit seulement selon A1 (30E), à la fois pour A1 et A2 (7A et 20B). Pour 7A (« devez-vous cracher quand vous toussez ? »), la raison vient du fait que pour presque tous les participants, le pictogramme « cracher » représente un « thermomètre » et donc de la « fièvre » ou de la « température » (à cause des lignes représentant le mouvement). Pour 20B (« avez-vous pris votre température ? »), la phrase est confondue avec « avez-vous de la température ? ». Pour 30E (« étiez-vous en quarantaine en tant que contact proche d'un cas confirmé de coronavirus ? », Figure 6.15), A2 a estimé que P1 (-p) avait compris. Bien que la réponse de ce participant était correcte, A1 l'a jugée incorrecte, car P1 a reçu beaucoup d'aide de la part de la facilitatrice (-e), plus de deux questions supplémentaires (« ça peut être quoi ? », « je veux savoir si tu as dû faire quoi ? », « à cause de quoi ? »).

```
[30 étiez-vous en quarantaine en tant que contact proche d'un cas confirmé de coronavirus ? (être en
quarantaine | vous avez été dans le passé | être en contact avec une personne qui a la covid | ?)]

-e : et la dernière phrase
-p : c'est la la maison
-e : hum hum
-p : rouge c'est l'heure non ?
-e : hum hum oui
-p : à moi... con contaminé puis voilà
-e : hum hum ça peut être quoi ? si tu penses au covid
-p : c'est rester à à la maison
-e : rester à la maison donc ça veut dire quoi être à la maison, si, je veux savoir si tu as du faire quoi ?
-p : être en confinement
-e : hum hum à à cause de quoi ?
-p : à cause du... con contaminé
-e : quelqu'un t'as contaminé ou tu étais en contact avec quelqu'un
-p : oui
-e : et puis tu dois rester à la maison
-p : oui
-e : hum hum ok donc ça c'était la dernière phrase
```

FIGURE 6.15 – Retranscription : exemple d'aide de la facilitatrice (P1, phrase 30E)

Enfin (Figure 6.16), on peut aussi noter que les plus grands écarts par phrase entre les deux annotateurs (A1 < A2) sont de : 30 % (10A et 29E). La 21B est la seule phrase où A1 (30 %) > A2 (20 %). Pour les 10A et 29B, ce sont des tous des cas où A1 a estimé que le sens de la phrase était incorrect et A2 correct. A1 a noté que la facilitatrice les avait aidé, en posant une ou deux questions supplémentaires. Nous reviendrons sur ces résultats dans la discussion (cf. section 2.7).

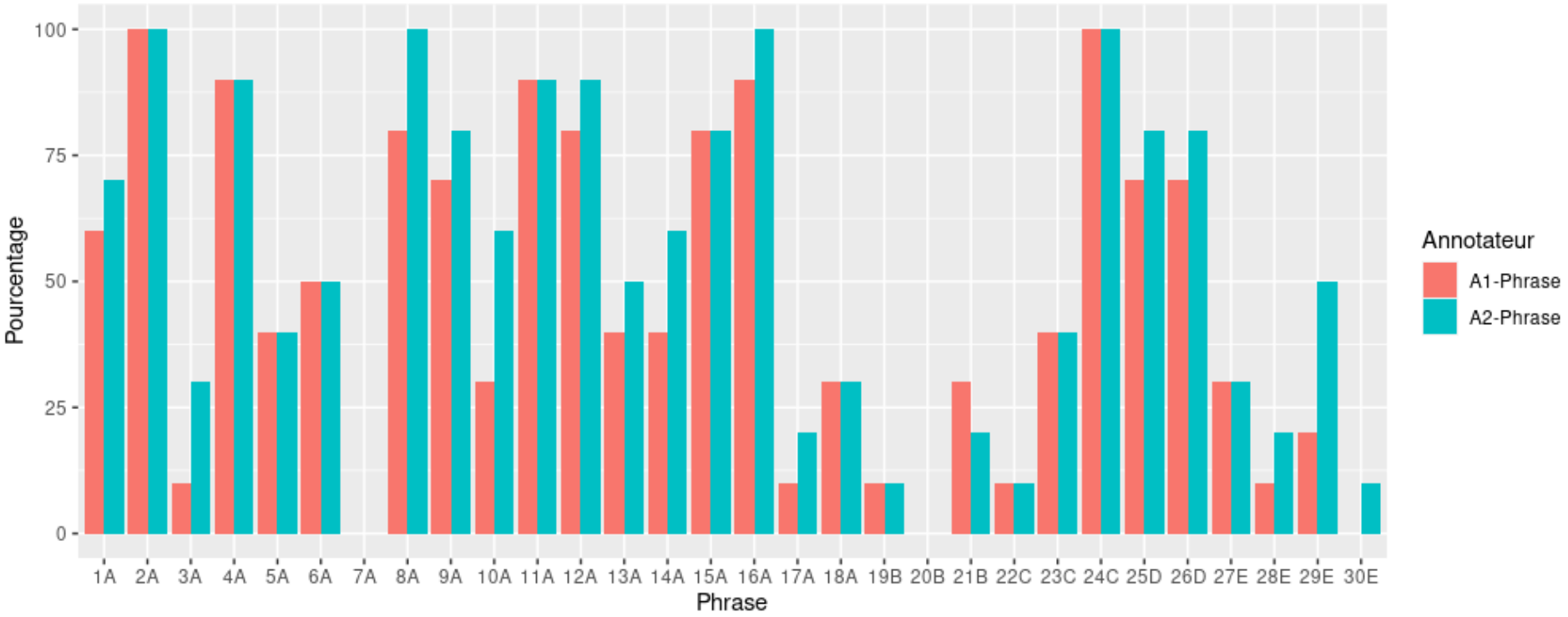


FIGURE 6.16 – Compréhension des phrases de PictoDr par phrase avec la section COVID-19 et par annotateur

Nombre de pictogrammes	A1-Phrase (#300)	A2-Phrase (#300)	Moyenne	Écart-type
1 (n = 200)	**55,00** %	**60,00** %	**57,50** %	**3,53**
2 (n = 60)	38,33 %	50,00 %	44,16 %	8,24
3 (n = 40)	12,50 %	22,50 %	17,50 %	7,07

TABLE 6.23 – H2 : Compréhension des phrases de PictoDr par nombre de pictogrammes par phrase pour tous les participants selon deux annotateurs (A1-A2)

Pour **H2**, la Table 6.23 présente les pourcentages de réponses correctes – selon A1 et/ou A2 – des phrases par nombre de pictogrammes dans la bulle sans « ? » (1, 2, 3). Ceux avec un seul pictogramme sont les plus élevés (57,5 % en moyenne selon les deux annotateurs). Plus le nombre de pictogrammes augmente, moins les phrases sont comprises (17,5 %). Selon notre modèle (11) pour A1, un nombre de pictogrammes plus élevé a en effet un impact négatif extrêmement significatif sur la compréhension des phrases en pictogrammes (coefficient de régression de V1 de -1,01***; OR = 0,36; χ^2 = (ddl = 1, n = 300) = 24,797***).

Variable dépendante – A1-Phrase (#300)							
Effets fixes	**Coefficient**	**Erreur type**	**IC à 95 %**	**OR**	**ddl**	χ^2	**VIF**
intercept	2,09***	0,49	[1,12–3,07]	8,13	–	–	–
section	–	–	–	–	4	34,42***	1,65
→ section[B]	-2,75***	0,59	[-3,92–-1,59]	0,06	–	–	–
→ section[C]	-0,67	0,43	[-1,52–0,17]	0,50	–	–	–
→ section[D]	1,74**	0,60	[0,56–2,92]	5,73	–	–	–
→ section[E]	-0,60	0,59	[-1,77–0,55]	0,54	–	–	–
V1	-1,44***	0,30	[-2,043–-0,83]	0,23	1	22,01***	1,65
Effets aléatoires	**Variance**	**Écart-type**	**IC à 95 %**	**OR**	***n***		
participant (intercept)	0,71	0,84	[0,46–1,53]	2,32	10		

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$ || A = Symptômes | B = Param. vitaux | C = Crit. de vulnérabilité | D = Habitudes de vie | E = Exposition

TABLE 6.24 – Modèle linéaire généralisé à effets mixtes pour prédire la compréhension des phrases de PictoDr pour une anamnèse du COVID-19

Le modèle (12) pour A1, présenté à la Table 6.24 et qui est plus complet, confirme que la catégorie de question liée au COVID-19 et le nombre de pictogrammes par phrase ont un impact significatif sur la compréhension des phrases (**H1**, **H2**). À noter que les sections qui ont une influence significative ne sont pas les mêmes que dans le modèle mixte (10), ici la catégorie D (Habitudes de vie) remplace E (Exposition).

Afin d'élargir l'étude, nous avons analysé le temps des évaluations. La Table 6.25 montre la durée des tests par participant, trié du plus rapide (rang 1) au plus lent (10). Le « Temps 1 » est la durée des enregistrements qui comprend les 30 phrases à évaluer. Le « Temps 2 » est la durée totale des enregistrements qui comprend les 30 phrases et les trois questions ouvertes (une au début, deux à la fin). La Figure 6.17 montre la compréhension des 30 questions de PictoDr selon A1 et le Temps 1 en secondes de chaque participant avec son niveau de langue.

Le plus rapide est P9 et le plus lent P5, soit les participants qui obtiennent respectivement les meilleurs et les pires résultats selon les deux annotateurs (71,66 % et

Participant	Temps 1 | Min. : sec.	Temps 1 | Sec.	Temps 2 | Min. : sec.	Temps 2 | Sec.	Rang
P1	22 : 16	1 336	23 : 51	1 431	6
P2	25 : 19	1 519	27 : 09	1 629	8
P3	20 : 07	1 207	22 : 19	1 339	4
P4	13 : 00	780	15 : 07	907	2
P5	29 : 40	1 780	31 : 24	1 884	10
P6	21 : 01	1 261	24 : 08	1 448	5
P7	18 : 08	1 088	19 : 57	1 197	3
P8	27 : 43	1 663	29 : 07	1 747	9
P9	**9 : 07**	**547**	**10 : 32**	**632**	**1**
P10	22 : 27	1 347	23 : 56	1 436	7
Moyenne	20 : 52	1 252,8	22 : 45	1 365	

Min. : sec. = Minute : seconde || Sec. = Seconde

TABLE 6.25 – Durée de l'évaluation de l'anamnèse par participant

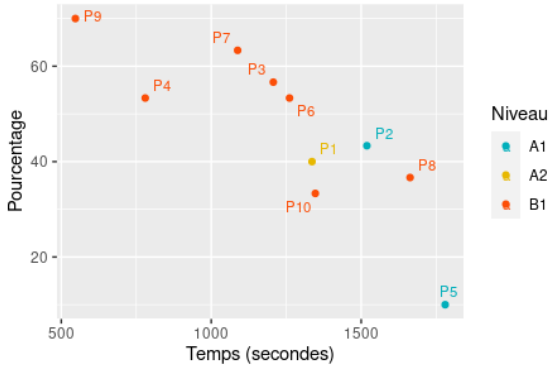


FIGURE 6.17 – Compréhension des phrases de PictoDr et temps en secondes par participant avec le niveau de langue selon un annotateur (A1)

16,66 %). P5 et P2 (dernier et avant-avant dernier) ont un niveau de langue A1. Il y a un faible écart de temps entre les deux participants suivants, P1 (A2) et P10 (B1). Quant à P8 (avant-dernier, B1), il a un débit de parole relativement lent par rapport aux autres.

Le classement des participants, du plus rapide au plus lent, est le même entre le Temps 1 et le Temps 2, mais il est un peu différent du classement des participants par score de compréhension, du plus élevé au plus bas (cf. Table 6.21). Nous notons que les premiers (P9), quatrièmes (P3), cinquièmes (P6) et dixièmes (P5) participants sont identiques. Pour confirmer cet effet, nous avons testé le modèle mixte A1-Phrase ~ temps + (1|participant), où « temps » est la durée de l'évaluation en secondes (Temps 1).

Il indique que plus le temps augmente, plus il y a un très petit impact négatif significatif sur la compréhension des phrases selon A1 (coefficient de régression du temps de -0,001*; OR = 0,99; χ^2 = (ddl = 1, n = 300) = 4,80*). Bien qu'un lien entre la vitesse et la compréhension semble évident, ces résultats peuvent également s'expliquer par la méthode de notre deuxième étude. Si la réponse du participant était fausse, la facilitatrice devait poser des questions supplémentaires et expliquer le sens correct de la phrase pour vérifier l'origine de l'erreur et nous aider à comprendre comment le participant comprenait les phrases en pictogrammes.

2.6.2 COVID-19, utilité des pictogrammes et amélioration

Nous détaillons maintenant les réponses à la première question qui a servi à contextualiser l'étude, ainsi qu'aux deux dernières questions des tests. La **première** portait sur les connaissances de la maladie du COVID-19. Différentes descriptions ont été données : « une petite... bête ou je sais pas quoi dis... toute petite comme ça... qui se promène dans le euh dans le corps » (P1), « alors la maladie du COVID hum c'est... je crois quand quand les animaux ils vont tout près c'est là qu'on choppe le virus [...] il y a aussi les animaux mais les êtres humains [...] peut-être hein » (P5).

Certains participants ont évoqué le fait d'avoir des « (graves) symptômes » (P2, P6, P7). Des participant ont énoncé spontanément (ou avec l'aide de la facilitatrice) : « avoir mal à la tête, euh je crois, après je crois que on a plus d'odorat, je pense qu'on a mal au ventre, euh on a mal à la tête, on est fatigué, on a plus d'odorat et euh euh on a de la fièvre » (P3), « des toux grasses, des rhumes ou des bobos » (P4), « soit euh la soit soit la toux soit la fièvre soit euh soit la gorge enflammée » (P5), « perte d'odorat ou de goût et puis euh mal à la tête euh vomissement euh rhume des fesses, de la fièvre, mal au cou » (P6). P7 a raconté sa propre expérience : « moi j'avais eu de la toux, du rhume, mal aux muscles, un peu de fièvre ; j'avais eu deux fois le COVID et euh j'étais vaccinée contre le COVID aussi ». P9 a fait remarquer que : « ça dépend de chaque personne ».

Des participants ont énoncé des particularités et/ou des conséquences de la maladie : « on peut tomber malade ou on peut aussi... transmettre... aux autres personnes » (P1), « si on présente des symptômes on doit faire euuh le test [...] et après faut attendre euh les résultats [...] si tu euh, si tu l'as ben il faut te mettre en en isole... en isolement » (P2), « une bactérie [...] qui se transmet » (P6). La facilitatrice a parfois posé une ou deux questions supplémentaires plus précises pour avoir un peu plus d'informations de la part des participants : « oui et puis on peut tomber malade de ça ? » (à P1), « qui fait quoi ? » (à P4), « hum hum les animaux ou pas seulement les animaux ? » (à P5), etc.

La Table 6.26 résume toutes les réponses des participants concernant la « maladie du COVID-19 ». Nous les avons regroupées en trois catégories : une première générale sur ce qu'est le COVID-19 (1-4), une deuxième sur les symptômes (5-21), une troisième plus liée à la crise sanitaire (22-27). Bien que chaque personne avait sa propre définition, nous retrouvons des caractéristiques communes entre les participants, des moins précises/correctes que d'autres (4, 18, 19, 20, 27), mais aussi beaucoup qui auront finalement fait l'objet de questions en pictogrammes, en particulier, celles liées aux symptômes, au vaccin et à la transmission de la maladie (les sections A, D et E du formulaire des HUG). Des participants ont fourni des explications plus ou moins longues : entre 1 (pour P10) à 10 (pour P6) caractéristiques du COVID-19. La moitié ont relevé la « fièvre » (5). Les autres qualificatifs fréquents sont : « un virus » (1), la présence de « symptômes » (6), dont la « toux » (7) et la « perte de l'odorat » (8).

Caractéristiques attribuées au COVID-19	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**#**
1) Un virus (qui est fort)	–	–	–	v	v	–	v	–	–	–	**3**
2) Une bactérie	–	–	–	–	–	v	–	–	v	–	2
3) Une maladie chronique | grave	–	–	–	–	–	–	–	v	–	–	1
4) Une toute petit bête qui se promène dans le corps	v	–	–	–	–	–	–	–	–	–	1
5) Fièvre	–	–	v	–	v	v	v	–	v	–	**5**
6) (Graves) symptômes	–	v	–	–	–	v	v	–	–	–	3
7) Toux (grasses)	–	–	–	v	v	–	v	–	–	–	3
8) Perte de l'odorat	–	–	v	–	–	v	–	–	v	–	3
9) Perte du goût	–	–	–	–	–	v	–	–	v	–	2
10) Rhume(s)	–	–	–	v	–	–	v	–	–	–	2
11) Mal à la tête	–	–	v	–	–	v	–	–	–	–	2
12) Mal au ventre	–	–	v	–	–	–	–	–	–	–	1
13) Mal aux muscles	–	–	–	–	–	–	v	–	–	–	1
14) Mal au cou	–	–	–	–	–	v	–	–	–	–	1
15) Gorge enflammée	–	–	–	–	v	–	–	–	–	–	1
16) Fatigué	–	–	v	–	–	–	–	–	–	–	1
17) Vomissement	–	–	–	–	–	v	–	–	–	–	1
18) Bobos	–	–	–	v	–	–	–	–	–	–	1
19) Cœur	–	–	–	–	–	–	–	v	–	–	1
20) Rhume des fesses	–	–	–	–	–	v	–	–	–	–	1
21) Ça dépend de chaque personne	–	–	–	–	–	–	–	–	v	–	1
22) Transmet(tre) (aux autres personnes)	v	–	–	–	–	v	–	–	–	–	**2**
23) Vacciné(e)	–	–	–	–	–	–	v	–	–	–	1
24) Si symptômes : test, attendre les résultats, isolement	–	v	–	–	–	–	–	–	–	–	1
25) En 2020, il y a eu pas mal de cas de décès	–	–	–	–	–	–	–	v	–	–	1
26) En 2020, quand on avait les masques	–	–	–	–	–	–	–	–	–	v	1
27) Quand les animaux | êtres humains ils vont tout près	–	–	–	–	v	–	–	–	–	–	1
# Réponses	2	2	5	4	5	**10**	7	3	5	1	**44**

TABLE 6.26 – Connaissances préalables de la maladie du COVID-19

La **deuxième** question concernait le ressenti des participants quant à l'utilité des pictogrammes (**H3**). Les avis étaient plutôt positifs comme pour l'autre étude (cf. section 1.6.3). En fonction des participants, la facilitatrice a formulé différemment la question avec parfois plusieurs sous-questions. Lorsqu'elle a demandé si les pictogrammes sont utiles (ou pourraient aider) à la compréhension (à l'hôpital ou chez le médecin) : « oui » (P1, P8, P9, P10), « oui [pour] savoir qu'est-ce que j'ai comme maladie [...] c'est pour savoir facilement aussi » (P4), « ouai [...] s'il y a des personnes qui vont bien ou pas bien » (P5). Elle a demandé aux participants s'ils aimeraient qu'à l'hôpital ou chez le médecin, ce dernier leur propose des pictogrammes comme cela : « oui [car] ça, c'est plus facile pour voir [car] ça se voit » (P1), « oui » (P5), « oui cela pourrait être utile » (P6).

Elle a également demandé si le participant aime bien utiliser les pictogrammes : « oui moi ça, moi ça, ça m'aide [car] ben il y a des images, il y a des images et tout ça » (P2), « oui » (P5, P8, P9, P10), « ça peut aider les gens, mais il faut il faut avoir plus de détails sur les sur les pictogrammes » (P7). Quant à P2, il a aussi répondu que « oui », il aimerait bien continuer à utiliser ce genre de pictogrammes. La facilitatrice a voulu savoir si les participants trouvaient les pictogrammes utiles pour quelqu'un qui ne parle pas la langue (ou pour d'autres personnes) : « certains oui, certains oui, pas tous, mais certains oui » (P3), « oui » (P4, P8).

P5 a toutefois reconnu qu'il y avait « peut-être, peut-être » des pictogrammes un

peu compliqués. P6 a donné un exemple de pictogrammes pas clairs : « quand on voit deux horloges qu'il y a rien et puis une autre qui est rouge, on ne sait pas si c'était sur le moment » (le concept temporel « *Post* » dans la phrase 30E) ; il fait référence au problème de la temporalité (passé, présent). Il a dit ne pas avoir compris « quand la personne a été [...] ou [...] était en quarantaine ». P10 a également reconnu que « certaines questions » étaient un peu plus difficiles à comprendre.

Nous avons aussi quelques cas où les pictogrammes ne sont pas jugés nécessaires pour le participant : « en fait j'ai pas forcément besoin de ça pour m'aider » (P3). Celui-ci préfère que « quelqu'un [lui] explique la situation [avec des mots] depuis les pictogrammes ». C'était également le cas de P7 : « pour moi, ben je sais lire ». Il a toutefois ajouté qu'il pense que cela pourrait être « facile » pour certaines catégories de personnes, à savoir, « pour les gens qui ne savent pas lire et puis les gens qui ont de la peine à comprendre les choses ». P9 a aussi donné l'exemple d'un « enfant comme ça, qui a de la peine à lire, il a de la peine à comprendre » : « si on explique avec des dessins ou des trucs comme ça, peut-être qu'il comprendra mieux ».

Enfin, la **troisième** question abordait de nouveau les éventuelles pistes d'amélioration. Bien qu'aucun participant n'ait tout compris (cf. Table 6.21), certains ont estimé qu'il ne faut pas améliorer les pictogrammes : « je crois que j'ai tout compris » (P1), « ça va, pas spécialement [...] ça peut jouer comme ça » (P4), « ça joue comme ça » (P8).

Ce n'était pas le cas de tous : « je pense qu'il y a deux ou trois images à améliorer euh [...] euh que les gens ben qu'ils comprennent, qu'ils comprennent mieux et tout ça quoi » (P2), « certaines choses [...] ouai certains pictogrammes étaient [...] compliqués pour moi de comprendre [...] où il y [en] a plusieurs » (P3), « mettre, par exemple euh, euh une flèche [...] [pour] indiquer les choses » (P7), « ça dépend sur quoi, mais il faudrait trouver des pictos qui correspondent un peu à tout le monde [...] vraiment trouver des choses, même mettre des smileys ou je ne sais pas, un truc un peu ludique [...] plus ludique » (P9). Nous retrouvons des difficultés liées au nombre plus élevé de pictogrammes et la demande d'ajouter des flèches. [31]

Les participants n'ont pas tous réussi à expliquer qu'est-ce qu'il fallait améliorer : « il faut améliorer encore un peu ouai [...] *long silence* mais améliorer mais après euh » (P5). Néanmoins, quelques-uns ont évoqué la piste de la multimodalité comme dans la première étude : « si on aurait mis les mots, par exemple, là, j'aurais compris » (P3), « faire des petites phrases [...] pour expliquer, pour être mieux sûr » (P6), « avec quand même un petit texte en-dessous [...] juste un petit mot simple, juste pour expliquer en gros ce que ça veut dire » afin que ce soit « plus facile » (P10).

31. À noter que nous avons vu dans l'autre étude que les pictogrammes avec des flèches étaient ceux qui avaient été les moins compris (cf. Table 6.9). Néanmoins, il s'agissait de pictogrammes composés uniquement de formes géométriques (« première fois/déjà », « causer », « encore/déjà », « depuis »).

2.7 Discussion

Dans cette partie, nous discutons de l'objectif et des hypothèses de recherche, ainsi que des réponses de nos participants concernant les phrases en pictogrammes.

Objectif et hypothèses de recherche

L'**objectif** principal de cette nouvelle étude était d'évaluer la compréhension de questions traduites en pictogrammes (sans texte), qui peuvent être posées lors d'une anamnèse médicale sur un cas précis, ici dans un secteur de dépistage de la maladie du COVID-19. Tout d'abord, nous avons vu que le sens global des phrases était mieux compris dans cette étude que dans la précédente : environ 50 % contre 25-31 % selon les deux mêmes annotateurs (cf. Tables 6.21 et 6.7). Il faut relativiser ces résultats. Même si les phrases entre les deux études ont été générées par deux versions du même système de traduction, elles étaient différentes (au niveau des phrases sources, du nombre de phrases, du type de phrase, du nombre de pictogrammes par phrase, etc.). Nous ne pouvons pas conclure que les améliorations observées sont dues à la nouvelle version de PictoDr, la version contextuelle. Le design expérimental était également un peu différent, la facilitatrice a posé la question différemment, avec des gestes, etc. Le contexte de l'étude a été amélioré. Avec un modèle mixte, nous avons vu qu'il n'y a pas d'effet d'apprentissage entre les deux évaluations (2023 et 2024). Il ne s'agit pas d'une étude longitudinale.

Hypothèses | Variables	**A1-Phrase (#300)**	**A2-Phrase (#300)**
H1 | Section COVID-19 : A ≠ B ≠ C ≠ D ≠ E (section)	***	***
H2 | Nombre de pictogrammes : 1 > 3 (V1)	***	***

* $p < 0{,}05$; ** $p < 0{,}01$; *** $p < 0{,}001$ || A = Symptômes | B = Param. vitaux | C = Crit. de vulnérabilité | D = Habitudes de vie | E = Exposition

TABLE 6.27 – Niveaux de significativité des résultats de compréhension des phrases par hypothèse de l'évaluation d'une anamnèse selon deux annotateurs (A1-A2)

Les niveaux de significativité des résultats quantitatifs par hypothèse sont résumés dans la Table 6.27 (H1 à H2).[32] Nous avons vu que des phrases sont facilement comprises comme celles sur les habitudes de vie ; d'autres sont difficilement compréhensibles, en particulier, celles qui concernent les demandes d'information sur les paramètres vitaux et l'exposition (**H1**). Nous avons observé une différence extrêmement significative entre des nombres de pictogrammes différents (**H2**). Enfin, nous avons récolté des données qualitatives, incluant des avis positifs sur la perception de l'utilité des – phrases en – pictogrammes (**H3**). Nous proposons de revenir brièvement sur des points liés à chacune de nos hypothèses.

32. Dans cette table, nous reportons aussi la significativité avec les annotations de A2, même si nous avons utilisé celles de A1 pour valider les deux premières hypothèses.

Durant la crise sanitaire, les médecins ont dû poser une série de questions différentes liées au dépistage du COVID-19. Par conséquent, nous avions plusieurs catégories (**H1**),[33] dont une plus importante avec plus de la moitié des questions en pictogrammes (18 pour la section A sur les symptômes). Les autres (B à E) comprenaient chacune 2 à 4 questions. Nous avons voulu traduire les phrases qui nous semblaient les plus représentatives de la maladie du COVID-19 pour un public non spécialiste. Cela était moins le cas des critères de vulnérabilité qui peuvent concerner beaucoup d'autres maladies que le COVID-19. Dans le formulaire des HUG, certaines phrases n'avaient pas de traduction correcte dans PictoDr (souvent à cause de pictogrammes Arasaac manquants ou jugés mal représentés par les concepteurs du système). Les sections significatives sont celles qui ont obtenu le meilleur (D) et les pires (B et E) pourcentages de compréhension, respectivement, une très générale (D sur les habitudes de vie) et deux plus spécifiques au domaine médical (B sur les paramètres vitaux) ou au COVID-19 (E sur l'exposition). Les phrases sources de la dernière section sont plus difficiles à comprendre même sans pictogrammes (28E, 29E et 30E) ; elles sont longues et difficilement simplifiables syntaxiquement et lexicalement, en gardant le même sens. En pictogrammes, ces phrases sont plus difficiles à représenter à cause du nombre plus élevé de détails nécessaires à la transmission du sens complet.

Le nombre de pictogrammes par phrase (**H2**) – qui n'était pas significatif dans la précédente étude –, est extrêmement significatif ici ; moins il y a de pictogrammes, plus il y aura de chances que la phrase soit comprise. Cela renforce les impressions des participants qui ont chaque fois montré des signes de découragement à la vue des phrases avec plus de pictogrammes. Dans les résultats qualitatifs de nos deux études (cf. sections 1.6.3 et 2.6.2), des participants ont confirmé à la facilitatrice que c'était plus clair « quand il y avait moins d'images » (P2 dans la première évaluation) ; et que les phrases plus difficiles à comprendre étaient celles avec « plusieurs » pictogrammes (P3 dans la deuxième). Ces résultats soutiennent ceux des focus groups de Sevens (2018, p. 72). En effet, elle avait observé que durant les tests utilisateurs, des séquences de pictogrammes longues et complexes avaient amené ses participants adultes avec une DI à ressentir davantage de difficultés à comprendre le sens des messages Facebook. Leur écran avait été complètement inondé de pictogrammes générés avec le système de base de Text-to-Picto. C'est d'ailleurs la raison pour laquelle elle avait conçu un module de simplification syntaxique et de compression pour la version néerlandaise du système (Sevens, 2018; Sevens *et al.*, 2017). Ici, PictoDr traduisait déjà uniquement les concepts principaux des phrases sans les mots de fonction (prépositions, articles, etc.).

En ce qui concerne les résultats sur la perception de l'utilité des – phrases en –

33. En ce qui concerne le formulaire COVID-19 des HUG, il n'est pas conçu spécifiquement pour les patients avec un handicap. Il est destiné aux médecins et aux infirmiers. Selon un médecin des HUG, il est utilisé en format papier, puis scanné dans le dossier médical du patient.

pictogrammes chez nos participants (cf. section 2.6.2), ils semblent confirmer notre hypothèse (**H3**). Il faut toutefois rester relativement prudent étant donné qu'il s'agit de données qualitatives sur un petit échantillon d'individus. Comme nous l'avons vu dans le chapitre 1 (section 3.2), les personnes avec une DI ont tendance à répondre « oui » quelle que soit la question. Aucun de nos participants n'a dit explicitement « non » ici, même si nous avons eu à nouveau des remarques en fonction de certains pictogrammes ou de certaines phrases. Au moins deux participants ont dit que ce n'était pas « forcément » utile pour eux et/ou que cela pouvait être utile pour d'autres catégories de personnes. Des participants ont encore suggéré les mêmes améliorations : la multimodalité, l'ajout de flèche et la clarification de certains pictogrammes. Une personne a aussi proposé d'ajouter des « pictogrammes qui correspondent un peu à tout le monde », des « smileys » ou un côté « plus ludique ». Quant à la première question de cette évaluation, elle avait pour but de contextualiser l'étude. Bien que les définitions des participants sur ce qu'est la maladie du COVID-19 étaient plus ou moins précises (cf. Table 6.26), nous avons vu que tous pouvaient comprendre le contexte de l'étude. Cela justifie le choix du scénario, celui-ci ne devait pas être un frein à la participation.

Dans nos deux études, nous avons voulu évaluer la compréhension de phrases en pictogrammes sans texte. Pour des utilisateurs avec une DI, nous pensons qu'il est indispensable de proposer une interface avec des options de personnalisation comme l'activation ou non de la multimodalité (du texte simple, la reconnaissance et la synthèse vocale, etc.), que ce soit pour le patient et/ou pour le médecin. Dans le système PictoDr que nous n'avons pas évalué (car il était encore en développement), plusieurs fonctionnalités pourraient permettre d'améliorer la compréhension des patients. Par exemple, le bouton qui permet de choisir entre plusieurs pictogrammes différents pour un même concept (cf. Figure 6.12), la possibilité de sélectionner la version compositionnelle ou la version non compositionnelle d'une même phrase, le fait de pouvoir agrandir chaque pictogramme en cliquant dessus, etc. L'évaluation du système avec son interface en situation réelle ouvre des perspectives de recherches intéressantes.

Analyse des réponses par phrase

Nous détaillons ici les réponses des – questions en – pictogrammes. Toutes les phrases en pictogrammes sont consultables dans l'annexe B.8.1. Les résultats quantitatifs par phrase sont disponibles dans la Figure 6.16. Les participants avec une DI avaient pour tâche d'expliquer les phrases avec leurs mots. Les deux annotateurs (A1 et A2) ont donc accepté les synonymes, paraphrases et reformulations. Par exemple, P6 avait dit « froid » puis « on tremble de froid » pour la phrase **4A** (« avez-vous des frissons ? »). Comme pour la première étude, les participants ne répondaient pas toujours sous la forme d'une question. Les deux annotateurs ne l'ont pas compté comme une erreur, étant donné qu'il ne s'agissait pas d'une expérience en conditions écologiques avec de vrais médecins. Les participants décrivaient ce qu'ils voyaient ou complétaient la question de la facilitatrice (-e : « qu'est-ce que je veux savoir » | -p : « si tu si je fume »). Ils ont formulé leurs réponses à toutes les personnes (« je », « tu », « il(s) », etc.).

La facilitatrice (A2) a aidé le participant au moins une fois dans plus de 51 % des phrases (153/300 ont une annotation « aide » selon A1). Le nombre de questions supplémentaires se répartit comme suit : 155 « 0 » (51,66 %), 68 « 1 » (22,66 %), 25 « 2 » (8,33 %) et 52 « 3++ » (17,33 %). Ce sont toutefois des métadonnées relativement subjectives, car il n'est pas toujours évident de savoir si les reformulations et paraphrases doivent être comptées comme des questions ou de l'aide et si elles ont eu un impact ou non sur la réponse des participants. Parfois, la facilitatrice répétait simplement la même question, le participant donnait alors une réponse identique, voire différente (ou répondait pour la première fois). Les disfluences (dont les nombreux « hum hum » de la facilitatrice) ne sont pas incluses dans ces calculs. Durant toute l'expérience, la facilitatrice pouvait aussi noter divers commentaires sur chaque phrase dans la grille de contrôle de chaque participant : des éléments problématiques et leur description.

Pour **1A** (« avez-vous le nez qui coule ? »), 6 participants ont d'abord pensé que le pictogramme représentant un nez qui coule signifiait un « rhume » (P1-P4, P6, P8). Il s'agit bien du nom du pictogramme Arasaac (sans la flèche). Néanmoins, la facilitatrice a chaque fois demandé une précision (« plus spécifiquement », « en particulier, plus en particulier »). Dans 5 cas, les participants ont ensuite donné la réponse correcte selon A1 et A2. À noter que P5 a confondu le nez avec une « poire ». Pour **2A** (« avez-vous mal à la gorge ? »), la phrase a été jugée correcte pour tous les participants par A1 et A2, étant donné qu'ils ont aussi accepté la « gorge irritée » (P4) et « (mal) au cou » chez 4 participants (P2, P5, P6, P8) pour le pictogramme représentant un « mal à la gorge ».

Contrairement à la précédente, **3A** (« avez-vous des douleurs musculaires ? ») a été très mal comprise (A1-A2 : 2/10), elle a nécessité beaucoup d'aide de la facilitatrice (pour 7 participants sur 10). Il s'agit d'une version compositionnelle (cf. Figure 6.12), incluant deux pictogrammes (pour les concepts UMLS « *Pain* » + « *Muscle* »). À cause

du premier pictogramme, 5 participants ont compris « mal à la tête » ou « migraines » ; 3 ont décrit l'expression du personnage (P1 : « nerveux [...] triste » ; P2 : « pas content » ; P10 : « quelqu'un qui ne sourit pas »). **4A** a bien été comprise (A1-A2 : 9/10). Beaucoup ont parlé du « froid » (ou utilisé d'autres mots appartenant à ce champ lexical).

Quant aux **5A** et **6A** (respectivement, « avez-vous une toux grasse ? » et « est-ce sanglant quand vous toussez ? »), nous retrouvons un problème de précision dans les réponses étant donné qu'il s'agit de représentations non compositionnelles (un seul pictogramme pour la toux avec des gouttes bleues pour la toux grasse | un avec un nuage rouge pour le sang). Le plus souvent, les participants ne trouvaient pas tout de suite la réponse attendue, la facilitatrice devait alors poser au moins 3 questions ou plus (« et plus précisément », « quel type de toux »), afin de l'obtenir. Nous avons déjà expliqué pourquoi **7A** n'a été comprise par aucun participant (cf. section 2.6.1). C'était déjà le cas dans la première étude avec Text-to-Picto et PictoDr (cf. Figure 6.6b, Ph10).

Pour **8A**, **9A** et **10A**, la facilitatrice a noté que dans quelques cas, certains participants n'avaient pas compris pas tout de suite. Ils pensaient parfois à d'autres choses ou hésitaient. Cela explique pourquoi **10A** est une des phrases avec un écart de 30 % entre A1 et A2 ; A1 a accepté moins de réponses. Pour **11A** (« avez-vous la diarrhée ? »), nous avons eu 7 participants qui ont dit la « diarrhée » (P2-P5, P7 et P9-P10 sans aide), 1 a eu du mal à identifier le type exact de « caca » (P1 a essayé tous les qualificatifs : « différent », « rouge », « normal », « dur », « facile », avant de donner la bonne réponse) – comptée comme incorrecte par A1 et A2 à cause du nombre élevé de questions supplémentaires –, tandis que les 2 autres ont été très précis (P6 : « dysenterie » ; P8 : « gastro »). Pour **12A** (« avez-vous mal au ventre ? »), A1 et A2 ont accepté toutes les réponses qui concernaient le ventre, mais pas « vers les reins » (P8).

Les expressions faciales peuvent être subjectives et ambiguës. Ainsi, pour **13A** (« êtes-vous fatigué ? »), nous avons « fatigué » comme seule réponse (P3-P4, P8-P9), mais aussi « pas content » (P1, P10), « triste » (P1, P2, P5) ou « pas bien (du tout) » (P6, P7). Pour **14A** (« avez-vous des difficultés à respirer ? »), les participants ont eu difficile à identifier le verbe « respirer » à cause de l'expression faciale et des flèches. En ce qui concerne l'anatomie avec **15A** (« avez-vous mal à la poitrine ? »), A1 et A2 ont aussi accepté le « thorax » (P2, P8) et d'autres organes internes comme le « cœur » (P9) ou le « poumon » (P7, P10). **16A** (« avez-vous mal à la tête ? ») a été très bien comprise, les participants ont vu que le bonhomme se tenait la tête. Pour **17A** (« avez-vous des allergies (connues) ? »), ils n'ont pas du tout identifié le symptôme correct. Chacun a donné plusieurs réponses, 8 ont reçu de l'aide. Les réponses les plus fréquentes étaient les « démangeaisons » ou l'« eczéma », des conséquences de l'allergie.

En ce qui concerne **18A** (« à quelle date ont commencé les symptômes ? »), elle comportait trois pictogrammes qui n'ont pas été compris. Pour le premier, la moitié ont vu un « calendrier » ou un « prochain rendez-vous » comme pour l'autre étude. Pour le

troisième, la facilitatrice a noté que les participants ne comprenaient pas le sens, 1 a dit qu'il pensait que l'image n'était pas claire. Une ligne de départ de course est en effet utilisée pour représenter « commencer », un verbe polysémique. Afin de comprendre la phrase, la personne doit reconnaître que c'est le début d'une course, être certain que c'est un verbe, trouver le bon et faire le lien avec les autres pictogrammes de la phrase. Selon A1 et A2, 3 participants ont dépassé le problème de la compositionnalité (P3, P7, P9). Pour **19B**, seul 1 participant a réussi à comprendre le sens du pictogramme créé pour cette étude (« *Tachypnea* »). En ce qui concerne **20B**, nous avons également déjà expliqué pourquoi elle n'a été comprise par aucun participant (cf. section 2.6.1).

Nous revenons sur la seule phrase où A1 (3/10) > A2 (2/10), **21B** (« à combien était votre température la dernière fois que vous l'avez mesurée ? »), A1 et A2 ont des avis divergents à cause de la deuxième partie de la phrase. P3 répond correctement « c'est savoir euh combien de température j'ai, si c'est élevé la température ou pas ». La facilitatrice a ensuite demandé « oui, mais à quel moment ? », P3 a répondu « le matin ou le soir, je sais pas ». Même si nous pouvons comprendre la raison pour laquelle A2 a jugé la phrase incorrecte, nous estimons que la deuxième partie de la phrase est facultative. La réponse du participant nous semble suffisante. En posant cette question, le médecin veut savoir combien il a de température (et pour le savoir, c'est évidemment la dernière fois qu'il l'a mesurée). Nous avons accepté sa réponse à cause de ce sous-entendu, c'est pour cela que l'on ne le représente pas en pictogramme dans PictoDr.

Nous avons aussi voulu voir si les concepts médicaux plus difficiles des **22C** et **23C** étaient compris (respectivement, « avez-vous eu un accident vasculaire cérébral ? » et « avez-vous eu une chimiothérapie ? »). Selon A1 et A2, 1 participant a correctement répondu pour le premier (P7) ; 4 pour le deuxième (P3, P7-P9). D'après les commentaires de A2, 6 ne sauraient pas ce qu'est un accident vasculaire cérébral (P1, P2, P5-P7, P10) ; 1 pour la chimiothérapie (P5). On peut relever la réponse descriptive pour **23C** de P5 (ex-P8) : « il tient un lampadaire », suivie de son interprétation « elle allume une lumière ». Les autres ont confondu la chimiothérapie avec la « pression » (P1), des « prises de sang » (P2) ou une « perfusion » (P6, P10). **24C** (« fumez-vous ? »), la dernière phrase de la catégorie sur les critères de vulnérabilité a été comprise par tous les participants selon A1 et A2. Le même pictogramme pour « fumer » avait été compris à 92,85 % (13/14) dans la première étude (cf. Table 6.10).

Les deux questions sur les habitudes de vie, les **25D** (« avez-vous trop de tension artérielle ? ») et **26D** (« avez-vous été vacciné contre le covid ? ») ont obtenu le même score de compréhension, 75 % en moyenne. Pour la première, la facilitatrice a dû les aider pour qualifier la tension (par exemple, « donc là, je veux savoir si tu as la pression comment ? ») ; pour la deuxième, tous les participants n'ont pas trouvé la phrase exacte. Le pictogramme pour le concept UMLS « *Too much* » est ambigu, 3 participants ont répondu que la tension était « basse » (P2, P5), voire « haute ou bien basse » (P4, P5).

En ce qui concerne la vaccination, P8 a répondu que le médecin aimerait savoir s'il a « besoin ou pas » du vaccin (jugé incorrect par A1 et correct par A2). P10 a dit « est-ce que je dois le faire ou pas, le vaccin » (incorrect pour A1 et A2, la phrase est au présent).

À noter que le pictogramme représentant le COVID-19 a toujours été bien reconnu (**26D** et **29E**). Nous pouvons supposer que c'est parce qu'il est écrit COVID-19 dessus et que nos participants savaient lire. Cela renforce l'idée qu'une utilisation multimodale (texte + image) peut être utile pour certaines personnes avec une DI (cf. sections 1.6.3 et 2.6.2). Pour **27E** (« avez-vous voyagé récemment ? »), la facilitatrice a noté pour 4 participants que « la temporalité est difficile à comprendre » (P1, P3, P8, P10), les participants ont difficile à interpréter la représentation des calendriers. Pour **28E** (« avez-vous été en contact avec une personne confirmée positive au coronavirus dans les quatorze derniers jours ? »), les participants ont des difficultés à savoir qui est contaminé, 9 ont eu besoin de l'aide et 1 seul a bien répondu selon A1. **29E** (« à quelle date avez-vous été en contact avec quelqu'un de positif au coronavirus ? ») est la deuxième phrase avec un écart de 30 % entre A1 et A2, dû ici à un problème de temporalité.

Nous avons déjà un peu discuté de **30E** (« étiez-vous en quarantaine en tant que contact proche d'un cas confirmé de coronavirus ? »), car il s'agit d'une phrase qui a nécessité de l'aide (cf. Figure 6.15). Dans 5 cas sur 10, la facilitatrice a posé au moins trois questions supplémentaires. Elle a noté plusieurs commentaires : « difficulté à mettre ensemble les images » (P2), « le participant ne sait pas quelle est la question mais comprend les images » (P4), « les images sont très difficiles à comprendre pour ce participant, pas sûr qu'il ait compris » (P5), « la temporalité est difficile à comprendre » (P6), « le participant considère cette phrase très compliquée » (P9). D'après un médecin des HUG, cette question n'est plus posée aujourd'hui. Il convient de noter que nous n'avons pas utilisé la dernière version du formulaire ; il a été un peu adapté.

Lorsque la réponse à la première question était incorrecte dans cette étude, la facilitatrice devait donner la bonne réponse et poser deux questions supplémentaires (cf. section 2.3).[34] Les participants répondaient plus souvent « oui » (c'est-à-dire que pour eux le pictogramme aide à comprendre la phrase) que « non » (« oui » : 57,14 % | « non » : 42,86 %), mais ils ne savaient pas toujours justifier en quoi le pictogramme pouvait les aider. Comme nous l'avons déjà mentionné, les personnes avec une DI répondent souvent « oui » aux questions fermées. Nous avions d'ailleurs envisagé de demander aux participants de répondre directement aux questions posées pour voir s'ils avaient compris la phrase, mais nous avons renoncé car presque toutes les phrases évaluées étaient des questions fermées (sauf **18A**, **21B**, **29E**). Ils auraient pu inventer une réponse, répondre au hasard et s'ils avaient répondu juste « oui » ou « non » plutôt que « non je ne fume pas », nous n'aurions pas pu déterminer s'ils comprennent.

34. Pour rappel, « est-ce que le(s) pictogramme(s) aide(nt) à comprendre le mot ou la phrase » et « comment comprenez-vous le mot ou la phrase avec le(s) pictogramme(s) ».

3 Discussion générale : une méthodologie

L'un des objectifs de ces deux études était de proposer une méthodologie pour évaluer des traductions – automatiques – en pictogrammes auprès d'adultes avec une DI. Nous avons donc testé une méthode non invasive qui a permis de récolter des données quantitatives et qualitatives utiles à nos objectifs et hypothèses de recherche (cf. sections 1.1 et 2.1), tout en étant adaptée à des adultes verbaux avec une DI ne connaissant pas les pictogrammes. La recherche participative et inclusive comprenant l'implication de personnes en situation de handicap nous semble indispensable. Nous avons toutefois été confronté à plusieurs contraintes, dont la nécessité de recueillir une quantité de données suffisante pour garantir des résultats significatifs, tout en réduisant la charge cognitive de nos participants (dont les efforts et un temps de concentration prolongé sont souvent plus limités que chez les personnes sans DI).

Une approche qualitative avec un nombre restreint de participants est souvent utilisée lorsqu'il s'agit d'évaluer des personnes avec une DI (Beail et Williams, 2014), que ce soit via des entretiens ou des focus groups. Étant donné la difficulté de recruter des personnes en situation de handicap, nous n'avons pas voulu imposer de conditions d'inclusion strictes. Tout comme Sevens (2018, p. 48), le critère d'être motivé à coopérer (et avoir les capacités) nous semblait suffisant. Nous avons souhaité adopter une position similaire à celle de Schwab (2021, p. 42) concernant l'implication de personnes en situation de handicap dans un autre projet lié à la CAA, ParticipAACtion : « les personnes avec des handicaps plus modérés, comme c'est le cas de la trisomie 21 par exemple, ont aussi des difficultés spécifiques avec la parole et souffrent du manque d'usage de CAA. Le choix d'un outil ou d'une méthode de CAA n'étant pas lié à l'étiologie du trouble mais aux compétences de la personne et de son environnement, il nous a paru incontournable de ne pas nous limiter à un syndrome spécifique afin de favoriser le transfert des connaissances et l'universalisation des usages ».

Pour le recrutement, nous ne voulions pas de critères d'exclusion tels que la restriction à un type de handicap, à un certain QI, à une connaissance préalable de la technologie ou de la CAA. Par exemple, l'utilisation d'un questionnaire en ligne – qui permet une plus large diffusion – ou la passation de tests utilisateurs via un logiciel de visioconférence ne nous semblaient pas être des solutions envisageables avec des personnes ayant une DI, alors que nous avions déjà de l'expérience avec ces méthodes pour évaluer des pictogrammes (Norré *et al.*, 2020, 2021a,b; Bouillon *et al.*, 2023). Dans un réel contexte d'utilisation, les médecins font face à des patients qui n'ont pas forcément de connaissances technologiques ou de la CAA. À noter qu'un focus group nous aurait plus éloigné de notre situation de communication, l'anamnèse. Une piste qui pourrait être envisagée pour analyser la compréhension de – phrases en – pictogrammes serait le recours à des techniques d'oculométrie (*eye-tracking*). En

effet, celles-ci sont déjà utilisées pour étudier l'accessibilité de sites web, mais aussi des questions de lisibilité et de compréhension écrite, incluant celle de documents en FALC – qui comprennent aussi des images –, auprès de personnes avec des déficiences cognitives, un Trouble du Spectre de l'Autisme (TSA), etc. (Yaneva *et al.*, 2015, 2016).

Cette thèse ayant débuté durant la pandémie de COVID-19 en 2020, nous ne savions pas si nous aurions le temps de faire des tests utilisateurs entre des adultes avec une DI et des médecins dans des hôpitaux ou des foyers. C'est pourquoi, une limitation de nos études est sans doute le nombre de participants et leur profil. Impliquer des personnes en situation de handicap est un processus qui nécessite du temps et des moyens. Cela passe par l'obtention d'un consentement éclairé – auprès du participant ou de son tuteur –, mais aussi par le besoin d'être entouré par des professionnels de la santé ou des aidants proches qui connaissent et comprennent la personne en situation de handicap. Il est en effet nécessaire de pouvoir s'adapter aux capacités de ces participants, en particulier, celles des personnes avec une DI plus sévère qui peuvent avoir des troubles du comportement ou un polyhandicap. Dans ces deux évaluations, nous avons fait en sorte de simplifier le plus possible le protocole expérimental pour la facilitatrice et nos participants (la tâche, le temps de l'expérience, le matériel, etc.).

Comme à notre connaissance il n'existait pas d'études auprès de personnes avec une DI sur la compréhension de phrases médicales traduites en pictogrammes (cf. chapitre 2, section 1.3.3), nous avons fait le choix de proposer une méthodologie, qui est une adaptation très libre de celle de Rodríguez Vázquez et Bouillon (2021), destinée à évaluer la traduction en FALC. Nos deux études présentent toutefois certaines limites. Quelques-unes ont déjà été prises en compte dans la deuxième évaluation, dont l'amélioration des traductions du système, du contexte (une anamnèse), voire de l'annotation (l'ajout de métadonnées). Nous aimerions insister sur le fait qu'il faut rester relativement prudent quant à l'analyse de certains de nos résultats. Nous n'avions par exemple pas de groupe contrôle, ce qui nous aurait permis de proposer une comparaison avec la compréhension de nos phrases médicales en pictogrammes auprès de personnes sans une DI. Les traductions ont été validées par nous, voire par deux autres concepteurs du système en ce qui concerne celles de PictoDr dans la deuxième étude.

Bien que nous ayons ajouté de nouvelles métadonnées (si le participant a été aidé et le nombre de questions supplémentaires), l'annotation manuelle reste une tâche subjective. Les deux annotateurs devaient juger si les – phrases en – pictogrammes étaient compris ou pas par nos participants. Comme l'avaient aussi fait remarquer Bulté *et al.* (2021, p. 199) pour leur étude, « *the evaluation procedure involves judgements that can be characterised as subjective, as concepts such as 'useful' and 'comprehensible' are clearly open to interpretation* ».[35] Étant donné que la facilitatrice n'avait

35. La procédure d'évaluation implique des jugements qui peuvent être qualifiés de subjectifs, car des concepts tels que « utile » et « compréhensible » sont clairement ouverts à l'interprétation.

pas d'expérience avec les pictogrammes et évaluait les participants durant les tests sur la base de nos phrases sources, [36] nous avons privilégié ici une évaluation sur la compréhensibilité avec deux choix : correct ou incorrect ; contrairement à Bulté *et al.* (2021), qui avaient évalué eux-mêmes leurs traductions en pictogrammes Sclera, Beta et Rand avec une échelle composée de quatre catégories au niveau des phrases (cf. chapitre 2, section 1.3.2). Nous avons simplifié l'annotation pour la facilitatrice.

Le fait de se baser sur une échelle comportant seulement deux critères de jugement présente toutefois un biais, puisque la compréhension partielle ou incertaine n'est pas représentée, alors que celle-ci était bien présente (cf. section 1.4), que ce soit parce que le participant ne comprenait pas tout ou parce ce que l'annotateur n'était pas capable de déterminer si le participant avait compris. Un autre biais peut concerner le nombre et le profil des annotateurs. Pour valider nos hypothèses, nous avons utilisé uniquement les annotations d'un annotateur, car elles étaient plus complètes. [37] Il s'agissait également de l'annotateur qui avait de l'expérience avec les pictogrammes. Nous pensons que si un médecin avait évalué la compréhension de ces phrases en pictogrammes, il aurait peut-être été plus sévère sur certains cas. Par exemple, un « mal à la gorge » n'a pas les mêmes conséquences qu'un « mal au cou » en termes de diagnostic et de traitement ; or, nous avons vu que les deux annotateurs l'ont toujours accepté.

Bien que nous ayons amélioré la contextualisation de l'étude (des phrases indépendantes à une anamnèse, la multimodalité avec l'explicitation de la tâche via des gestes, la formulation de la troisième à la première personne, etc.), il ne s'agissait toujours pas d'un test utilisateur en situation écologique entre un médecin et son patient. Néanmoins, recourir à une facilitatrice – responsable du bureau Langage simplifié et habituée à faire relire des textes difficiles à des adultes avec une DI qu'elle connaissait – comportait plusieurs avantages. Cela a mis les participants en confiance et a pu les encourager à donner plus de réponses, plus variées, par exemple, pour les questions ouvertes sur le COVID-19 ou les améliorations. Comme les participants ont l'habitude de donner leur avis sur la compréhension de textes, on peut supposer que le sentiment d'être jugé était moins présent. Il s'agissait d'une évaluation des traductions du système amélioré en un an, mais pas d'une évaluation du système complet en contexte.

Afin notamment de réduire la durée des expériences et ne pas dépendre de problèmes techniques (internet, panne, etc.), nous avons fait traduire en amont les phrases sources. Nous avons vérifié chacune des traductions pour ne pas faire évaluer celles qui comportaient des erreurs de précision ou de rappel. Le choix d'utiliser des traduc-

36. Nous ne pouvions pas demander à la facilitatrice de réécouter tous les enregistrements ou de relire toutes les retranscriptions post-éditées après les expériences.

37. Comme nous ne savions pas le temps que prendrait une évaluation, nous n'avons pas demandé à la facilitatrice d'évaluer tous les pictogrammes – ce qui permettait aussi de diminuer la charge cognitive des participants. Nous aurions pu compter uniquement les pictogrammes et les phrases où les deux annotateurs étaient d'accord, mais cela aurait réduit davantage nos données qui étaient déjà limitées.

tions "parfaites" (pour les concepteurs) nous éloigne encore une fois d'une situation réelle, mais nos objectifs étaient d'évaluer si des adultes avec une DI pouvaient comprendre des – phrases en – pictogrammes (sans texte) indépendamment du système et de déterminer des facteurs qui influencent la compréhension pour un contexte médical donné. Nous avons aussi pu proposer des améliorations concrètes concernant les pictogrammes. Et comme nous l'avons vu dans nos études, les personnes avec une DI avaient déjà des difficultés à comprendre des phrases qui nous semblaient "bien" traduites. Nous pensons qu'à ce stade du développement, cela n'avait pas de sens d'évaluer des phrases dont il manquait des pictogrammes ou avec des erreurs de traduction.

Le choix du matériel utilisé pour les évaluations a généré une ou deux contraintes. En effet, nous n'avons pas pu évaluer certaines phrases de PictoDr. La Faculté de Traduction et d'Interprétation de Genève a créé de nouveaux pictogrammes, dont des pictogrammes animés pour des concepts temporels (par exemple, « *Hour* », « *Since when* », etc.). Une étude a montré que les pictogrammes animés pouvaient améliorer la compréhension des verbes chez des élèves avec une DI (Fujisawa *et al.*, 2011). Nous n'avons pas pu tester ce type d'images avec notre méthodologie, car nos phrases en pictogrammes étaient affichées dans un document PDF pour que la facilitatrice puisse facilement les présenter individuellement et les agrandir si besoin. Comme nous l'avons vu, nos participants ont eu difficile à comprendre la temporalité, les – phrases avec des – pictogrammes statiques représentant des calendriers et des horloges. Il nous semble donc intéressant de pouvoir évaluer si des pictogrammes animés améliorent ou non la compréhension des phrases médicales en pictogrammes auprès d'adultes avec une DI.

En conclusion de cette partie, nous ne prétendons pas avoir proposé la meilleure méthodologie, mais une parmi d'autres. Nous pensons qu'en fonction des objectifs, des hypothèses de recherche, ainsi que du public cible, il est possible d'améliorer certains points. C'est pourquoi, nous nous permettons d'insister sur des limites déjà évoquées. D'une part, les participants retenus n'avaient pas de DI très sévère. Il serait intéressant d'adapter ces études, afin d'impliquer des personnes ayant une DI plus sévère, voire un polyhandicap. D'autre part, les participants n'avaient pas une connaissance préalable des pictogrammes Arasaac, ce qui influence évidemment grandement nos résultats. Cela s'explique par la diversité des systèmes iconiques en usage dans différents pays (cf. chapitre 1). Nous pensons que pour favoriser l'implémentation de ce type de système, il faudrait que les utilisateurs, leur entourage, le personnel soignant (médecins, infirmiers, mais aussi logopèdes et ergothérapeutes) réfléchissent ensemble sur le développement ou la validation d'un grand ensemble de pictogrammes médicaux libres de droit les plus facilement interprétables possible qui puisse être utilisé par tous dans un grand nombre de situations différentes. Concevoir des interfaces personnalisables et adaptables aux besoins de divers utilisateurs avec une DI pourrait également permettre d'améliorer leur compréhension.

4 Conclusion : vers de nouvelles recherches ?

L'objectif principal de ce chapitre était de répondre à la question de recherche suivante : Text-to-Picto et PictoDr, des systèmes de traduction automatique du français vers les pictogrammes produisent-ils une sortie utile et compréhensible au niveau des pictogrammes et des phrases en pictogrammes pour des adultes avec une DI ?

Les résultats concernant la compréhensibilité des pictogrammes et la compréhensibilité des phrases en pictogrammes montrent que les traductions générées par les deux systèmes ne peuvent pas être utilisées telles quelles pour la communication médecin-patient. Bien que la compréhensibilité des phrases soit plus élevée dans les conditions de la deuxième étude, nous pensons que les traductions des systèmes actuels doivent encore être améliorées. En effet, nous avons constaté qu'il y avait des phrases et des pictogrammes qui étaient mieux compris que d'autres, alors que certains prêtaient clairement à confusion. Or, dans le contexte d'utilisation proposé, à savoir le domaine médical, les enjeux sont élevés et les erreurs coûteuses (Wołk *et al.*, 2017). Un médecin doit pouvoir avoir l'assurance d'être compris en toutes circonstances, même si nous avons vu que ce n'est pas non plus toujours le cas actuellement.

Dans ce chapitre, nous avons aussi proposé un corpus et une méthodologie originale d'évaluation d'interactions médicales traduites en pictogrammes, qui comprend des entretiens entre une facilitatrice et des adultes avec une DI. Nous avons montré que plusieurs facteurs peuvent influencer la compréhension des – phrases en – pictogrammes : le niveau de représentation (phrase ou pictogramme), le niveau de langue, la nature de la traduction, le type de phrase à traduire, la catégorie morphosyntaxique du mot à traduire, mais aussi le type de la question ou le nombre de pictogrammes par phrase. À notre connaissance, tous ces facteurs n'avaient pas encore été étudiés chez des adultes avec une DI, qui ne connaissent pas les ensembles de pictogrammes proposés. Bien que ces personnes aient souvent des difficultés à communiquer avec un médecin, recourir uniquement à des phrases traduites en pictogrammes sans texte n'apparaît pas comme une solution pour tous les patients avec une DI.

Nous avons vu qu'il n'était pas toujours facile de déterminer avec certitude si les personnes avec une DI comprenaient ou non les – phrases traduites en – pictogrammes. Les personnes interrogées ont toutefois eu plus de difficultés à comprendre le sens global des phrases que le sens de certains pictogrammes isolés. La première étude a mis en évidence le problème de la compositionnalité de sens qui passe dès lors par un nécessaire apprentissage des pictogrammes, mais aussi l'importance de la multimodalité et du potentiel amélioratif des pictogrammes pour communiquer. La deuxième étude a montré que l'utilisation des pictogrammes pour des adultes avec une DI offre des perspectives de recherches en vue d'améliorer le déroulement d'une anamnèse, en particulier, la communication médecin-patient avec une DI.

Conclusion

Nous avons proposé un travail sur la traduction automatique du français vers les pictogrammes pour améliorer la communication entre médecins et patients avec une Déficience Intellectuelle (DI) en milieu hospitalier. Pour ce faire, nous nous sommes appuyée à la fois sur le Traitement Automatique des Langues (TAL) et sur le domaine de la Communication Alternative et Améliorée (CAA) appliqué à la communication médicale. Dans cette conclusion, nous revenons d'abord sur les principales contributions (section 1). Dans un second temps, nous présentons la démo de Text-to-Picto (section 2), avant de refermer ce travail en discutant ses limites et ses perspectives (section 3).

1 Contributions

Dans le chapitre 1, nous avons commencé par décrire le contexte de notre étude, à savoir la CAA, avant d'aborder les thématiques centrales de notre étude : les pictogrammes et la communication entre médecins et patients avec une DI. Nous avons vu que la CAA inclut un ensemble de systèmes qui remplacent ou améliorent la communication des personnes, enfants et adultes, en situation de handicap moteur, langagier et/ou intellectuel, pouvant ainsi contribuer à améliorer leur qualité de vie et celle de leur entourage. Parmi ces méthodes, on retrouve des supports papiers ou technologiques avec des images. Les pictogrammes sont regroupés dans des banques d'images (libres de droit ou non), qui visent à faciliter la communication dans la vie quotidienne, mais aussi dans d'autres situations, telles que les consultations médicales. Dans le domaine médical, il a été montré que les patients avec une DI rencontrent des difficultés à comprendre les médecins, s'exprimer et à communiquer leur douleur. La communication s'appuyant sur des pictogrammes apparaît dès lors comme une piste, mais, bien que conçus pour faciliter l'accès à l'information et à la communication, les pictogrammes ne sont pas universels. Ils nécessitent souvent aux personnes avec une DI des années d'apprentissage auprès de professionnels de la santé (logopèdes, ergothérapeutes, etc.). Nous avons constaté que différents supports avec des images sont déjà utilisés dans les hôpitaux pour l'auto-évaluation (des fiches de dialogue, des échelles de la douleur, des tableaux de communication, etc.).

Dans le chapitre 2, nous avons présenté les tâches de TAL qui font l'objet de ce travail : la traduction automatique et la simplification automatique de textes. Nous avons d'abord vu qu'en raison du manque de corpus français en pictogrammes, nous ne pouvions pas utiliser les approches actuelles en traduction automatique, qui nécessitent des données pour entraîner des modèles. La traduction neuronale n'était d'ailleurs pas utilisée par les systèmes de traduction vers les pictogrammes avec des méthodes de TAL jusqu'il y a peu. En effet, les outils intégrant des techniques de TAL se limitaient à effectuer une analyse linguistique superficielle en vue de faire le lien avec une base de pictogrammes, tandis qu'un ou plusieurs modules s'occupent de tâches annexes : résumé automatique de texte, simplification automatique de textes, désambiguïsation lexicale, etc. Nous avons montré qu'aucun de ces systèmes n'était conçu pour la traduction automatique de phrases médicales françaises vers les pictogrammes. En ce qui concerne leur évaluation, nous avons également vu qu'il y avait peu d'études utilisateurs, en particulier auprès de personnes avec une DI.

Dans les autres chapitres de cette thèse, nous avons contribué à deux objectifs : **(1)** nous avons développé et amélioré Text-to-Picto, afin de proposer un système de traduction automatique du français vers les pictogrammes pour la communication entre médecins et patients avec une DI ; **(2)** nous avons évalué les traductions d'instructions et de questions médicales en pictogrammes chez des adultes avec une DI. Nous avons ainsi testé différentes approches de développement et d'évaluation de phrases traduites en pictogrammes (automatique, manuelle, études utilisateurs).

Bien que libre de droit et multilingue, le système de traduction automatique vers les pictogrammes Text-to-Picto de Sevens (2018), Vandeghinste *et al.* (2017) ne proposait ni le français comme langue source, ni l'ensemble de pictogrammes Arasaac comme langue cible. Aucun corpus adapté au français – médical – et à différents ensembles de pictogrammes (Arasaac, Sclera, voire Beta) n'était librement disponible. Dans le chapitre 3, nous avons donc adapté Text-to-Picto au français (objectif 1, sous-objectif 1). Nous l'avons étendu à un grand ensemble de pictogrammes libres de droit de plus en plus utilisés en CAA (sous-objectif 2). Nous avons évalué notre système de base sur des corpus représentant trois cas d'utilisation : des emails, des livres pour enfants et des phrases médicales. [38] Nos premiers résultats étaient comparables à l'état de l'art, même si l'évaluation a montré certaines limites que nous détaillerons plus loin.

Dans le chapitre 4, nous avons exploré des façons de traduire une phrase médicale en pictogrammes en déterminant des stratégies de traduction sur le corpus de BabelDr auprès d'une experte en CAA pour développer le système (sous-objectif 3).

38. Comme nous l'avons déjà mentionné, notre Corpus Médical a été inclus au sein du corpus libre de droit Propicto sur ORTOLANG (Macaire *et al.*, 2023a, 2024a,c). Il a également été intégré au projet Pantagruel (https://pantagruel.imag.fr), qui vise le développement et l'évaluation de modèles de langue inclusifs et multimodaux – écrit, oral, pictogrammes – pour le français (Segonne *et al.*, 2024a,b), adaptés à différents domaines d'application (général, clinique, etc.).

Par exemple, la recherche dans la base d'images d'un pictogramme dont le nom est un synonyme, un hyperonyme, etc. du mot de la phrase à traduire. Nous avons constaté que ces stratégies étaient comparables à certaines opérations de simplification. Afin de spécialiser Text-to-Picto pour la traduction de dialogues médicaux, nous en avons donc implémentées sous la forme de simples règles et nous les avons évaluées (sous-objectif 4). Nous avons aussi analysé l'impact de la substitution lexicale sur la traduction de phrases médicales françaises vers les pictogrammes. Nous avons montré que celle-ci n'était pas efficace, que ce soit en utilisant une ressource générale ou un dictionnaire de paraphrases médicales, cela n'avait pas été testé. Bien que les grands modèles de langue génératifs puissent effectuer diverses tâches, nous avons vu qu'ils ne permettaient pas encore de traduire une phrase en pictogrammes de CAA. Nous avons donc proposé une autre façon de les exploiter à cette fin. Nous avons évalué l'impact de la simplification automatique de textes générée avec des grands modèles de langue génératifs sur la traduction automatique vers les pictogrammes avec Text-to-Picto (sous-objectif 5). Dans certains cas, cette influence est positive (produisant une traduction meilleure) ; dans d'autres, elle est neutre (équivalente), voire négative (pire). Elle reste toutefois limitée, en raison des limites de notre méthodologie et de l'ambiguïté de certains mots dans la phrase source, qui produisent un pictogramme inadapté.

Dans le chapitre 5, nous avons vu qu'aucune étude sur un système de traduction automatique vers les pictogrammes n'avait évalué spécifiquement l'impact de techniques de désambiguïsation lexicale en comparant différents – types de – modèles de langue sur la traduction de mots polysémiques en pictogrammes. Nous avons donc testé une méthode de désambiguïsation lexicale relativement simple, mais adaptée à nos données (sous-objectif 6). Contrairement à d'autres systèmes de traduction vers les pictogrammes (Vaschalde, 2018; Macaire *et al.*, 2023b), notre méthode permet de désambiguïser les noms, les verbes, les adjectifs et les adverbes. Nous avons constitué le premier corpus d'évaluation pour la traduction automatique en pictogrammes de mots polysémiques dans des phrases médicales. Nous avons évalué l'impact de différentes variables : des – types de – modèles de langue, des configurations d'encodage, des – types de – corpus d'entraînement (généraux et/ou médicaux), ainsi que les synsets de plusieurs WordNets français. Les résultats des modèles Word2Vec avec WOLF sur des cas problématiques ont montré des améliorations significatives concernant la précision des mots polysémiques traduits en pictogrammes Arasaac par rapport à ceux traduits avec Text-to-Picto sans désambiguïsation. Avec notre méthode, les modèles BERT peuvent améliorer la précision si l'on encode des phrases associées aux synsets. Toutefois, le rappel était plus faible, car les synsets du WOLF ne comprennent pas assez de phrases représentant des usages. Les grands modèles de langue génératifs ont permis de désambiguïser plusieurs phrases en pictogrammes en utilisant un prompt simple sans exemple avec les mêmes données, tout ceci n'avait pas encore été testé.

Dans le chapitre 6, nous avons finalement évalué des traductions médicales en pictogrammes auprès d'adultes avec une DI dans le cadre de deux études utilisateurs. Nous avons utilisé des modèles linéaires généralisés à effets mixtes pour expliquer la compréhensibilité des pictogrammes. La première étude sur des interactions médicales hors contexte a montré que plusieurs variables ont un impact significatif (objectif 2, sous-objectif 1) : le niveau de représentation (phrase ou pictogramme) ; et en fonction de celui-ci, le niveau de langue, les choix du système de traduction, le type de phrases (interrogatif ou non) et la catégorie morphosyntaxique du mot à traduire. Nous avons évalué pour la première fois des traductions médicales générées par Text-to-Picto et PictoDr auprès de personnes avec une DI. Nous avons mis en évidence le problème de la compositionnalité : les pictogrammes isolés ont été mieux compris que les phrases. Dans la deuxième étude, nous avons évalué plus précisément les traductions de questions en pictogrammes en vue d'une anamnèse du COVID-19 (sous-objectif 2). Nos résultats ont montré que les questions plus spécifiques et celles avec plus de pictogrammes étaient plus difficiles à comprendre. Cependant, les pictogrammes ont été perçus comme utiles dans un contexte médical, y compris pour des personnes verbales. En effet, nous avons vu que les personnes interrogées rencontrent parfois des difficultés de communication avec un médecin. Nous avons donc proposé un corpus et une méthodologie originale avec une facilitatrice, afin d'évaluer directement la compréhension de pictogrammes et la compréhension de phrases en pictogrammes auprès d'adultes verbaux avec une DI. Elle a permis de répondre à nos sous-objectifs et nos hypothèses de recherche, tout en étant adaptée à nos participants (sous-objectif 3).

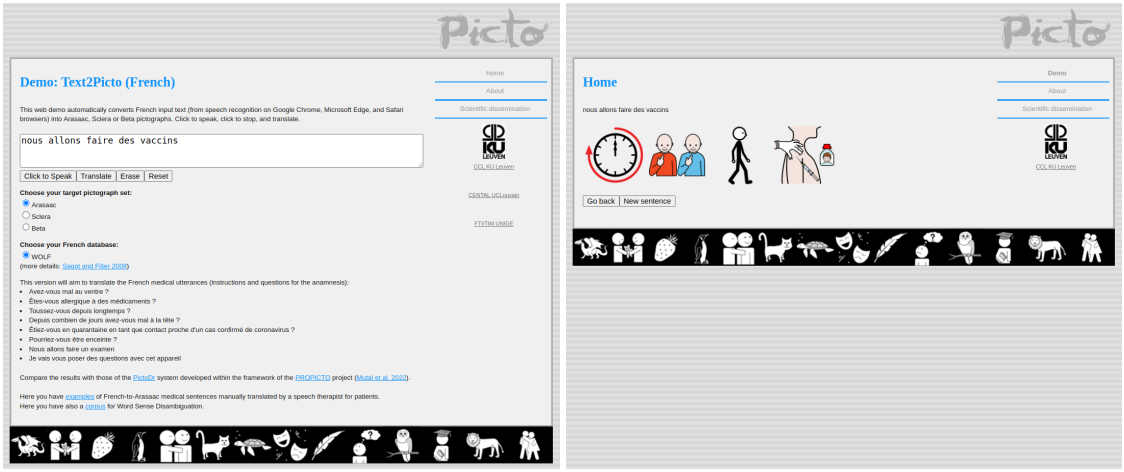
2 Démo de Text-to-Picto

Comme pour les autres versions de Text-to-Picto (Sevens, 2018), nous avons mis la démo de la version 2 du système Text-to-Picto français en ligne (Figure 6.18). [39] Elle propose les trois ensembles de pictogrammes (Arasaac, Sclera et Beta), mais pas nos adaptations liées à la substitution lexicale, ni la simplification automatique via les grands modèles de langue génératifs et elle n'est pas reliée à notre module de désambiguïsation lexicale. Nous avons toutefois ajouté une nouvelle modalité d'entrée.

Pour en faire un système speech-to-picto simple et proposer une deuxième modalité d'entrée à l'utilisateur, nous avons utilisé une *Application Programming Interface (API)* JavaScript de reconnaissance vocale. [40] L'API permet de dicter la phrase à faire traduire automatiquement par Text-to-Picto en exploitant la reconnaissance vocale intégrée dans le navigateur. Elle fonctionne sur Google Chrome, Microsoft Edge

39. Le design du site web a été créé par Leen Sevens, que nous remercions (https://text2picto.ccl.kuleuven.be/text2picto_french/web/french/index_fr_speech.php).

40. Nous remercions également Johanna Gerlach et Hubert Naets pour leur aide.



(a) Entrée : phrase à traduire (Français) (b) Sortie : phrase traduite (Arasaac)

FIGURE 6.18 – Interface de Text-to-Picto

et Safari. Les mots reconnus sont transcrits dans la zone de texte sur le site au fur et à mesure de la dictée jusqu'à ce que l'utilisateur clique sur le bouton « *Click to Stop* ». La phrase est ensuite traduite en pictogrammes si l'utilisateur clique sur « *Translate* », il peut donc la modifier avant. La possibilité de corriger la phrase en amont de la traduction est nécessaire en raison des nombreuses erreurs qui peuvent provenir de la reconnaissance vocale (Vaschalde, 2018). Dans cette thèse, nous n'avons pas développé ce point, mais il serait possible de lier Text-to-Picto à un système de reconnaissance vocale plus robuste et libre de droit comme Whisper d'OpenAI (Radford *et al.*, 2023). La traduction automatique de la parole vers des pictogrammes est néanmoins une tâche complexe, qui soulève des problèmes spécifiques (Ormaechea Grijalba *et al.*, 2023a,b; Macaire *et al.*, 2024b; Ionescu *et al.*, 2024). Tout comme PictoDr ou BabelDr, cette démo pourrait par exemple être exploitée à des fins de recherche : les chercheurs pourraient utiliser le système pour générer, voire « collecter des données et comparer différentes méthodes de communication » (Norré *et al.*, 2021a).

3 Limites et perspectives

Tout au long de cette thèse, nous avons déjà évoqué des limites et des pistes de recherche supplémentaires. Nous proposons toutefois d'en reprendre quelques-unes ici pour encourager d'autres personnes à participer aux recherches sur la traduction automatique vers les pictogrammes ou à contribuer plus généralement à l'amélioration de l'accès à la communication des personnes en situation de handicap.

Concernant les chapitres 1 et 2, notre état de l'art sur la traduction automatique et la simplification automatique de textes ne se voulait pas exhaustif, ces domaines étant très vastes. Cependant, il nous a permis de faire plusieurs constatations. Il nous

semble tout d'abord important de souligner que les recherches en CAA et en TAL sont pluri/interdisplinaires et ne cessent de se développer. Il y a toutefois très peu d'études qui impliquent l'inclusion de beaucoup de participants avec une DI, ceux-ci peuvent avoir un profil très divers. C'est le cas du Trouble du Spectre de l'Autisme (TSA), mais aussi d'autres types de handicap. Les résultats de ces études sont difficilement généralisables et comparables entre eux. Nous avons aussi constaté qu'il y a plus d'études anglosaxones ou américaines (où la CAA est plus développée), qui sont également plus centrées sur les enfants que sur les adultes. En TAL, les recherches évoluent très rapidement, à l'image de la sortie rapide des grands modèles de langue génératifs. Les études sur la traduction automatique de phrases vers les pictogrammes, qui combinent le TAL, la CAA pour le médical, en particulier pour le français sont relativement peu nombreuses et mériteraient, à notre sens, d'être davantage mises en évidence.

Dans Text-to-Picto, nous aurions pu utiliser d'autres métadonnées de l'API d'Arasaac. [41] Notre système de traduction vers les pictogrammes fonctionne avec le WOLF (Sagot et Fišer, 2008), une base lexico-sémantique générale, qui a déjà plus d'une quinzaine d'années. Elle ne comprend donc pas les néologismes comme ceux liés au COVID-19. Nous pensons que la qualité des synsets du WOLF pourrait être améliorée. Il serait possible de la remplacer par la ressource créée pour PictoDr (Gerlach *et al.*, 2023), qui lie des phrases médicales à des concepts de l'*Unified Medical Language System (UMLS)*, eux-mêmes reliés à des pictogrammes Arasaac. Concernant les métriques automatiques, nous aurions pu employer METEOR (Banerjee et Lavie, 2005), qui prend en compte les synonymes et qui a été utilisé dans la tâche ToPicto (Ionescu *et al.*, 2024), ainsi que par Macaire *et al.* (2024a,b). Nous avons en fait choisi d'utiliser les mêmes métriques automatiques et les mêmes procédures d'évaluation que Sevens (2018), afin de pouvoir comparer nos résultats de la version française de Text-to-Picto aux siens. Nous aurions pu toutefois utiliser l'outil SacreBLEU (Post, 2018), afin de calculer les scores BLEU et leur significativité. Bien que nous ayons fait en sorte d'évaluer Text-to-Picto sur des corpus d'évaluation plus grands que ceux de Sevens (2018), la taille de nos corpus reste encore fort limitée, par manque de ressources. De même, nous n'avons utilisé qu'une traduction de référence pour les évaluations automatiques, alors que nous avons vu qu'il existe plusieurs stratégies différentes pour traduire une phrase en pictogrammes. Certaines de nos évaluations manuelles n'ont été réalisées que par un seul annotateur, alors qu'il est toujours préférable d'en avoir plusieurs. Toujours concernant l'évaluation, nous n'avons pas proposé d'études utilisateurs après avoir conçu le système de base de Text-to-Picto, contrairement à Sevens (2018). Or, il est utile d'avoir l'avis de potentiels utilisateurs tout au long du processus de développement.

41. Par exemple, les catégories ou les tags plutôt que les synsets. Il convient de noter que cette API permet maintenant de directement extraire les synsets de différentes versions du WordNet. Il ne serait donc plus nécessaire de passer par les index collaboratifs interlingues comme nous l'avons proposé.

Dans ce travail, nous avons dû nous limiter à la traduction en pictogrammes d'un ensemble de phrases médicales. C'est pourquoi, nous avons décidé d'utiliser un corpus utilisés pour le système BabelDr. Il s'agissait de phrases sans ellipses et sans erreurs. Or, les médecins doivent en faire comme tout le monde. Un correcteur orthographique pourrait être ajouté. Notre corpus n'a pu être traduit en pictogrammes que par une seule logopède, étant donné la difficulté de se déplacer durant la crise sanitaire. Dû au manque de données en pictogrammes, nous n'avons donc pas utilisé l'approche de traduction neuronale, mais une plus simple qui repose sur des ressources linguistiques et un petit nombre de règles adaptées aux spécificités de notre corpus, constitué principalement de questions médicales relativement simples. Sur d'autres phrases, nous pouvons nous attendre à des traductions encore plus complexes. Avec une approche non neuronale, d'autres modules pourraient être explorés pour la traduction automatique de phrases médicales en pictogrammes, comme l'annotation automatique des expressions multi-mots, la prise en compte des ellipses, voire l'extraction de concepts et des techniques de résumé automatique. Nous avons en effet vu que que de courtes séquences de pictogrammes avaient un impact positif significatif sur la compréhension. D'autres perspectives pourraient être la construction d'un corpus synthétique en vue d'entraîner un modèle neuronal ou la collecte de données auprès des personnes concernées. Ces données pourraient par exemple être des phrases en pictogrammes utilisées par des personnes avec une DI dans diverses situations médicales, validées par elles et/ou par un grand nombre de professionnels (para)médicaux.

Dans le chapitre 5, nous avons conçu un module de désambiguïsation lexicale, mais nous ne l'avons pas intégré dans Text-to-Picto. Il s'agit d'une méthode relativement simple et nous pensons que des améliorations seraient possibles. Nous avons déjà évoqué la possibilité de recourir au modèle français pré-entraîné pour cette tâche de Vial *et al.* (2019), lequel a déjà été utilisé par Vaschalde (2018), Vaschalde *et al.* (2018). Des adaptations seraient toutefois nécessaires. Dans les chapitres 4 et 5, nous avons également testé des grands modèles de langue génératifs pour la simplification automatique et la désambiguïsation lexicale. Comme nous voulions comparer plusieurs modèles, nous avons utilisé la plateforme du LMSys et pas l'API de chaque modèle qui permet d'accéder aux dernières versions payantes. Nous avons testé des prompts simples sans exemple. Des améliorations pourraient être proposées (en utilisant des demandes différentes ou plus précises, etc.). Nous avons toutefois relevé certaines limites des grands modèles de langue génératifs. Les exploiter dans le domaine médical ou de la CAA peut être intéressant dans le futur, mais pourrait aussi soulever certaines questions liées à la confidentialité et la diffusion de données privées. Nous n'avons malheureusement pas pu aborder les aspects éthiques dans cette thèse. Néanmoins, nous aimerions mentionner qu'il existe déjà des études sur l'évaluation des risques de ces technologies langagières liées à la CAA et au TAL sur le long

terme (Lefeuvre-Halftermeyer *et al.*, 2016; Antoine *et al.*, 2014). Nous pensons que des réflexions éthiques seront d'autant plus nécessaires pour le domaine médical et de l'aide au handicap avec les avancées rapides des grands modèles de langue génératifs.

Concernant les études utilisateurs du chapitre 6, nous avons déjà détaillé les limites, en particulier celles liées à la méthodologie que nous avons proposée (cf. section 3). Nous pouvons insister sur le fait que nous avions peu de participants, tous avec une DI légère. Nous n'avions pas non plus d'information sur leur handicap. De plus, ils utilisaient tous le langage oral et pas la CAA ; ils ne connaissaient donc pas les pictogrammes utilisés. À Genève, nous avons rencontré des infirmières qui travaillent dans un service de consultation pour les personnes avec une DI. Elles nous disaient qu'elles utilisaient des images issues de SantéBD, mais qu'une des difficultés qu'elles rencontraient était que parfois leurs patients avaient déjà leurs propres images et systèmes de CAA. Pour elles, cela pourrait être plus simple si les foyers et institutions utilisaient le même matériel de communication. Nous avons aussi vu qu'il était difficile d'évaluer si les personnes avec une DI – même verbales – comprenaient ou non les phrases en pictogrammes. Un moyen pour obtenir des informations complémentaires sur la façon dont ils appréhendent une phrase en pictogrammes pourrait passer par des techniques d'oculométrie (*eye-tracking*), des jeux, voire via l'enregistrement de situations de la vie quotidienne par des proches consentants, comme c'est respectivement le cas des projets français AAC4All,[42] GazePlay (Schwab *et al.*, 2018) et ParticipAACtion.[43] Nous pouvons par exemple nous demander s'il est possible et nécessaire de représenter certains concepts médicaux ou qualificatifs pour toutes les personnes avec une DI.

Enfin, nous concluons cette thèse en insistant sur le fait que les traductions médicales de Text-to-Picto et PictoDr pourraient encore être améliorées pour qu'elles soient plus compréhensibles par des adultes verbaux avec une DI qui n'ont pas appris à utiliser les pictogrammes pour communiquer. Dans cette thèse, nous avons tenté de montrer que traduire automatiquement des phrases médicales en pictogrammes est différent que traduire des emails, voire des histoires pour enfants. Les enjeux de la communication orale entre un médecin et un patient ne sont pas les mêmes que ceux de la communication écrite, surtout sur des questions qui pourraient concerner directement la santé de potentiels utilisateurs.

42. Ce projet a pour but de trouver des solutions pour une CAA adaptée à tous (https://www.aac4all.org). Un *work package* concerne l'évaluation des pictogrammes. Il consiste en la conception, ainsi que la réalisation de protocoles expérimentaux en psycholinguistique et neurolinguistique (par exemple, via l'électroencéphalographie ou l'*eye-tracking*) sur la compréhension et l'apprentissage des pictogrammes par des enfants en situation de handicap, dont des patients avec des troubles langagiers, cognitifs, moteurs, etc. Les protocoles et corpus de données seront également publiés.

43. Il a pour but de « répertorier les modalités et les compétences communicatives des personnes en situation de handicap complexe, associant des déficits perceptifs, moteurs et cognitifs » (Schwab, 2021, p. 42). Afin de pouvoir étudier davantage leurs façons de communiquer, un corpus audio-visuel sera créé. Il est prévu que les proches filment ces personnes dans leur quotidien. Le projet adopte une « approche participative et interdisciplinaire » (https://www.participaaction.interaaction.com).

Annexe A

Corpus

Nous donnons un aperçu des corpus français que nous avons construits et utilisés pour le développement, l'amélioration et l'évaluation du système Text-to-Picto :

- Annexe A.1 : un extrait des corpus en pictogrammes Arasaac.
- Annexe A.2 : le corpus traduit en pictogrammes Arasaac par la logopède.
- Annexe A.3 : le corpus simplifié avec des grands modèles de langue génératifs.
- Annexe A.4 : le corpus pour évaluer la désambiguïsation lexicale.

1 Corpus en pictogrammes Arasaac : Email, Livre, Médical

Nous donnons un aperçu du contenu des corpus (Norré *et al.*, 2021d), que nous avons adaptés ou constitués pour évaluer le système de base de Text-to-Picto (cf. chapitre 3) : le Corpus Email (section 3.1), le Corpus Livre (section 3.2) et le Corpus Médical (section 3.3).

#	Corpus Email		Corpus Livre		Corpus Médical	
	Source	**Référence – Arasaac**	**Source**	**Référence – Arasaac**	**Source**	**Référence – Arasaac**
#1	bonjour comment vas-tu?	bonjour comment être tu?	Abel a perdu son doudou	perso avoir perdre_6 son nounours	avez-vous très mal au ventre?	avoir vous très mal_de_ventre_?
#2	qu'as-tu fait aujourd'hui?	qu'as-tu_fait aujourd'hui?	il est l'heure d'aller dormir	être le heure de aller dormir_2	marchez-vous tous les jours?	marcher vous souvent jour?
#3	salut *perso*	salut perso	Abel est inquiet	perso être inquiet	avez-vous eu une radiothérapie?	avoir vous une radiographie?
#4	arrête une fois de m'envoyer stp	arrêter une_fois_de_plus de envoyer moi s'il_vous_plaît	il ne retrouve plus son doudou	il trouver_3 pas son nounours	depuis combien de jours avez-vous de la fièvre?	depuis combien de jour avoir vous de la fièvre_?
#5	viens-tu à ma fête du 22 juin ou pas je dois savoir je peux dire à ma maman	venir tu à ma fête du 22 Juin ou pas je obliger savoir je pouvoir dire à ma maman	il demande à sa sœur Bellina	il demander_4 à sa sœur perso	avez-vous froid?	avoir vous froid_3?
#6	rien pour *perso*	rien pour perso	as-tu vu mon doudou?	avoir tu voir mon nounours?	avez-vous déjà été enceinte?	avoir vous encore_une_fois être enceinte?
#7	salut *perso* tu as un beau chien comment s'appelle le chien ce n'est pas Lassie *perso*	salut perso tu avoir un beau_2 chien comment_t'appelles-tu le chien ce être pas Lassie perso	je ne le retrouve plus et j'aimerais tant le serrer contre moi	je le trouver_3 pas et je aimer_2 beaucoup le prendre_11 contre moi	comment prenez-vous le médicament?	comment prendre vous le médicament?
#8	salut *perso* de mon papa je peux aller à ta fête d'adieu *perso* est-ce que ce n'est pas bien ou très bien *perso* salutations de ta meilleure amie *perso*???	salut perso de mon papa je pouvoir aller à ta fête de au_revoir perso ce être mal ou très bien perso salut de ta meilleur amie perso???	non je ne l'ai pas vu, répond-elle, mais je vais chercher avec toi	non je le avoir pas voir répondre elle mais je aller chercher avec toi	de quelle couleur sont vos selles?	de_quelle_couleur_est-il_? vos caca?
#9	salut *perso* tu as envoyé une lettre et rien	salut perso tu avoir envoyer_2 et rien	Abel cherche son doudou derrière les rideaux.	perso chercher son nounours derrière les rideau	saignez-vous souvent du nez?	sang vous souvent du nez?
#10	salut *perso* comment va ton frère?	salut perso comment être ton frère?	il ne s'y trouve pas	il être pas	vous êtes-vous blessé au visage?	vous être maltraiter au visage?

2 Corpus traduit en pictogrammes par une logopède

Nous présentons les 100 phrases canoniques de BabelDr (corpus source), traduites manuellement en pictogrammes Arasaac par la logopède belge (corpus de référence).[1] Elles ont servi à adapter la partie du Corpus Médical de référence utilisée pour entraîner et évaluer la version 2 du système Text-to-Picto (cf. chapitre 4, sections 2 et 3). Le corpus a été décrit dans Norré *et al.* (2022). Nous donnons les opérations de transformations syntaxiques et lexico-sémantiques que nous avons notées :

- **suppression** : *delete*;
- **insertion** : *insert*;
- **déplacement** : *move*;
- **synonyme** : *synonym*;
- **hyperonyme** : *hyperonym*;
- **hyponyme** : *hyponym*;
- **antonyme** : *antonym*;
- **changement de POS** : *pos*;
- **remplacement d'expression multi-mots par un mot unique** : *single*;
- **autre remplacement** : *replacement.*

#	Source	Référence – Arasaac	Opération(s)
#1	avez-vous des troubles de l'odorat?	vous nez?	delete replacement
#2	à quelle date ont commencé les symptômes?	quel date commencer symptomes?	delete
#3	respirez-vous trop rapidement?	vous respirer augmenter_2 rapidement?	move replacement
#4	quelle était votre température maximale?	thermomètre_médical?	delete replacement
#5	avez-vous une maladie respiratoire chronique?	vous maladie respirer toujours?	delete pos replacement
#6	avez-vous déjà fait une crise cardiaque?	passé vous infarctus?	delete replacement single move
#7	avez-vous trop de tension artérielle?	vous hypertension?	delete single
#8	pouvez-vous indiquer combien de personnes vivent avec vous?	combien personne vivre_2 avec vous?	delete
#9	avez-vous été en contact avec une personne confirmée positive au coronavirus dans les quatorze derniers jours?	passé 14 jour vous parler_2 avec individu plus virus?	delete move replacement hyponym single hyperonym synonym
#10	étiez-vous en quarantaine en tant que contact proche d'un cas confirmé de coronavirus?	vous isoler avec individu plus virus?	delete hyponym hyperonym

1. https://text2picto.ccl.kuleuven.be/text2picto_french/web/french/examples_arasaac.php

#11	avez-vous une toux sèche?	vous toux goutte pas?	delete replacement antonym
#12	avez-vous des allergies connues?	vous allergie allergie_2 allergie_3?	delete insert
#13	est-ce sanglant quand vous toussez?	vous tousser_3 sang?	delete move pos
#14	à combien était votre température la dernière fois que vous l'avez mesurée?	quand vous prendre_la_température thermomètre_médical combien?	delete move replacement single
#15	êtes-vous confus?	vous égal confus?	move replacement
#16	avez-vous des boutons sur la peau?	vous boutons peau?	delete
#17	pouvez-vous énumérer ou écrire les noms de médicaments auxquels vous êtes allergiques?	vous écrire prénom médicaments égal allergie_2?	delete hyponym replacement pos
#18	votre cœur bat-il de manière anormale?	électrocardiogramme mal?	delete replacement
#19	nous vous contacterons par téléphone pour les résultats	nous vous parler_au_téléphone_2 pour information_2	delete single hyperonym
#20	vous devez porter un masque	vous obliger porter_un_masque_2	hyponym single
#21	bonjour je suis le docteur	bonjour_2 je égal docteur	delete replacement
#22	je vais m'occuper de vous aujourd'hui	futur aujourd'hui je s'occuper_2 vous	delete move
#23	je vais vous poser des questions avec cet appareil	futur je vous demander avec appareil	delete single
#24	je vais tester votre audition	futur je voir_2 audiométrie	delete hyperonym replacement
#25	ça ne fera pas mal	faire_mal pas	delete move single
#26	de quel pays venez-vous?	vous où_allez_vous_? quel pays?	delete move replacement
#27	quel âge avez-vous?	vous égal âge?	delete move replacement
#28	avez-vous déjà vu votre dermatologue pour ce cas?	passé vous voir_2 dermatologue?	replacement delete move
#29	êtes-vous déjà venu aux urgences pour ce cas?	passé vous où_allez_vous_? dans hôpital?	replacement hyperonym move
#30	avez-vous eu un scanner?	vous scanner?	delete
#31	comprenez-vous ce que je dis?	vous comprendre_2 je dire_3?	delete move
#32	des gens dans votre entourage sont-ils malades?	gens famille égal malade?	delete hyperonym replacement
#33	quand vous urinez, est-ce que ça brûle?	quand vous uriner_3 brûler_2?	delete
#34	ouvrez les yeux, s'il vous plaît	ouvrir_les_yeux svp	single replacement
#35	je vais palper le ventre	futur je palper ventre	delete
#36	êtes-vous sourd d'une oreille?	vous sourd_3 un oreille?	delete synonym
#37	merci beaucoup	merci_2 beaucoup_2	
#38	la douleur est-elle difficile à situer?	douleur égal difficile situer?	delete replacement
#39	pouvez-vous décrire la douleur?	vous dire_3 douleur?	delete hyperonym
#40	les vertiges sont-ils fréquents?	faire_tourner_la_tête beaucoup_2?	delete replacement
#41	avez-vous mal à la tête après avoir bu?	vous mal_de_tête après boire_2?	delete single
#42	la douleur à la tête augmente-t-elle quand vous êtes fatigué?	quand vous égal fatigué_3 douleur_à_la_tête augmenter_2?	delete move replacement
#43	mangez-vous beaucoup de viande rouge?	vous manger_3 beaucoup_2 viande rouge?	delete move
#44	depuis combien de jours avez-vous des problèmes de respiration?	depuis combien jour vous mal respirer?	delete replacement pos
#45	avez-vous souvent des aphtes?	vous souvent boutons dans bouche?	delete insert
#46	avez-vous une pharyngite?	vous pharyngite?	delete

#47	êtes-vous grippé en ce moment?	maintenant vous égal grippe?	delete move replacement pos
#48	avez-vous le nez qui coule?	vous se_moucher_le_nez_2?	delete single
#49	montrez-moi une position où vous avez moins mal	montrer_2 position_2 vous faire_mal baisser_2	delete replacement move
#50	je vais examiner votre gorge	futur je examiner_4 gorge	delete
#51	la douleur à la tête est-elle toujours la même?	douleur_à_la_tête toujours égal?	delete move replacement synonym
#52	savez-vous où vous vous trouvez?	vous avoir_une_idée_2 où vous égal?	move replacement
#53	la durée de vos règles est-elle normale?	temps règle_2 égal ok?	delete hyperonym replacement synonym
#54	avez-vous actuellement un traitement pour le sida?	maintenant vous traitement pour sida?	delete move synonym
#55	avez-vous déjà avorté?	passé vous avortement?	replacement delete move pos
#56	voyez-vous double?	vous voir_2 doubler_2?	move pos
#57	prenez-vous de l'insuline?	vous injecter_de_l'insuline?	move single
#58	avez-vous trop de sucre dans le sang?	vous beaucoup_2 sucre dans sang?	delete synonym
#59	êtes-vous anxieux?	vous égal anxieux_2?	move replacement
#60	la douleur à la tête vous a-t-elle réveillé?	douleur_à_la_tête réveiller_2 vous?	delete move single
#61	avez-vous été vacciné contre la varicelle?	vous égal vacciner contre varicelle?	delete replacement
#62	devez-vous cracher quand vous toussez?	quand vous tousser_3 vous obliger cracher_3?	move hyponym
#63	de quelle oreille entendez-vous le moins bien?	quel oreille vous entendre_2 baisser_2?	delete move replacement
#64	portez-vous un appareil auditif?	vous avoir appareil_auditif?	delete move hyperonym single
#65	la douleur irradie-t-elle vers la nuque?	douleur où_allez_vous_? vers nuque?	delete replacement
#66	pouvez-vous me montrer avec le doigt où la douleur s'est déplacée?	montrer_2 où douleur où_allez_vous_??	delete replacement
#67	combien de temps dure la douleur?	combien durer douleur?	delete
#68	avez-vous mal quand vous toussez?	quand vous tousser_3 vous faire_mal?	delete move replacement
#69	respirez profondément la bouche ouverte, s'il vous plaît	respirer bouche bien svp	delete move replacement
#70	je vais vous lever la jambe	futur je lever_2 jambe	delete
#71	l'ordinateur va traduire mes questions	ordinateur traduire question	delete
#72	la gorge est rouge?	gorge égal rouge?	delete replacement
#73	la bouche vous brûle?	bouche brûler_2?	delete
#74	avez-vous observé des vésicules dans la bouche?	vous observer_2 boutons dans bouche?	delete hyperonym
#75	toussez-vous depuis longtemps?	depuis longtemps vous tousser_3?	move
#76	est-ce que vos expectorations sont blanches?	toux égal blanc?	delete replacement
#77	vous êtes-vous mordu la langue?	vous mordre_2 langue?	delete
#78	je vais vous demander de marcher, s'il vous plaît	futur vous marcher_3 svp	delete replacement
#79	montrez-moi quel œil est concerné?	montrer_2 œil mal?	delete replacement
#80	y a-t-il des maladies particulières dans votre famille?	famille avoir maladies?	delete move

#81	votre douleur dans la poitrine est-elle fluctuante?	douleur_de_poitrine égal augmenter_2 baisser_2?	delete single replacement insert
#82	la douleur dans la poitrine est-elle récurrente?	douleur_de_poitrine égal toujours?	delete single replacement
#83	avez-vous plus de trente-huit de fièvre?	vous augmenter_2 38 fièvre_3?	delete replacement synonym
#84	suez-vous plus que d'habitude?	vous suer_2 augmenter_2?	delete move replacement
#85	les vomissements sont-elles soulagées par des médicaments?	vomir_2 baisser_2 par médicaments?	delete pos replacement
#86	vomissez-vous?	vous vomir_2?	move
#87	avez-vous des ulcères à l'estomac ou au duodénum?	vous avoir trou estomac? intestin?	delete move hyperonym insert
#88	êtes-vous allergique à des aliments?	vous égal allergie aliment?	delete pos
#89	vous pouvez vous rhabiller	vous s'habiller_2	delete hyperonym
#90	ne bougez pas la tête, s'il vous plaît	bouger_les_mains pas tête svp	delete replacement
#91	à quelle fréquence viennent les douleurs dans la poitrine?	quel encore_une_fois où_allez_vous_? douleur_de_poitrine?	delete replacement single
#92	la douleur est causée par le froid?	froid_3 augmenter_2 douleur?	delete move replacement
#93	c'est la première fois que vous avez fait un malaise?	première fois vous s'évanouir_3?	delete single
#94	devez-vous vous tenir aux meubles pour marcher?	vous obliger tenir_2 meubles pour marcher_3?	delete move
#95	suivez-vous un régime alimentaire?	vous régime aliment?	delete pos
#96	avez-vous des brûlures d'estomac?	vous estomac brûler_2?	delete move pos
#97	l'écoulement d'oreille est-il jaune?	écoulement oreille égal jaune?	delete replacement
#98	sentez-vous quelque chose quand je touche ici?	quand je toucher ici vous faire_mal?	delete replacement move
#99	je vais appuyer ici	futur je toucher ici	delete hyperonym
#100	je vais vous demander de vous coucher sur le ventre, s'il vous plaît	futur vous se_coucher_3 sur ventre svp	delete replacement

3 Corpus simplifié avec des grands modèles de langue génératifs

Ce corpus contient les 100 phrases canoniques que nous avons fait simplifier automatiquement avec 8 grands modèles de langue génératifs (gemini-pro-dev-api, claude-2.1, command-r, gpt-3.5-turbo-0613, llama-2-70b-chat, mistral-medium, qwen1.5-72b-chat, vicuna-33b), avant de les faire traduire automatiquement en pictogrammes Arasaac avec la version 2.1 du système Text-to-Picto (cf. chapitre 4, section 5). Ces phrases sources ont également été utilisées dans l'évaluation manuelle de Text-to-Picto sans et avec la substitution lexicale via la ressource ReSyf (cf. section 4.2.1), seuls les mots de contenu qui ne se trouvaient pas dans le dictionnaire du système ont été évalués (ici en gras).
Légende après traduction avec Text-to-Picto : Traduction pire (rouge) | Traduction équivalente (orange) | Traduction meilleure (vert).

#	Phrase à simplifier|traduire	gemini-pro-dev-api	claude-2.1	command-r	gpt-3.5-turbo-0613	llama-2-70b-chat	mistral-medium	qwen1.5-72b-chat	vicuna-33b
#1	**mangez**-vous **normalement**?	mangez-vous habituellement?	mangez-vous habituellement?	mangez comme d'habitude?	mangez-vous de manière habituelle?	mangez-vous comme d'habitude?	mangez-vous normalement?	avez-vous une alimentation normale?	mangez-vous bien?
#2	avez-vous **toujours mal** au **même endroit**?	votre douleur est-elle toujours localisée au même endroit?	la douleur est-elle toujours au même endroit?	est-ce que la douleur persiste au même endroit?	as-tu toujours mal au même endroit?	avez-vous toujours mal au même lieu?	votre douleur est-elle toujours au même endroit?	est-ce que le mal de tête persiste au même endroit?	avez-vous mal souvent au même endroit?
#3	avez-vous eu un **bleu**?	vous êtes-vous fait un bleu?	vous êtes-vous fait mal?	avez-vous eu une bosse?	avez-vous eu une contusion?	avez-vous eu mal?	avez-vous eu une contusion?	avez-vous été blessé(e)?	as-tu eu un coup?
#4	la douleur au ventre est-elle **intermittente**?	le mal de ventre est-il intermittent?	est-ce que la douleur au ventre vient et repart?	est-ce que la douleur au ventre vient et s'en va?	la douleur au ventre est-elle parfois présente et parfois absente?	la douleur au ventre est-elle temporaire?	la douleur à l'estomac est-elle irrégulière?	est-elle la douleur au ventre intermittente?	est-ce que la douleur au ventre est intermittente?
#5	pouvez-vous me montrer avec le doigt où est la **douleur**?	où est la douleur? montrez-moi.	où avez-vous mal? montrez-moi avec le doigt.	montrez-moi avec votre doigt où vous avez mal.	pouvez-vous me montrer où est la douleur avec votre doigt?	pouvez-vous me montrer où est la douleur?	peux-tu montrer du doigt l'endroit qui te fait mal?	pouvez-vous indiquer où est la douleur en me désignant le lieu avec votre doigt?	pouvez-vous indiquer avec le doigt où est la douleur?
#6	êtes-vous en **bonne santé**?	ça va?	êtes-vous en forme?	êtes-vous en forme?	êtes-vous en forme?	comment allez-vous?	allez-vous bien?	êtes-vous en bonne santé?	es-tu en bonne santé?
#7	avez-vous de l'**asthme**?	souffrez-vous d'asthme?	êtes-vous asthmatique?	souffrez-vous d'asthme?	as-tu de l'asthme?	do you have asthma?	avez-vous de l'asthme?	as-tu de l'asthme?	as-tu de l'asthme?
#8	avez-vous **bu** de l'**alcool**?	avez-vous consommé de l'alcool?	avez-vous consommé de l'alcool?	buvez-vous de l'alcool?	avez-vous consommé de l'alcool?	have you drunk alcohol?	vous avez bu de l'alcool?	as-tu bu de l'alcool?	as-tu bu de l'alcool?
#9	quand vous **urinez**, est-ce que ça **brûle**?	avez-vous des brûlures en urinant?	avez-vous une sensation de brûlure en urinant?	est-ce que ça vous brûle quand vous urinez?	est-ce que ça brûle quand vous urinez?	does peeing burn?	est-ce que ça brûle quand vous urinez?	est-ce que vous ressentez de la brûlure lors de la miction?	est-ce que ça brûle quand vous urinez?
#10	la douleur à la tête est-elle **constante**?	est-ce que le mal de tête est continu?	as-tu mal à la tête tout le temps?	est-ce que la douleur à la tête persiste?	la douleur à la tête est-elle continue?	la tête vous fait-elle mal tout le temps?	la douleur à la tête est-elle continue?	est-ce que la douleur à la tête est constante?	est-ce que la douleur à la tête est constante?
#11	la douleur au ventre **augmente**-t-elle après avoir **mangé**?	le mal de ventre s'aggrave-t-il après les repas?	avoir mal au ventre après manger?	est-ce que ma douleur à l'estomac empire après avoir mangé?	la douleur au ventre augmente-t-elle après le repas?	does belly pain get worse after eating?	est-ce que la douleur à l'estomac s'aggrave après les repas?	est-ce que la douleur au ventre s'aggrave après avoir mangé?	la douleur au ventre s'aggrave-t-elle après manger?
#12	vos **maux** de **tête** sont-ils **causés** par le **bruit**?	le bruit provoque-t-il vos maux de tête?	avez-vous mal à la tête à cause du bruit?	le bruit cause-t-il vos maux de tête?	le bruit cause-t-il vos maux de tête?	les maux de tête sont-ils causés par le bruit?	est-ce que le bruit donne mal à votre tête?	est-ce que le bruit cause vos maux de tête?	est-ce le bruit qui cause tes migraines?

#13	en quelle **année** êtes-vous **arrivé** en **Suisse**?	quand êtes-vous arrivé en Suisse?	quelle année êtes-vous arrivé en Suisse?	quand êtes-vous venu en Suisse?	quand êtes-vous arrivé en Suisse?	quelle année avez-vous arrivé en Suisse?	depuis quelle année êtes-vous en Suisse?	quel année êtes-vous venu en Suisse?	dans quelle année êtes-vous venu en Suisse?
#14	la **douleur** est-elle liée à une **activité sexuelle**?	la douleur est-elle causée par le sexe?	avez-vous ressenti de la douleur pendant les rapports intimes?	est-ce que faire l'amour fait mal?	la douleur est-elle liée au sexe?	est-ce que la douleur est causée par une activité sexuelle?	est-ce que la douleur est causée par une activité sexuelle?	est-ce que la douleur est liée à l'activité sexuelle?	est-ce que la douleur est due à une activité sexuelle?
#15	avez-vous des problèmes d'**incontinence urinaire**?	avez-vous des fuites urinaires?	avez-vous des fuites urinaires?	souffrez-vous d'incontinence urinaire?	avez-vous des problèmes d'incontinence urinaire?	avez-vous des problèmes de vessie?	avez-vous des problèmes de fuite d'urine?	souffrez-vous d'incontinence urinaire?	as-tu des problèmes pour contrôler tes urines?
#16	avez-vous **voyagé récemment**?	avez-vous voyagé dernièrement?	avez-vous voyagé dernièrement?	avez-vous fait du voyage récemment?	avez-vous récemment voyagé?	avez-vous voyagé ces derniers temps?	avez-vous fait un voyage récemment?	avez-vous voyagez récemment?	as-tu voyagé récemment?
#17	vous a-t-on **déjà** fait une **gastroscopie**?	avez-vous déjà subi une gastroscopie?	vous a-t-on déjà examiné l'estomac avec un tube et une caméra?	avez-vous déjà subi un examen gastroscopique?	avez-vous déjà subi une gastroscopie?	avez-vous déjà subi une gastroscopie?	as-tu déjà eu une gastroscopie?	avez-vous déjà subi une gastroscopie?	as-tu déjà eu une gastroscopie?
#18	avez-vous fait un **malaise**?	vous êtes-vous évanoui?	vous êtes-vous senti mal?	êtes-vous mal?	avez-vous eu un malaise?	vous avez eu un malaise?	vous êtes-vous évanoui(e)?	as-tu eu un malaise?	as-tu eu un malaise?
#19	êtes-vous **stressé**?	vous stressez?	vous stressez?	stressez-vous?	êtes-vous anxieux?	are you stressed?	avez-vous du stress?	est-ce que vous êtes stressé?	es-tu stressé?
#20	devez-vous **cracher** quand vous **toussez**?	faut-il cracher en toussant?	faut-il cracher si vous toussez?	toussez-vous le besoin de cracher?	faut-il cracher quand on tousse?	do you spit when you cough?	faut-il cracher en toussant?	est-ce que vous devez cracher lorsque vous toussez?	doit-on cracher lorsqu'on tousse?
#21	êtes-vous **étudiant**?	tu étudies?	êtes-vous étudiant?	voulez-vous étudier?	êtes-vous un étudiant?	are you a student?	êtes-vous étudiant?	êtes-vous étudiant(e)?	es-tu étudiant?
#22	avez-vous des **brûlures** d'**estomac**?	souffrez-vous de brûlures d'estomac?	avez-vous des brûlures à l'estomac?	avez-vous des problèmes d'estomac?	avez-vous des brûlures d'estomac?	vous avez des brûlures?	souffrez-vous de brûlures d'estomac?	avez-vous des brûlures d'estomac?	as-tu des brûlures d'estomac?
#23	êtes-vous **déjà venu** aux **urgences** pour ce cas?	avez-vous déjà consulté les urgences pour ce problème?	êtes-vous déjà venu ici pour ce problème?	avez-vous déjà été aux urgences pour cela?	êtes-vous déjà venu aux urgences pour cela?	avez-vous déjà été aux urgences pour cela?	avez-vous déjà été aux urgences pour ce problème?	avez-vous déjà été aux urgences pour ce problème?	as-tu déjà visité les urgences pour cette situation?
#24	avez-vous de la **fièvre**?	avez-vous de la température?	êtes-vous fiévreux?	vous avez de la fièvre?	as-tu de la fièvre?	do you have a fever?	avez-vous chaud?	as-tu de la fièvre?	as-tu la fièvre?
#25	nous allons faire un **électrocardiogramme**	on va faire un ECG	nous ferons un électrocardiogramme.	nous allons faire un ECG.	nous allons passer un électrocardiogramme.	nous allons faire un ECG.	nous allons passer un électrocardiogramme.	nous réaliserons un électrocardiogramme.	nous allons faire un ECG.
#26	je **reviens** dans quelques minutes	je reviens bientôt	je reviens bientôt.	je serai de retour dans un instant.	je reviens bientôt.	je reviens bientôt.	je serai de retour sous peu.	je suis de retour dans quelques instants.	je reviens bientôt
#27	nous allons faire une **prise** de **sang**	on va te piquer le doigt	nous ferons une prise de sang.	on va prendre votre sang.	nous allons faire une analyse sanguine.	nous allons prendre du sang.	nous allons prendre votre sang.	nous allons réaliser une ponction sanguine.	nous allons prendre du sang.
#28	avez-vous des **douleurs musculaires**?	souffrez-vous de douleurs musculaires?	avez-vous mal aux muscles?	souffrez-vous de douleurs musculaires?	avez-vous mal aux muscles?	have you got muscle pain?	avez-vous mal aux muscles?	experiences any muscle pain?	as-tu des douleurs musculaires?
#29	avez-vous **déjà** eu ce **type** de **douleur**?	avez-vous déjà ressenti cette douleur?	avez-vous déjà ressenti ce genre de douleur?	avez-vous déjà ressenti cette douleur?	avez-vous déjà ressenti cette douleur?	avez-vous déjà eu mal comme ça?	avez-vous déjà eu mal de cette façon?	avez-vous déjà ressenti cette douleur?	as-tu déjà eu cette douleur-là?
#30	depuis combien d'**années**?	depuis quand?	combien d'années?	depuis quand?	depuis quand?	quel est le nombre d'années?	depuis combien d'années?	depuis quand?	depuis quand?
#31	la **douleur irradie**-t-elle au niveau des **testicules**?	la douleur descend-elle dans les testicules?	est-ce que vous avez mal aux testicules?	la douleur s'étend-elle aux testicules?	la douleur se propage-t-elle aux testicules?	la douleur se propage-t-elle jusqu'aux testicules?	est-ce que ça fait mal aux testicules?	est-ce que la douleur se répand aux testicules?	est-ce que la douleur se fait sentir aux testicules?

#32	avez-vous **besoin** d'une **antalgie**?	avez-vous besoin d'un antidouleur?	avez-vous besoin de médicaments contre la douleur?	voulez-vous un analgésique?	avez-vous besoin d'un analgésique?	avez-vous mal?	avez-vous besoin d'un médicament pour soulager la douleur?	besoin d'une antalgie?	besoin-tu d'un analgésique?
#33	avez-vous des **antécédents médicaux**?	avez-vous des problèmes de santé passés?	avez-vous eu des problèmes de santé par le passé?	vous avez des problèmes de santé?	avez-vous des antécédents médicaux?	avez-vous des problèmes de santé dans votre famille?	avez-vous des antécédents de santé?	avez-vous des problèmes de santé passés?	as-tu des antécédents médicaux?
#34	avez-vous **aussi** mal au dos?	votre dos vous fait-il mal aussi?	avez-vous mal au dos vous aussi?	vous avez mal au dos, aussi?	as-tu mal au dos aussi?	avez-vous mal au dos?	vous avez mal au dos aussi?	est-ce que vous avez mal au dos aussi?	as-tu mal au dos?
#35	pouvez-vous **bouger** le **pied**?	bouge ton pied?	pouvez-vous déplacer votre pied?	bougez votre pied, s'il vous plaît?	pouvez-vous déplacer le pied?	pouvez-vous bouger votre pied?	pourriez-vous déplacer votre pied?	pouvez-vous déplacer votre pied?	pouvez-vous bouger la jambe?
#36	avez-vous **maigri**?	as-tu perdu du poids?	avez-vous perdu du poids?	vous avez perdu du poids?	as-tu perdu du poids?	avez-vous perdu du poids?	as-tu maigri?	as-tu perdu du poids?	as-tu maigri?
#37	avez-vous un **rhume**?	êtes-vous enrhumé?	êtes-vous enrhumé?	êtes-vous enrhumé?	as-tu un rhume?	do you have a cold?	est-ce que vous avez un rhume?	as-tu le rhume?	as-tu un rhume?
#38	avez-vous **assez bu dernièrement**?	as-tu bu suffisamment ces derniers temps?	avez-vous bu beaucoup récemment?	avez-vous bien bu?	avez-vous bu suffisamment récemment?	avez-vous bu suffisamment récemment?	avez-vous bien bu dernièrement?	assez bu récemment?	as-tu bu récemment?
#39	avez-vous des problèmes **urinaires**?	avez-vous des difficultés à uriner?	avez-vous des difficultés à uriner?	vous avez des problèmes pour uriner?	avez-vous des soucis avec votre urine?	do you have urinary issues?	avez-vous des problèmes de miction?	experiences des difficultés urinaires?	as-tu des soucis pour uriner?
#40	êtes-vous **allergique** à des **médicaments**?	avez-vous des allergies médicamenteuses?	avez-vous des allergies à des médicaments?	avez-vous déjà eu des réactions allergiques à des médicaments?	êtes-vous allergique à des médicaments?	are you allergic to meds?	avez-vous des allergies médicamenteuses?	avez-vous des allergies aux médicaments?	es-tu allergique à des médicaments?
#41	avez-vous une **toux sèche**?	avez-vous une toux sans mucus?	toussez-vous sans cracher?	vous toussez-vous sec?	avez-vous de la toux sèche?	avez-vous une toux?	avez-vous une toux sans expectoration?	avez-vous de la toux sèche?	as-tu une toux sèche?
#42	**consommez**-vous **régulièrement** de la **drogue**?	prenez-vous de la drogue?	je préfère ne pas donner de conseils ou de recommandations concernant la consommation de drogues.	prends-tu souvent des drogues?	prenez-vous régulièrement de la drogue?	prendre de la drogue?	prends-tu souvent de la drogue?	est-ce que vous prenez de la drogue souvent?	consommes-tu régulièrement de la drogue?
#43	pouvez-vous me montrer ce qui vous **amène**?	pouvez-vous m'expliquer votre présence ici?	qu'est-ce qui vous amène ici?	montrez-moi ce qui vous a amené ici.	pouvez-vous me montrer pourquoi vous êtes ici?	pourriez-vous me dire ce qui vous a amené(e) ici?	qu'est-ce qui vous amène ici?	pouvez-vous me dire pourquoi vous êtes ici?	pouvez-vous me dire ce qui vous préoccupe?
#44	vos **urines** sont-elles **rouges**?	vos urines sont-elles de couleur rouge?	vos urines sont-elles de couleur rouge?	est-ce que votre urine est rouge?	vos urines sont-elles de couleur rouge?	les urines sont-elles rouges?	vos urines sont-elles rouges?	avez-vous des urines rouges?	est-ce que tes urines sont rouges?
#45	avez-vous **déjà** eu une **infection** des **reins**?	avez-vous déjà eu une pyélonéphrite?	avez-vous déjà eu une infection rénale?	avez-vous déjà eu une infection rénale?	avez-vous déjà eu une infection rénale?	avez-vous déjà eu une infection urinaire?	avez-vous déjà eu une infection des reins?	as-tu déjà eu une infection rénale?	as-tu déjà eu une infection rénale?
#46	quelle est l'**intensité** de la **douleur** sur une échelle de zéro à dix, zéro étant le **minimum** et dix le **maximum**?	sur une échelle de 0 à 10, quelle est l'intensité de votre douleur?	sur une échelle de 0 à 10, 0 étant aucune douleur et 10 étant la pire douleur imaginable, à quel niveau évaluez-vous votre douleur?	sur une échelle de un à dix, un étant le minimum et dix le maximum, quelle est l'intensité de votre douleur?	quelle est l'intensité de la douleur sur une échelle de zéro à dix?	qu'est-ce que la douleur, de 0 à 10?	à quelle note évalues-tu l'intensité de ta douleur, sur une échelle de 0 à 10? (avec 0 étant l'absence totale de douleur et 10 la douleur la plus intense possible)	sur une échelle de zéro à dix, comment évaluez-vous l'intensité de la douleur, avec zéro étant le plus faible et dix le plus fort?	quelle est l'intensité de la douleur sur une échelle de zéro à dix?
#47	avez-vous des **difficultés** à **avaler**?	avez-vous du mal à avaler?	avez-vous du mal à avaler?	avez-vous de la difficulté à avaler?	avez-vous du mal à avaler?	avez-vous de la difficulté à manger?	avez-vous des problèmes pour avaler?	est-ce que vous avez du mal à avaler?	prenez-vous mal à avaler?

#48	nous allons faire une **analyse** d'**urine**	on va faire un test d'urine	nous allons analyser de l'urine.	nous allons procéder à un examen des urines.	nous allons analyser l'urine.	nous allons analyser une urine.	nous ferons une analyse urinaire.	nous allons analyser l'urine.	nous allons analyser l'urine.
#49	je vais vous **prescrire** un **médicament**	je vais vous donner un médicament.	je vais vous donner un médicament sur ordonnance.	je vais vous donner un médicament à prendre.	je vais vous donner un médicament.	je vous prescrirai un médicament.	je vais prescrire un médicament pour vous.	je vais vous donner un médicament.	je vais te prescrire un médicament.
#50	avez-vous des **nausées**?	vous sentez-vous nauséeux?	êtes-vous malade?	souffrez-vous de nausée?	avez-vous des nausées?	vous avez des nausées.	avez-vous la nausée?	est-ce que vous avez des nausées?	as-tu des nausées?
#51	**prenez**-vous un **médicament**?	prenez-vous des médicaments?	vous prenez un médicament?	prenez-vous des médicaments?	prenez-vous un médicament?	prenez-vous des médicaments?	prenez-vous un médicament?	est-ce que vous prenez un médicament?	prends-tu un médicament?
#52	avez-vous **mal** quand vous **urinez**?	avez-vous des douleurs en urinant?	avez-vous de la douleur en urinant?	vous avez mal en urinant?	ressentez-vous de la douleur lors de la miction?	vous avez mal en urinant?	avez-vous mal en urinant?	est-ce que vous avez mal lors de la miction?	as-tu mal lorsqu'tu urines?
#53	avez-vous **pris** des **médicaments** contre la **fièvre**?	avez-vous pris des antipyrétiques?	avez-vous pris un médicament pour la fièvre?	vous avez pris quelque chose pour la fièvre?	avez-vous pris des médicaments pour la fièvre?	avez-vous pris des médicaments pour la fièvre?	avez-vous pris un antipyrétique	avez-vous pris des médicaments pour la fièvre?	as-tu pris des médicaments pour la fièvre?
#54	y a-t-il du **sang** dans les **urines**?	les urines contiennent-elles du sang?	y a-t-il du sang dans votre urine?	est-ce que les urines contiennent du sang?	y a-t-il du sang dans les urines?	est-ce qu'il y a du sang dans les urines?	les urines contiennent-elles du sang?	est-ce qu'il y a du sang dans les urines?	est-ce qu'il y a du sang dans les urines?
#55	avez-vous **mangé** quelque chose d'**inhabituel**?	avez-vous mangé quelque chose d'inhabituel?	avez-vous mangé quelque chose de différent/d'inaccoutumé?	avez-vous mangé quelque chose de différent?	avez-vous mangé quelque chose de différent?	avez-vous mangé quelque chose de différent?	avez-vous mangé un plat inhabituel?	avez-vous mangé quelque chose d'exceptionnellement différent?	as-tu essayé quelque chose de différent?
#56	quand avez-vous été **à selles** pour la **dernière fois**?	quand avez-vous eu votre dernier mouvement de selles?	je m'excuse, je ne fournirai pas de suggestions pour simplifier cette phrase.	à quel moment avez-vous eu votre dernière selle?	quand avez-vous fait vos dernières selles?	quand avez-vous été à cheval pour la dernière fois?	quand avez-vous visité selles dernièrement?	quand vous êtes allé(e) à selles pour la dernière fois?	quand est-ce que tu as été à selles pour la dernière fois?
#57	votre **cœur** bat-il de manière **anormale**?	votre rythme cardiaque est-il irrégulier?	votre cœur bat-il bizarrement?	votre cœur bat-il de façon irrégulière?	bat-il votre cœur de manière anormale?	bat-il votre cœur de manière étrange?	votre rythme cardiaque est-il anormal?	est-ce que votre cœur bat de manière anormale?	est-ce que ton cœur bat irrégulièrement?
#58	**toussez**-vous?	avez-vous de la toux?	avez-vous la toux?	êtes-vous en train de tousser?	toussez-vous?	est-ce que vous toussez?	toussez-vous?	tousseriez-vous?	est-ce que vous toussez?
#59	avez-vous des **reflux gastriques**?	souffrez-vous de brûlures d'estomac?	avez-vous des brûlures d'estomac?	vous avez des brûlures d'estomac?	avez-vous des problèmes d'acidité gastrique?	avez-vous des problèmes de reflux?	avez-vous des reflux acides?	experienciez-vous des reflux gastriques?	as-tu des reflux gastriques?
#60	de combien de **mètres** êtes-vous **tombé**?	quelle est la hauteur de votre chute?	de quelle hauteur êtes-vous tombé?	combien de mètres avez-vous chutés?	à quelle distance êtes-vous tombé?	combien de mètres avez-vous tombé?	quelle est la hauteur de votre chute?	de quelle hauteur êtes-vous tombé?	de combien de mètres avez-vous chuté?
#61	**urinez**-vous plus que d'habitude?	avez-vous une fréquence urinaire accrue?	faites-vous pipi plus souvent qu'avant?	vous urinez-vous beaucoup ces derniers temps?	urinez-vous plus souvent?	vous urinez-t-il plus souvent qu'à l'habitude?	passez-vous plus souvent aux toilettes?	est-ce que vous urinez plus souvent que d'habitude?	est-ce que vous pissez plus souvent que d'ordinaire?
#62	avez-vous la **tête** qui **tourne**?	êtes-vous étourdi?	avez-vous des vertiges?	votre tête tourne-t-elle?	votre tête tourne-t-elle?	vous vous sentez déorienté?	avez-vous le vertige?	est-ce que vous avez la nausée?	est-ce que vous avez l'impression que votre tête tourne?
#63	des **gens** dans votre **entourage** sont-ils **malades**?	y a-t-il des malades dans votre entourage?	quelqu'un de proche de vous est-il malade?	y a-t-il des personnes malades près de vous?	y a-t-il des malades dans votre entourage?	est-ce que certains de vos proches sont malades?	y a-t-il des malades dans votre entourage?	est-ce que des personnes autour de vous sont malades?	est-ce que des personnes de votre entourage sont malades?

#64	avez-vous un **gonflement** du **pénis**?	avez-vous un pénis enflé?	je m'excuse, je ne fournirai pas de suggestions pour simplifier cette phrase.	y a-t-il un gonflement de votre pénis?	avez-vous une érection?	I apologize, but I cannot provide a simplification of that phrase as it is not appropriate to discuss or ask about someone's genitalia or any potential medical issues related to them. It's important to respect people's privacy and boundaries, and to avoid asking personal or intrusive questions. Instead, I suggest focusing on asking respectful and appropriate questions that are relevant to the conversation or topic at hand. If you have any other questions or concerns, feel free to ask, and I will do my best to assist you.	votre pénis est-il enflé?	est-ce que votre pénis est gonflé?	as-tu un gonflement du pénis?
#65	avez-vous **aussi mal** au niveau des **reins**?	vos reins vous font-ils mal?	avez-vous mal aux reins?	vous avez mal aux reins?	avez-vous mal aux reins aussi?	avez-vous mal aux reins?	avez-vous mal aux reins?	avez-vous mal aux reins?	as-tu mal au dos?
#66	avez-vous **mangé** des **produits crus**?	avez-vous consommé des aliments non cuits?	avez-vous mangé quelque chose de cru?	avez-vous mangé des aliments crus?	avez-vous mangé des aliments non cuits?	have you eaten raw? [emoji][emoji]	avez-vous mangé des aliments crus?	avez-vous mangé des aliments crus?	as-tu consommé des aliments crus?
#67	avez-vous des **frissons**?	avez-vous froid?	avez-vous froid?	frissons-tu?	as-tu des frissons?	des frissons? [smiley]	est-ce que vous frissonnez?	est-ce que vous avez des frissons?	as-tu froid?
#68	c'est la **première fois** que ça vous **arrive**?	est-ce la première fois que cela vous arrive?	est-ce la première fois que cela se produit?	est-ce que c'est la première fois que vous vivez cela?	c'est votre première fois que ça vous arrive?	ça vous arrive pour la première fois?	c'est votre première fois pour cela?	est-ce que cela vous arrive pour la première fois?	est-ce la première fois pour vous?
#69	la **douleur dans la poitrine diminue**-t-elle avec le temps?	la douleur thoracique s'atténue-t-elle avec le temps?	la douleur dans la poitrine s'atténue-t-elle avec le temps?	est-ce que la douleur à la poitrine s'atténue au fil du temps?	la douleur dans la poitrine diminue-t-elle au fil du temps?	la douleur thoracique s'atténue-t-elle avec le temps?	la douleur à la poitrine s'estompe-t-elle avec le temps?	est-ce que la douleur à la poitrine s'atténue avec le temps?	la douleur poitrine s'atténue-t-elle avec le temps?
#70	avez-vous **déjà vu** votre **médecin** traitant pour ce cas?	avez-vous déjà consulté votre médecin pour ce problème?	avez-vous consulté votre médecin à propos de ce cas?	avez-vous déjà consulté votre médecin pour cela?	avez-vous déjà consulté votre médecin traitant pour cela?	avez-vous déjà vu votre médecin pour cette raison?	avez-vous déjà vu votre docteur pour ce problème?	avez-vous consulté votre médecin pour ce problème?	as-tu déjà consulte ton médecin pour ce problème?
#71	la **douleur au ventre irradie**-t-elle vers les **dents**?	la douleur abdominale se propage-t-elle aux dents?	est-ce que la douleur au ventre se propage vers les dents?	la douleur au ventre rayonne-t-elle vers les dents?	la douleur au ventre se propage-t-elle aux dents?	la douleur au ventre se propage-t-elle jusqu'aux dents?	la douleur à l'estomac atteint-elle les dents?	est-ce que la douleur au ventre se répand jusqu'aux dents?	douleur ventre → dents?
#72	avez-vous de la **famille** en **Suisse**?	avez-vous des proches en Suisse?	avez-vous de la famille en Suisse?	vous avez de la famille qui vit en Suisse?	avez-vous des proches en Suisse?	do you have family in Switzerland?	avez-vous de la famille en Suisse?	avez-vous de la famille en Suisse?	as-tu des parents en Suisse?
#73	avez-vous **déjà** eu une **maladie sexuellement transmissible**?	avez-vous déjà contracté une IST?	avez-vous déjà eu une MST?	avez-vous déjà attrapé une infection sexuellement transmissible?	avez-vous déjà eu une MST?	avez-vous déjà eu une MST?	avez-vous déjà contracté une maladie sexuellement transmissible?	as-tu déjà contracté une MST?	as-tu déjà eu une MST?

#74	avez-vous de la **bronchite chronique**?	souffrez-vous de bronchite chronique?	avez-vous une bronchite chronique?	souffrez-vous de bronchite à long terme?	avez-vous une bronchite chronique?	do you have chronic bronchitis?	avez-vous une bronchite chronique?	avez-vous bronchite chronique?	as-tu de la bronchite chronique?
#75	avez-vous de la **faiblesse musculaire**?	avez-vous des muscles faibles?	êtes-vous faible musculairement?	souffrez-vous de faiblesse musculaire?	avez-vous des muscles faibles?	have you got muscle weakness?	avez-vous des faiblesses musculaires?	avez-vous des muscles faibles?	as-tu des problèmes de muscles faibles?
#76	avez-vous **déjà** eu un **caillot sanguin**?	avez-vous déjà eu une thrombose?	avez-vous déjà fait un caillot de sang?	avez-vous déjà eu un problème de caillot sanguin?	avez-vous déjà eu un caillot sanguin auparavant?	avez-vous déjà eu un caillot de sang?	avez-vous déjà eu un caillot?	avez-vous déjà eu un caillot de sang?	as-tu déjà eu un caillot de sang?
#77	avez-vous eu une **opération** de l'**appendicite**?	avez-vous été opéré de l'appendicite?	vous a-t-on opéré de l'appendicite?	avez-vous subi une opération pour enlever votre appendice?	avez-vous subi une appendicectomie?	avez-vous eu une appendicite opérée?	avez-vous subi une appendicectomie?	as-tu eu une opération d'appendicite?	as-tu eu une opération pour l'appendicite?
#78	avez-vous reçu un **coup** à la **tête**?	vous avez été frappé à la tête?	vous êtes-vous blessé à la tête?	vous avez eu un coup sur la tête?	avez-vous été frappé à la tête?	have you hit your head?	vous êtes-vous pris un coup à la tête?	as-tu été frappé à la tête?	as-tu reçu un coup sur la tête?
#79	est-ce que vos **expectorations** sont **vertes**?	vos expectorations sont-elles vertes?	crachez-vous du vert?	vos expectorations sont-elles vertes?	vos expectorations sont-elles vertes?	do you expect green? [emoji]	vos crachats sont-ils verts?	avez-vous des expectorations vertes?	est-ce que tes expectorations sont vertes?
#80	à combien était votre **température** la **dernière fois** que vous l'avez **mesurée**?	quelle était votre température lors de votre dernière prise?	quelle était votre dernière température mesurée?	la dernière fois que vous avez pris votre température, à combien était-elle?	quelle était votre température lors de votre dernière mesure?	quelle était votre température la dernière fois où vous l'avez mesurée?	quelle était votre dernière température?	quel était votre température la dernière fois que vous l'avez vérifiée?	quelle était votre température la dernière fois?
#81	à quelle **date** ont **commencé** les **symptômes**?	quand ont commencé les symptômes?	quand est-ce que les symptômes ont commencé?	quand avez-vous commencé à ressentir les symptômes?	quand les symptômes ont-ils commencé?	quand les symptômes ont commencé?	depuis quand les symptômes sont-ils présents?	quand sont apparus les symptômes?	quand ont commencé les symptômes?
#82	avez-vous **déjà** eu une **transfusion**?	avez-vous déjà reçu du sang?	avez-vous déjà reçu du sang?	avez-vous déjà subi une transfusion?	avez-vous déjà été transfusé?	avez-vous déjà eu une transfusion?	avez-vous déjà eu une transfusion sanguine?	avez-vous déjà subi une transfusion sanguine?	as-tu déjà eu une transfusion?
#83	avez-vous des problèmes de **thyroïde**?	souffrez-vous de problèmes thyroïdiens?	votre thyroïde vous pose-t-elle des problèmes?	votre thyroïde fonctionne-t-elle correctement?	avez-vous des soucis thyroïdiens?	avez-vous des problèmes de thyroïde?	avez-vous des troubles thyroïdiens?	avez-vous des problèmes avec votre thyroïde?	as-tu des problèmes de thyroïde?
#84	avez-vous du **cholestérol**?	avez-vous un taux de cholestérol élevé?	avez-vous trop de cholestérol?	vous avez-vous déjà fait tester votre taux de cholestérol?	as-tu du cholestérol?	do you have cholesterol?	avez-vous un taux de cholestérol élevé?	avez-vous du cholestérol?	as-tu du cholestérol?
#85	avez-vous du **mal** à **dormir** la **nuit**?	avez-vous des difficultés à dormir?	avez-vous des difficultés à vous endormir le soir?	dormez-vous bien la nuit?	dormez-vous mal la nuit?	do you have trouble sleeping?	avez-vous des difficultés à dormir la nuit?	est-ce que vous avez des difficultés à dormir pendant la nuit?	difficulté à dormir la nuit?
#86	avez-vous eu des **épisodes** de **paralysie**?	avez-vous déjà été paralysé?	avez-vous déjà été paralysé?	avez-vous déjà été paralysé?	avez-vous été paralysé?	avez-vous eu des problèmes de paralysie?	avez-vous eu des paralysies?	avez-vous expérimenté de la paralysie?	as-tu eu des fois où tu n'arrives pas à bouger?
#87	avez-vous eu un **examen radiologique** du **thorax**?	avez-vous passé une radiographie des poumons?	avez-vous passé une radio des poumons?	avez-vous eu des rayons X du thorax?	avez-vous passé une radio du thorax?	avez-vous eu une radiographie thoracique?	avez-vous eu une radio du thorax?	avez-vous subi une radiographie de votre poitrine?	as-tu eu un examen X du thorax?
#88	avez-vous **saigné**?	as-tu saigné?	saignez-vous?	saignez-vous?	avez-vous eu une saignée?	saigné?	avez-vous saigné?	avez-vous versé du sang?	as-tu saigné?
#89	avez-vous **vomi** à **plusieurs reprises**?	avez-vous vomi plusieurs fois?	je vous propose de reformuler la phrase de manière plus neutre.	vous avez vomi souvent?	avez-vous vomi plusieurs fois?	avez-vous vomi plusieurs fois?	avez-vous souvent vomi?	avez-vous vomis plusieurs fois?	as-tu vomi plusieurs fois?
#90	**buvez**-vous **régulièrement** des boissons **alcoolisées**?	consommez-vous de l'alcool régulièrement?	consommez-vous de l'alcool régulièrement?	buvez souvent des boissons alcoolisées?	buvez-vous souvent de l'alcool?	boivent-vous souvent des boissons alcoolisées?	consommez-vous souvent des boissons alcoolisées?	consommez-vous souvent des boissons alcoolisées?	boissons alcoolisées : buvez-vous régulièrement?

#91	de quelle **couleur** sont vos **vomissements**?	quelle est la couleur de vos vomissements?	je m'excuse, je ne fournirai pas de réponse pour cette question.	quelle couleur ont vos vomissements?	quelle est la couleur de vos vomissements?	quelle couleur ont vos vomissements?	quelle est la couleur de vos vomissements?	quelle couleur ont vos vomissements?	quelle couleur ont tes vomissements?
#92	des **gens** dans votre **entourage** ont-ils les **mêmes symptômes**?	avez-vous des proches présentant des symptômes similaires?	est-ce que des gens autour de vous ont les mêmes symptômes?	y a-t-il des gens dans votre vie qui ont les mêmes symptômes que vous?	est-ce que certaines personnes de votre entourage ont les mêmes symptômes?	les gens autour de vous ont-ils les mêmes symptômes?	y a-t-il d'autres personnes autour de vous avec les mêmes symptômes?	avez-vous des proches qui ont les mêmes symptômes?	personnes proches souffrent-elles des mêmes symptômes?
#93	êtes-vous **essoufflé**?	vous êtes à bout de souffle?	avez-vous le souffle court?	êtes-vous sans souffle?	êtes-vous fatigué?	are you out of breath?	avez-vous le souffle court?	êtes-vous fatigué(e)?	es-tu essoufflé?
#94	êtes-vous **grippé** en ce moment?	avez-vous la grippe?	avez-vous la grippe en ce moment?	êtes-vous actuellement enrhumé?	êtes-vous actuellement grippé?	are you feeling under the weather right now?	êtes-vous malade de la grippe actuellement?	est-ce que vous avez la grippe maintenant?	es-tu atteint de grippe actuellement?
#95	je vais **palper** le **ventre**	je vais examiner le ventre	je vais toucher le ventre.	je vais toucher mon ventre.	je vais toucher le ventre.	je vais toucher mon ventre.	je vais examiner l'abdomen par palpation.	je vais examiner le ventre en touchant doucement.	je vais palper le ventre
#96	la **couleur** de vos **urines** est-elle **normale**?	vos urines sont-elles de couleur normale?	vos urines ont-elles une couleur normale?	est-ce que votre urine a une couleur normale?	vos urines sont-elles normalement colorées?	est-ce que la couleur de vos urines est normale?	la couleur de vos urines est-elle habituelle?	est-ce que vos urines ont une couleur normale?	est-ce que votre urine a une couleur normale?
#97	la **douleur** a-t-elle **commencé progressivement**?	la douleur est-elle apparue graduellement?	la douleur s'est-elle installée petit à petit?	la douleur est-elle apparue progressivement?	la douleur a-t-elle commencé de manière progressive?	la douleur a commencé lentement?	la douleur a-t-elle commencé lentement?	est-ce que la douleur est apparue progressivement?	a-t-elle commencé progressivement?
#98	la douleur au ventre est-elle **pulsatile**?	le mal de ventre est-il pulsé?	est-ce que la douleur au ventre est rythmique?	est-ce que la douleur à l'estomac est pulsée?	la douleur au ventre pulse-t-elle?	la douleur au ventre bat-elle?	la douleur à l'estomac est-elle en pulsations?	est-ce que la douleur au ventre est pulsatile?	est-ce que la douleur au ventre palpite?
#99	les **vertiges** sont-ils **permanents**?	les vertiges sont-ils durables?	les vertiges durent-ils toujours?	les vertiges sont-ils continus?	les vertiges sont-ils constants?	les vertiges sont-ils durables?	sont-ils permanents, ces vertiges?	sont-ils des vertiges permanents?	est-ce que les vertiges durent toujours?
#100	nous allons **prévoir** une **consultation** chez un **médecin** spécialiste	nous allons consulter un spécialiste.	nous prendrons rendez-vous avec un médecin spécialiste.	nous allons voir un médecin spécialiste.	nous allons prendre rendez-vous chez un médecin spécialiste.	nous allons consulter un médecin spécialiste.	nous prévoyons de voir un spécialiste.	nous irons consulter un médecin spécialiste.	nous allons planifier une visite au médecin spécialiste.

4 Corpus à désambiguïser

En plus des 100 lemmes, de la phrase à désambiguïser issue de BabelDr et des synsets corrects (entre 1 à 4 par lemme) ci-dessous (cf. chapitre 5, section 3.1),[2] ce corpus – disponible aux formats TSV et HTML – comprend également : les catégories du discours de ces lemmes (NOM : -n, VER : -v, ADJ : -a, ADV : -r), les identifiants des bons pictogrammes Arasaac associés à ces synsets, parfois plusieurs autres exemples de phrases, leurs lemmes tokenisés par TreeTagger, ainsi que des exemples de sens/-pictogrammes (in)corrects générés par Text-to-Picto sans désambiguïsation lexicale (en mode AZ et ZA). Le corpus est aussi présenté dans Norré *et al.* (2023).

#	Lemme	Phrase à désambiguïser	Synset(s) correct(s)
#1	alcool	avez-vous bu de l'**alcool**?	07884567-n
#2	crème	avez-vous appliqué une **crème** sur la lésion?	03128519-n
#3	perdre	avez-vous **perdu** de la sensibilité?	02287618-v 02287789-v
#4	opérer	il ne faut pas vous **opérer**	00083809-v
#5	café	buvez-vous beaucoup de **café**?	07929519-n 07929351-n
#6	eau	quand vous sentez la sonde dans votre gorge, avalez un peu d'**eau**	07935504-n 14845743-n
#7	interprète	nous allons appeler un **interprète**	10212501-n
#8	répondre	veuillez **répondre** par oui ou par non, en hochant la tête ou avec des gestes	00815686-v 00636279-v
#9	doigt	suivez mon **doigt** avec les yeux sans bouger la tête, s'il vous plaît	05566504-n
#10	voir	**voyez**-vous double?	02129289-v
#11	cœur	nous allons écouter votre **cœur**	05388805-n
#12	vert	suivez la ligne **verte**	00375969-a
#13	faire	je vais vous montrer comment **faire** l'injection d'insuline	01619014-v 01621555-v 02560585-v
#14	examiner	êtes-vous d'accord que je vous **examine**?	02131279-v
#15	appliquer	avez-vous **appliqué** une crème sur la lésion?	01363648-v
#16	chambre	nous allons vous installer dans une **chambre**	02821627-n
#17	penser	avez-vous **pensé** à arrêter de boire?	00628491-v 00608808-v 00708538-v 01637166-v
#18	boire	avez-vous pensé à arrêter de **boire**?	01171183-v
#19	mesurer	nous allons vous **mesurer**	00647094-v
#20	prévoir	nous allons **prévoir** une consultation chez le dentiste	00708538-v
#21	corps	avez-vous observé des grains de beauté sur le **corps**?	05216365-n 05217168-n
#22	protéger	vous **protégez**-vous la peau du soleil?	01128193-v
#23	utiliser	**utilisez**-vous de la crème solaire?	01158872-v
#24	laver	il faut vous **laver** tout le corps avec la lotion pendant deux jours	00036362-v
#25	sentir	les lésions **sentent**-elles mauvais?	02124748-v
#26	foie	avez-vous une inflammation du **foie**?	05385534-n
#27	gratter	avez-vous les yeux qui **grattent**?	02119874-v
#28	pilule	prenez-vous la **pilule**?	03936895-n

2. https://text2picto.ccl.kuleuven.be/text2picto_french/web/french/wsd_arasaac.php

#29	venir	êtes-vous déjà **venu** aux urgences pour ce cas?	01849221-v
#30	beaucoup	mangez-vous **beaucoup** de viande rouge?	00059171-r
#31	souvent	avez-vous **souvent** des aphtes?	00035058-r
#32	savoir	**savez**-vous où vous vous trouvez?	00594621-v
#33	document	pourriez-vous me montrer un **document** avec votre adresse?	06477371-n 06470073-n
#34	entendre	de quelle oreille **entendez**-vous le moins bien?	02169702-v 02171039-v
#35	traduire	l'ordinateur va **traduire** mes questions	00959827-v
#36	famille	y a-t-il des maladies particulières dans votre **famille**?	07970406-n
#37	soulager	les vomissements sont-elles **soulagées** par des médicaments?	00064095-v
#38	vomir	**vomissez**-vous du sang?	00076400-v
#39	mois	combien de kilos avez-vous perdu ces six derniers **mois**?	15207872-n 15206296-n
#40	campagne	habitez-vous à la **campagne**?	08645033-n
#41	consommer	**consommez**-vous régulièrement de la drogue?	01156834-v
#42	visage	suez-vous au niveau du **visage**?	05600637-n
#43	légume	mangez-vous des **légumes**?	07707451-n 12212361-n
#44	alimentation	avez-vous des troubles de l'**alimentation**?	07555863-n
#45	tomber	êtes-vous **tombé**?	01972298-v 01984317-v 01984119-v
#46	marcher	avez-vous des difficultés à **marcher**?	01904930-v 01882170-v 01835496-v
#47	aujourd'hui	combien de fois avez-vous vomi **aujourd'hui**?	00048475-r
#48	vouloir	**voulez**-vous vous coucher ici?	01825237-v 00709205-v
#49	autre	avez-vous d'**autres** problèmes de santé?	02069355-a 02070188-a
#50	endroit	avez-vous reçu un coup à l'**endroit** de la douleur?	05611822-n
#51	asseoir	êtes-vous souvent **assis**?	02692882-v
#52	pâte	mangez-vous beaucoup de riz ou de **pâtes**?	07698915-n
#53	poitrine	avez-vous très mal dans la **poitrine**?	05554405-n
#54	drôle	avez-vous un **drôle** de goût dans la bouche?	00968010-a
#55	opération	avez-vous eu une **opération** du ventre?	00671351-n 03850245-n
#56	mètre	de combien de **mètres** êtes-vous tombé?	13659162-n
#57	âge	quel **âge** ont vos enfants?	15145171-n
#58	enfant	quel âge ont vos **enfants**?	09918554-n
#59	amener	pouvez-vous me montrer ce qui vous **amène**?	01999798-v
#60	masque	vous devez porter un **masque**	03725035-n
#61	contraception	prenez-vous un moyen de **contraception**?	00852181-n
#62	membre	avez-vous des fourmillements dans les **membres**?	05559908-n
#63	repos	avez-vous des vertiges au **repos**?	15274074-n
#64	petit	urinez-vous souvent en **petite** quantité?	01391351-a 02386612-a
#65	travail	votre **travail** se passe-t-il bien?	00584367-n 00582388-n
#66	rincer	vous **rincez**-vous la bouche après avoir utilisé le médicament?	00037138-v
#67	côte	vous êtes-vous déjà cassé une **côte**?	05279026-n
#68	question	l'ordinateur va traduire mes **questions**	07193596-n 07196682-n
#69	avant	votre dos s'est déjà bloqué **avant**?	00060939-r
#70	partie	avez-vous des difficultés à bouger une **partie** de votre corps?	05867413-n
#71	maison	avez-vous des animaux à la **maison**?	03544360-n

#72	cheville	avez-vous eu un bleu sur la **cheville**?	05578442-n
#73	seringue	utilisez-vous des **seringues** usagées?	04376876-n
#74	soleil	vous protégez-vous la peau du **soleil**?	11485367-n 09450163-n
#75	poids	quel est votre **poids**?	05026843-n
#76	température	quelle était votre **température** maximale?	05011790-n
#77	organe	avez-vous reçu une greffe d'**organes**?	05297523-n
#78	seconde	vos douleurs durent-elles quelques **secondes**?	15246853-n
#79	place	la douleur change-t-elle de **place**?	00721431-n 08621598-n 05611822-n
#80	cuisse	avez-vous parfois mal à l'arrière de la **cuisse**?	05562249-n
#81	terminer	avez-vous **terminé** les antibiotiques?	02609764-v
#82	donner	nous allons vous **donner** de l'oxygène avec un masque	02199590-v 02230772-v
#83	accouchement	votre **accouchement** s'est-il bien passé?	14048441-n
#84	lentille	il faut enlever les **lentilles**	03094159-n
#85	rester	vous devez **rester** couché pour le moment	02637592-v
#86	soulever	avez-vous **soulevé** quelque chose de lourd?	01974062-v
#87	date	à quelle **date** ont commencé les symptômes?	15159583-n
#88	fatiguer	la douleur à la tête augmente-t-elle quand vous êtes **fatigué**?	00075021-v
#89	scanner	nous allons faire un **scanner**	00901476-n
#90	vie	avez-vous déjà eu une fracture dans votre **vie**?	15140744-n
#91	situer	la douleur est-elle difficile à **situer**?	01759326-v
#92	sœur	votre mère ou votre **sœur** a-t-elle eu un cancer du sein?	10602985-n
#93	actuellement	avez-vous **actuellement** un traitement pour le sida?	00048475-r
#94	heure	depuis combien d'**heures** avez-vous mal?	15129927-n
#95	présenter	si des membres de votre entourage vivant sous le même toit **présentent** les mêmes symptômes, ils doivent consulter	02148788-v
#96	canne	je vais vous prescrire une paire de **cannes** anglaises	04546194-n
#97	mouvement	la douleur au ventre augmente-t-elle avec certains **mouvements**?	06876309-n
#98	faible	vos jambes sont-elles **faibles**?	02324397-a 02325097-a
#99	flotter	vos selles **flottent**-elles?	01904293-v
#100	tourner	avez-vous l'impression que tout **tourne** autour de vous?	02089984-v

Annexe B

Questionnaires de recherche

Cette partie comprend au moins un exemplaire de tous les documents que nous avons conçus pour les deux études utilisateurs du chapitre 6. La première évaluation (section 1) fait appel aux annexes 1 à 6; la deuxième (section 2) aux annexes 7 à 10 :

- Annexe B.1 : les phrases et pictogrammes évalués, incluant toutes les phrases en pictogrammes des deux questionnaires (1.1), un exemple d'une liste de phrases et de pictogrammes (1.2), un exemple montrant comment les phrases en pictogrammes étaient présentées au participant (1.3), ainsi qu'un exemple d'une liste de contrôle pour un participant (1.4).
- Annexe B.2 : la grille des participants.
- Annexe B.3 : le guide du déroulement des séances.
- Annexe B.4 : les instructions détaillées.
- Annexe B.5 : le formulaire de consentement.
- Annexe B.6 : un exemple de retranscriptions manuelles et automatiques.
- Annexe B.7 : le formulaire COVID-19 créé par les Hôpitaux Universitaires de Genève (HUG), qui a été partiellement traduit en pictogrammes avec PictoDr.
- Annexe B.8 : les phrases évaluées liées COVID-19, incluant toutes les phrases en pictogrammes (8.1), la liste de phrases et de pictogrammes (8.2), ainsi qu'un exemple montrant comment les phrases en pictogrammes étaient présentées au participant (8.3).
- Annexe B.9 : le guide du déroulement des séances.
- Annexe B.10 : les instructions détaillées.

1 Évaluation de deux traductions – Phrases

1.1 Toutes les phrases en pictogrammes

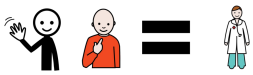
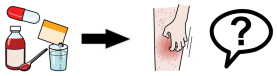
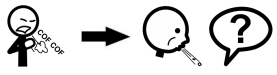
Variables	Text-to-Picto	PictoDr
1 | 4 | 5 | 7 | 8	BONJOUR JE SUIS LE MÉDECIN  bonjour | je | **être** | médecin/docteur	 bonjour | je | **médecin/docteur**
1 | 3 | 5 | 6	NOUS ALLONS FAIRE DES VACCINS **[futur]** | nous | faire | vaccin	 nous | **(faire) vaccin**
1 | 4 | 5 | 7 | 8	VOUS NE POUVEZ PAS BOIRE, PAS MANGER, NI FUMER vous | **ok** | non | boire | non | manger | non | fumer vous | **non** | boire | et | manger | et | fumer
1 | 3 | 5 | 7 | 8	JE VAIS VOUS PRESCRIRE DES GOUTTES **[futur]** | je | prescrire | goutte	 je | prescrire | **goutte**
1 | 4 | 5 | 6	VOUS DEVEZ (RE)FAIRE LE BANDAGE vous | **devoir** | (re)faire | bandage	 vous | **(re)faire le bandage**
1 | 5 | 6 | 7 | 8	MANGEZ-VOUS BEAUCOUP DE VIANDE ROUGE ? vous | manger | beaucoup | viande | **rouge** | ? vous | **manger** | beaucoup | viande rouge | ?
1 | 3 | 5 | 6 | 7	AVEZ-VOUS DÉJÀ FAIT UNE CRISE CARDIAQUE ? **[passé]** | vous | déjà | faire | crise cardiaque | ? **déjà** | vous | (faire) crise cardiaque | ?
1 | 5 | 7 | 8	AVEZ-VOUS DE LA FIÈVRE DEPUIS DEUX JOURS ? vous | fièvre/température | **depuis** | deux | jour | ? vous | fièvre/température | deux | **jour** | ?
1 | 2 | 4 | 5 | 7 | 8	ÊTES-VOUS ALLERGIQUE À DES MÉDICAMENTS ? vous | être | allergie | **médicament** | ?  médicament | causer | **allergie** | ?
1 | 2 | 4 | 5 | 7 | 8	DEVEZ-VOUS CRACHER QUAND VOUS TOUSSEZ ? vous | **devoir** | cracher | quand | vous | tousser | ?  **tousser** | causer | cracher | ?

TABLE B.1 – Questionnaire 1 : traductions de Text-to-Picto et PictoDr

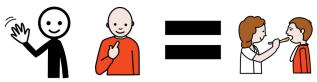
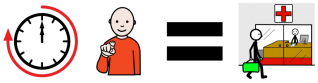
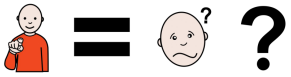
Variables	Text-to-Picto	PictoDr
1 | 4 | 5 | 7 | 8	BONJOUR JE SUIS L'INFIRMIÈRE  bonjour | je | **être** | infirmière	 bonjour | je | **infirmière**
1 | 3 | 5 | 7 | 8	NOUS ALLONS ÉCOUTER VOTRE CŒUR **[futur]** | nous | écouter | cœur	 nous | écouter | **cœur**
1 | 4 | 5 | 7 | 8	IL NE FAUT PAS VOUS OPÉRER vous | **devoir** | non | vous | opérer/opération	 vous | **non** | opérer/opération
1 | 3 | 5 | 6	JE VAIS EXAMINER VOTRE GORGE **[futur]** | je | examiner | gorge	 je | **examiner gorge**
1 | 3 | 4 | 5 | 8	VOUS ALLEZ ÊTRE HOSPITALISÉ  **[futur]** | vous | être | hospitalisé/rester hôpital	 vous | **hospitalisé/rester hôpital**
1 | 4 | 5 | 7	ÊTES-VOUS CONFUS ?  vous | **être** | confus |?	 vous | **confus** |?
1 | 5 | 6	AVEZ-VOUS LES YEUX ROUGES ? vous | œil/yeux | **rouge** |?	 vous | **œil/yeux rouge(s)** |?
1 | 3 | 5 | 6 | 7	AVEZ-VOUS DÉJÀ FUMÉ DU TABAC ? **[passé]** | vous | déjà | fumer | tabac/cigarettes |?	 **déjà** | vous | fumer |?
1 | 5 | 7 | 8	AVEZ-VOUS MAL DEPUIS DEUX SEMAINES ? vous | avoir mal | **depuis** | deux | semaine |?	 vous | avoir mal | deux | **semaine** |?
1 | 5 | 6	COMPRENEZ-VOUS CE QUE JE DIS ? vous | **comprendre** | je/moi | dire |?	 vous | **comprendre ce que je dis** |?

TABLE B.2 – Questionnaire 2 : traductions de Text-to-Picto et PictoDr

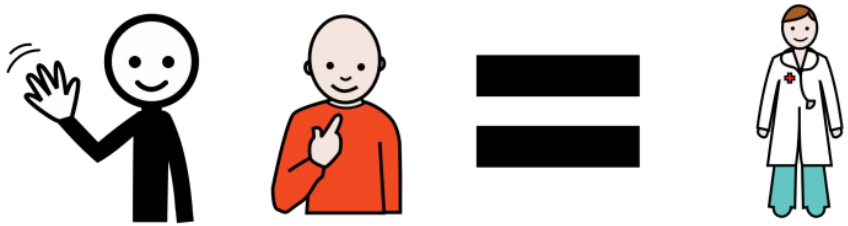
1.2 Liste de phrases et de pictogrammes : exemple

Questionnaire A (P1)

En bas à droite, figure le numéro de la phrase (1-20), suivi du numéro du pictogramme à évaluer dans la phrase. Exemple : 1-3, phrase 1 et picto n°3 (être/suis) à demander au participant.

N° Phrase	Sens de la phrase évaluée (sens des pictogrammes individuels)	N° Picto évalué
0-ex.	avez-vous mal au ventre quand vous marchez ? OU (est-ce que) marcher donne mal au ventre ? (**marcher** | causer/donner | mal_au_ventre | ?)	1
1	bonjour je suis le médecin (bonjour/salut | je/moi/médecin/docteur | **être/suis** | médecin/docteur)	3
2	vous devez (re)faire le bandage (vous | **devoir/devez** | (re)faire | bandage)	2
3	êtes-vous allergique à des médicaments ? (vous/patient | être/êtes | allergie/allergique | **médicament** | ?)	4
4	devez-vous cracher quand vous toussez ? (**tousser** | causer/donner | cracher | ?)	1
5	avez-vous de la fièvre depuis deux jours ? (vous | fièvre/température | deux | **jour** | ?)	4
6	vous devez (re)faire le bandage (vous | **(re)faire_le_bandage**)	2
7	vous ne pouvez pas boire, pas manger, ni fumer (vous | **ok/pouvoir** | pas/non | boire | pas/non | manger | pas/non | fumer)	2
8	mangez-vous beaucoup de viande rouge ? (vous | **manger** | beaucoup/souvent | viande_rouge | ?)	2
9	avez-vous de la fièvre depuis deux jours ? (vous | fièvre/température | **depuis** | deux | jour | ?)	3
10	avez-vous déjà fait une crise cardiaque ? (**"passé"** | vous | déjà | faire | crise_cardiaque | ?)	1
11	devez-vous cracher quand vous toussez ? (vous | **devoir/devez** | cracher | quand | vous | tousser | ?)	2
12	êtes-vous allergique à des médicaments ? (médicament | causer/donner | **allergie/allergique/gratter/boutons** | ?)	3
13	nous allons faire des vaccins (nous/médecin | **faire_vaccin/vacciner**)	2
14	bonjour je suis le médecin (bonjour/salut | je/moi/médecin/docteur/je_suis | **médecin/docteur**)	3
15	avez-vous déjà fait une crise cardiaque ? (**déjà/une_fois** | vous | crise_cardiaque | ?)	1
16	mangez-vous beaucoup de viande rouge ? (vous | manger | beaucoup/souvent | viande | **rouge** | ?)	5
17	vous ne pouvez pas boire, pas manger, ni fumer (vous | **pas/non** | boire | ni/et/pas | manger | ni/et/pas | fumer)	2
18	je vais vous prescrire des gouttes (je | prescrire | **goutte**)	3
19	nous allons faire des vaccins (**"futur"** | nous/médecin | faire/allons | vaccin/piqûre)	1
20	je vais vous prescrire des gouttes (**"futur"** | je | prescrire | goutte)	1

1.3 Phrase en pictogrammes : exemple de « bonjour je suis le médecin » avec Text-to-Picto



1-3

1.4 Liste de contrôle : exemple

Liste de contrôle A

Info séance

Date de la séance : ...../...../2023 Heure (début) : ..... Heure (fin) : .....

Participant : P1

Instructions

- Pour chaque phrase sous la forme de pictogrammes, veuillez toujours 1) vérifier si le participant a compris le sens global de la phrase (oui : ✓ | non : ✖ | ne sait pas : -) en lui demandant ce que dit le médecin ; 2) vérifier s'il a compris le sens du pictogramme à évaluer (voir "Questions de compréhension/satisfaction" et "Questionnaire") ; 3) vérifier s'il a compris le sens des pictogrammes avec des personnes ("je/moi/nous"/etc.).
- Notez, s'il vous plaît, d'autres éléments problématiques signalés par le participant ou observés par vous-même en tant que facilitatrice de l'étude. Indiquez le numéro de ces éléments dans la phrase à évaluer et le type de problème rencontré (voir exemples). Toutes remarques additionnelles seront appréciées.

Phrase	Le sens de la phrase est-il correct ?	Le sens du pictogramme à évaluer est-il correct ?	Autre(s) élément(s) problématique(s) (n° des pictogrammes dans la phrase à évaluer)	Description du problème	Remarques additionnelles
1a-exemple	✓	✖	(les pictos) 2 et 4	compréhension du sens des pictogrammes 'avoir' et 'cause'	le participant ne les signale pas, mais quand je pose la question..., sa réponse est fausse
2b-exemple	✖	✓	(le picto) 3	confusion : le participant hésite entre deux sens pour le picto 3 ('manger' ou 'se brosser les dents')	le participant a signalé qu'il hésitait (ou j'ai vu qu'il a hésité)
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

2 Évaluation de deux traductions – Grille des participants

Liste des participants

N° Participant	Niveau	Âge (tranche)	Sexe	Handicap	Commentaires (handicap, autres façons de communiquer, etc.)	Déjà utilisé les pictogrammes et/ou images ? Si oui, fréquence ?	Séance	Consentement pour l'enregistrement	Consentement participation à l'étude
P0-exemple	A1	36-45	homme	dyslexie	pas de spécificité reconnue, difficultés visuelles, autres	oui, souvent	10 mai 2023	oui	oui
P1									
P2									
P3									
P4									
P5									
P6									
P7									
P8									
P9									

3 Évaluation de deux traductions – Guide du déroulement

Guide pour la facilitatrice de l'étude

1. Au début de la séance, veuillez, s'il vous plaît :	Fait ?
• Présenter le contexte de l'étude : dire que le participant est un patient et que vous êtes un médecin, qu'il va devoir dire s'il comprend la phrase que vous allez lui donner sous la forme d'images.	
• Présenter la phrase exemple au participant.	
• Demander au participant son consentement (oral et écrit) pour participer à l'étude, être enregistré (audio seulement) et permettre la conservation de ses données (à demander déjà lors du recrutement).	
• Noter l'heure de début dans le document "Liste de contrôle".	
• Démarrer le logiciel d'enregistrement audio.	

2. Pendant la séance, veuillez, s'il vous plaît :	Fait ?
• Parcourir le questionnaire sur une tablette avec le participant : présenter la phrase en pictogrammes, sans donner de pistes sur son sens (si besoin, les agrandir).	
• Poser des questions de compréhension préparées à l'avance après chaque phrase (voir document "Questions de compréhension/satisfaction") : 1) que veut dire cette phrase (sens global de la phrase) ; 2) que veut dire ce pictogramme-ci (document "Questionnaire") en le montrant avec le doigt ; 3) qui sont les personnes représentées.	
• Noter si les réponses aux questions sont correctes dans le document "Liste de contrôle".	
• Si nécessaire, poser d'autres questions que vous trouvez pertinentes pour vérifier la compréhension de certains pictogrammes et les noter dans la "Liste de contrôle".	
• Poser les questions de satisfaction.	

3. A la fin de la séance, veuillez, s'il vous plaît :	Fait ?
• Arrêter le logiciel d'enregistrement audio.	
• Sauvegarder le fichier audio, si possible, avec le nom « seance-JJMM » (par exemple « seance-1005 » si la séance a eu lieu le 10 mai).	
• Noter l'heure de finalisation de la séance dans le document "Liste de contrôle".	
• Compléter le document "Liste des participants".	
• Vérifier après chaque séance la qualité de l'enregistrement (voix des participants).	
• Sauvegarder/scanner les documents d'évaluation de la séance ("Liste de contrôle" et "Liste des participants").	
• Envoyer les fichiers ci-dessous à Mme Magali Norré (par e-mail ou WeTransfer si le fichier est trop lourd) : ○ Documents d'évaluation ("Liste de contrôle" et "Liste des participants") ○ Fichiers audio	

4 Évaluation de deux traductions – Instructions

Questions de compréhension/satisfaction

Instructions générales

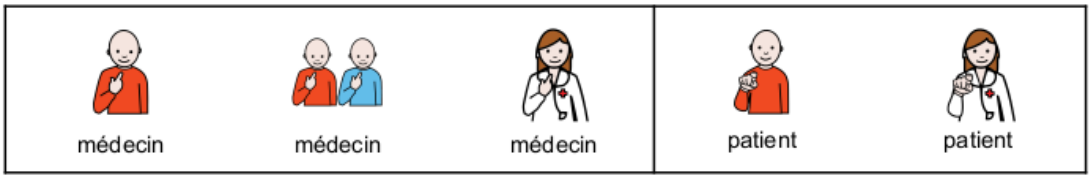
Au début de la séance, veuillez, s'il vous plaît, dire que le participant est un patient et que vous êtes un médecin qui doit lui poser quelques questions/instructions sur son état de santé, qu'il va devoir dire s'il comprend la phrase que vous allez lui demander sous la forme d'images affichées sur une tablette, une à la fois. Précisez que les phrases ne suivent pas un ordre logique. Veuillez également expliquer la tâche en montrant la phrase 0 à chacun des participants.

Pendant la séance, pour chaque phrase de chaque questionnaire (cf. document annexé), veuillez, s'il vous plaît, toujours vérifier :

1) la compréhension du sens global de la phrase : par exemple, avec "que veut dire cette phrase" ou "quelle(s) information(s) comprenez-vous "
2) la compréhension d'au moins un pictogramme prédéfini : par exemple, "qu'est-ce que ça / que signifie / veut dire cette image-ci" (en la montrant avec le doigt)

- Si le participant répond correctement à 1/2 (c'est-à-dire que sa réponse correspond au sens dans le document "Questionnaire", les synonymes et formulations équivalentes sont acceptées), marquez la cellule du document "Liste de contrôle" avec un ✓
- Si la réponse du participant est fausse, marquez la cellule avec un ✖
- Si le participant ne répond pas, laissez la cellule vide

3) la compréhension des pictogrammes avec des personnes 'je/moi//nous' qui représentent le médecin et 'vous' qui le représente lui en tant que patient, c'est-à-dire vérifier si le participant identifie bien qui est le locuteur de la phrase et à qui elle est destinée ; s'il ne comprend pas ces pictogrammes (ou hésite), indiquez le dans sa "Liste de contrôle"



Vous pouvez inviter le participant à :

- Demander d'agrandir un ou plusieurs pictogramme(s) en particulier (zoom)
- Signaler les pictogrammes qu'il ne comprend pas
- Expliquer avec ses mots la signification des pictogrammes (ex. : "hospitaliser" = rester à l'hôpital)

Avant la fin de la séance, veuillez, s'il vous plaît, poser les questions ci-dessous :

- Est-ce que vous rencontrez habituellement des difficultés à exprimer / comprendre des informations (médicales / de votre médecin) ?
- Aimez-vous utiliser ces (traductions en) pictogrammes ? Et aimeriez-vous continuer à utiliser ces pictogrammes lors d'une consultation médicale avec le médecin ou à l'hôpital ? Pourquoi ?
- Pensez-vous que c'est utile ? Et/ou que faut-il améliorer ?

5 Évaluation de deux traductions – Consentement

UNIVERSITÉ DE GENÈVE UCLouvain

Questionnaire Pictogrammes Santé

L'étude

L'étude vise à étudier la compréhension de **phrases médicales en pictogrammes**

Le questionnaire

Vous allez répondre à des **questions** sous la forme de pictogrammes

Vous allez expliquer **20 phrases**

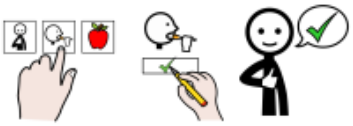
- si les phrases / pictogrammes sont faciles à comprendre et pourquoi ✓
- ou si les phrases / pictogrammes sont difficiles à comprendre et pourquoi ✗

Merci pour votre participation

L'étude est réalisée en collaboration avec l'Université de Genève et l'Université catholique de Louvain dans le cadre d'un travail (thèse) de Norré Magali

Le projet évalue des phrases en pictogrammes

- J'ai compris le but du projet
- J'accepte de répondre au questionnaire
- Je réponds à ce questionnaire volontairement



Les résultats sont anonymes

Votre nom ne sera jamais diffusé

- J'accepte d'être enregistré (audio)
- J'accepte que les résultats soient publiés dans des articles

Le questionnaire n'est pas obligatoire

- Je peux arrêter le questionnaire quand je veux
- Je ne dois pas m'expliquer
- J'ai lu / on m'a lu ces informations

Signature

ARASAAC

6 Évaluation de deux traductions – Retranscriptions

```
-e : Là... ?
-p : Là le médecin qui dit que sa patiente elle a des problèmes de yeux mais on sait pas quoi.
-e : Uh huh. Si on regarde les yeux peut-être on peut deviner qu'est-ce qu'elle a aux yeux ?
-p : Elle a les yeux rouges
-e : Uh huh, donc, peut-être le médecin demande quoi ?
-p : Euhhh, « pourquoi t'as les yeux rouges ? »
-e : Uh huh. Oui.
```

(a) Retranscription manuelle : PictoDr (P9, phrase 3)

```
-p : Et là peut-être... Là je viens de trouver pourquoi il... le médecin... a... ne rien trouvé
pourquoi il avait des yeux rouges, c'est de la peinture... comment il s'est mis la peinture
dans les yeux
-e : Ouais, donc le médecin pose la question...
-p : Ouais
-e : ... pour savoir si le patient...
-p : Comment il s'est mis euh... le
-e : Comment il s'est mis la peinture dans les yeux ?
-p : Ouais
-e : Uh huh. Ok.
```

(b) Retranscription manuelle : Text-to-Picto (P9, phrase 18)

```
25
00:03:09,000 --> 00:03:18,000
Là, le médecin dit que sa patiente a des problèmes des yeux, mais on ne sait pas quoi.

26
00:03:18,000 --> 00:03:22,000
Si on regarde les yeux, peut-être on peut deviner qu'est-ce qu'elle a aux yeux?

27
00:03:22,000 --> 00:03:27,000
Elle a les yeux rouges. Donc peut-être que le médecin demande quoi?

28
00:03:27,000 --> 00:03:35,000
Pour toi, tu as les yeux rouges. Parfait. Et ici? Ah, les cigarettes.
```

(c) Retranscription automatique : PictoDr (P9, phrase 3)

```
173
00:15:33,000 --> 00:15:45,000
Là, peut-être, j'ai trouvé pourquoi le médecin ou le client a trouvé pourquoi il avait les yeux rouges
à cause de la peinture.

174
00:15:45,000 --> 00:15:51,000
La question est comment il s'est mis la peinture dans les yeux.

175
00:15:51,000 --> 00:15:52,000
Ouais.

176
00:15:52,000 --> 00:15:56,000
Donc, le médecin pose la question.

177
00:15:56,000 --> 00:15:57,000
Ouais.

178
00:15:57,000 --> 00:16:03,000
Savoir si le patient a comment il s'est mis la peinture dans les yeux.

179
00:16:03,000 --> 00:16:04,000
Ouais.
```

(d) Retranscription automatique : Text-to-Picto (P9, phrase 18)

FIGURE B.1 – Retranscriptions manuelles et automatiques : exemple de la phrase « avez-vous les yeux rouges? »

7 Évaluation d'une anamnèse – Formulaire COVID-19

Hôpitaux Universitaires Genève	Direction des soins / Département de Médecine Communautaire		
Formulaire pour la prise en soin d'un patient stable suspect d'une infection à SARS-CoV-2			
Responsable du document : F. Jacquerioz, T. Delieutraz Validé par J. Salamun, E. Satin, H. Spechbach	***Type de document :*** Protocole	***Domaine :*** Soins infirmiers, soins médico-délégués	
Date de création : 27.02.2020	***Date d'entrée en vigueur :*** ***28.02.2020***	***Date de mise à jour :*** 02.04.2020	***N° de version :*** 1.9

COLLABORATEUR HUG :
oui ☐ non ☐
service : _____

Si oui en contact avec patients :
oui ☐ non ☐

Soignants hors HUG : oui ☐ non ☐
Lieu : _____

Etiquette patient

NOM

PRENOM

DATE DE NAISSANCE

Tél

DEPISTAGE AVANT HOSPITALISATION : **oui ☐ non ☐** Service : _____
DEPISTAGE ASYMPTOMATIQUE : **oui ☐ non ☐** Raison : _____

Profession : _____
Employeur : _____
Lieu : _____

Date :

Langue parlée : français albanais anglais arabe espagnol farsi dari portugais roumain tigrinien autre :_____

Date de début des symptômes :

A. SYMPTÔMES :

Rhinorrhée/*Runny nose* **Odynodysphagie**/*sore throat* **Myalgies**/*muscular pain* **Frissons**/*chills*

Toux sèche/*dry cough* **Toux productive**/*productive cough* **Crachats rosés**/*pink sputum*

Fièvre (anamnestique ou objectivée)/*Fever* **Anosmie** (perte d'odorat)/*anosmia (loss of smell)*

Symptômes digestifs (nausées/diarrhées/douleurs abdo)/*abdominal symptoms* **asthénie** /*fatigue*

Dyspnée (peine à respirer)/*difficulty breathing* **Douleurs thoraciques**/ *thoracic pain*

céphalées/*headache* Autre : _____

Allergies :

Traitements habituels : _____

B. PARAMETRES VITAUX / SIGNES CLINIQUES :

Dyspnée (stridor, SDR)	oui ☐		non ☐
Tachypnée (> 24/min)	oui ☐	……../min	non ☐
Température (axillaire > 38,5°C)	oui ☐	……. °C	non ☐
Désaturation (< 94% AA)	oui ☐	……. %	non ☐

Si OUI à l'un de ces trois symptômes : contacter le médecin pour suite de prise en charge

Hôpitaux Universitaires Genève	Direction des soins / Département de Médecine Communautaire		
Formulaire pour la prise en soin d'un patient stable suspect d'une infection à SARS-CoV-2			
Responsable du document : F. Jacquerioz, T. Delieutraz Validé par J. Salamun, E. Satin, H. Spechbach	***Type de document :*** Protocole	***Domaine :*** Soins infirmiers, soins médico-délégués	
Date de création : 27.02.2020	***Date d'entrée en vigueur :*** ***28.02.2020***	***Date de mise à jour :*** 02.04.2020	***N° de version :*** 1.9

C. CRITÈRES DE VULNÉRABILITÉ/facteurs de risque :

Maladie respiratoire chronique :	oui ☐	non ☐
Cardiopathie :	oui ☐	non ☐
ATCD d'infarctus :	oui ☐	non ☐
Immunosuppression :	oui ☐	non ☐
Greffé :	oui ☐	non ☐
Chimiothérapie récente :	oui ☐	non ☐
Cancer :	oui ☐	non ☐
Diabète :	oui ☐	non ☐
Grossesse:	oui ☐	non ☐
Patient prenant plus de 3 traitements médicamenteux :	oui ☐	non ☐
Tabagisme (si tabac actif, UPA : _____)	oui ☐	non ☐

D. HABITUDES DE VIE :

Comorbidité : HTA	oui ☐	non ☐
Prise récente (7 derniers jours) d'AINS (eg ibuprofen, diclofenac) :	oui ☐	non ☐
Vaccin grippe (saison 2019-2020) :	oui ☐	non ☐
Patient SDF :	oui ☐	non ☐
Foyer (migrants/réfugiés) :	oui ☐	non ☐
Résident EMS :	oui ☐	non ☐
Environnement psycho-social difficile :	oui ☐	non ☐

Autre structure de vie collective : _____

Nombre de personnes vivant dans le même appartement/foyer: _____

E. EXPOSITION (lien épidémiologique) :

1. **Avez-vous voyagé dans les 14 derniers jours ?** **oui ☐ non ☐**
   Have you travelled in the last 14 days?
   Si oui :
   Pays :_____, ville : _____Nombre de jour : _____
   Pays :_____, ville : _____Nombre de jour : _____

   Quelles activités sociales ou professionnelles avez-vous faites dans ces pays ? :

2. **Avez-vous été en contact avec une personne confirmée positive au SARS-CoV-2 dans les 14 derniers jours ?**
   Have you been in close contact with a SARS-CoV-2 confirmed patient in the last 14 days?
   oui ☐ non ☐
   Lieu : Date :

3. **Etiez-vous en quarantaine en tant que contact proche d'un cas confirmé SARS-CoV-2 ?**
   oui ☐ non ☐

FROTTIS : oui ☐ non ☐

Veuillez préciser si frottis non effectué : _____

Pour le patient : besoin d'une ordonnance ou d'un arrêt de travail ? **Oui ☐ non ☐**

Si oui : contacter le médecin

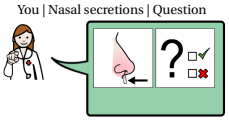
Intervention du médecin (statut, complément d'anamnèse) :

8 Évaluation d'une anamnèse – Phrases

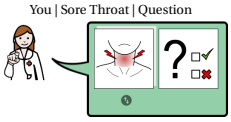
8.1 Toutes les phrases en pictogrammes

PictoDr

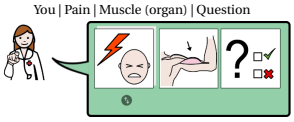
1A) AVEZ-VOUS LE NEZ QUI COULE?
You | Nasal secretions | Question



2A) AVEZ-VOUS MAL À LA GORGE?
You | Sore Throat | Question



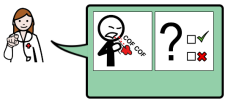
3A) AVEZ-VOUS DES DOULEURS MUSCULAIRES?
You | Pain | Muscle (organ) | Question



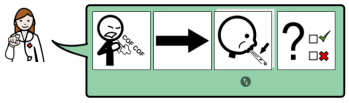
4A) AVEZ-VOUS DES FRISSONS?
You | Chills | Question

5A) AVEZ-VOUS UNE TOUX GRASSE?
You | Productive Cough | Question

6A) EST-CE SANGLANT QUAND VOUS TOUSSEZ?
You | Hemoptysis | Question



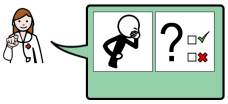
7A) DEVEZ-VOUS CRACHER QUAND VOUS TOUSSEZ?
You | Coughing | Causes | Spitting | Question



8A) AVEZ-VOUS DE LA FIÈVRE?
You | Fever | Question

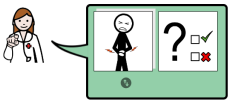
9A) AVEZ-VOUS DES TROUBLES DE L'ODORAT?
You | Dysosmia | Question

10A) AVEZ-VOUS DES NAUSÉES?
You | Nausea | Question



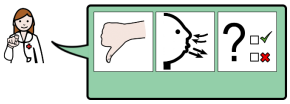
11A) AVEZ-VOUS LA DIARRHÉE?
You | Diarrhea | Question

12A) AVEZ-VOUS MAL AU VENTRE?
You | Abdominal Pain | Question

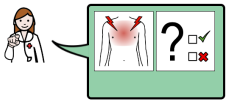


13A) ÊTES-VOUS FATIGUÉ?
You | Feeling tired | Question

14A) AVEZ-VOUS DES DIFFICULTÉS À RESPIRER?
You | Has difficulty doing (qualifier value) | Respiration | Question



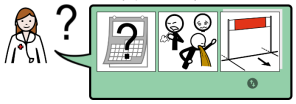
15A) AVEZ-VOUS MAL À LA POITRINE?
You | Chest Pain | Question



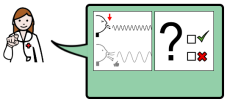
16A) AVEZ-VOUS MAL À LA TÊTE?
You | Headache | Question

17A) AVEZ-VOUS DES ALLERGIES (CONNUES)?
You | Known allergies or adverse drug reactions | Question

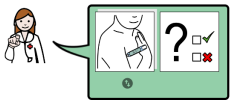
18A) À QUELLE DATE ONT COMMENCÉ LES SYMPTÔMES?
Date in time | Symptoms | Started



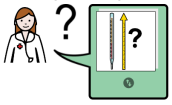
19B) RESPIREZ-VOUS TROP RAPIDEMENT?
You | Tachypnea |
Question



20B) AVEZ-VOUS PRIS VOTRE TEMPÉRATURE?
You | Past history of | Temperature taking |
Question

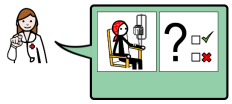


21B) À COMBIEN ÉTAIT VOTRE TEMPÉRATURE LA DERNIÈRE FOIS QUE VOUS L'AVEZ MESURÉE?
Degree findings

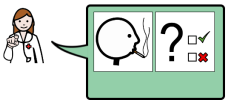


22C) AVEZ-VOUS EU UN ACCIDENT VASCULAIRE CÉRÉBRAL?
You | Past history of |
Cerebral vascular accident (CVA) / Stroke | Question

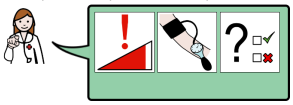
23C) AVEZ-VOUS EU UNE CHIMIOTHÉRAPIE?
You | Past history of | Chemotherapy |
Question



24C) FUMEZ-VOUS?
You | Smoking |
Question

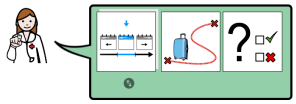


25D) AVEZ-VOUS TROP DE TENSION ARTÉRIELLE?
You | Too much | Blood Pressure | Question

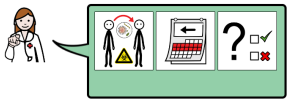


26D) AVEZ-VOUS ÉTÉ VACCINÉ CONTRE LE COVID?
You | Vaccinated (qualifier) |
COVID19 (disease) | Question

27E) AVEZ-VOUS VOYAGÉ RÉCEMMENT?
You | Recent | Travel |
Question

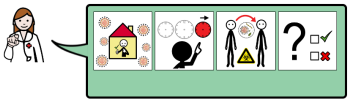


28E) AVEZ-VOUS ÉTÉ EN CONTACT AVEC UNE PERSONNE CONFIRMÉE POSITIVE AU CORONAVIRUS DANS LES QUATORZE DERNIERS JOURS?
You | Contact with and (suspected) exposure to COVID-19 | Past 2 Weeks | Question



29E) À QUELLE DATE AVEZ-VOUS ÉTÉ EN CONTACT AVEC QUELQU'UN DE POSITIF AU CORONAVIRUS?
Date in time | You | Contact with and (suspected) exposure to COVID-19 | COVID19 (disease)

30E) ÉTIEZ-VOUS EN QUARANTAINE EN TANT QUE CONTACT PROCHE D'UN CAS CONFIRMÉ DE CORONAVIRUS?
You | Quarantine | Post | Contact with and (suspected) exposure to COVID-19 | Question



A = Symptômes | B = Paramètres vitaux | C = Critères de vulnérabilité | D = Habitudes de vie | E = Exposition || Phrases sources suivies des concepts anglais d'UMLS (+ « You » | « Question »)

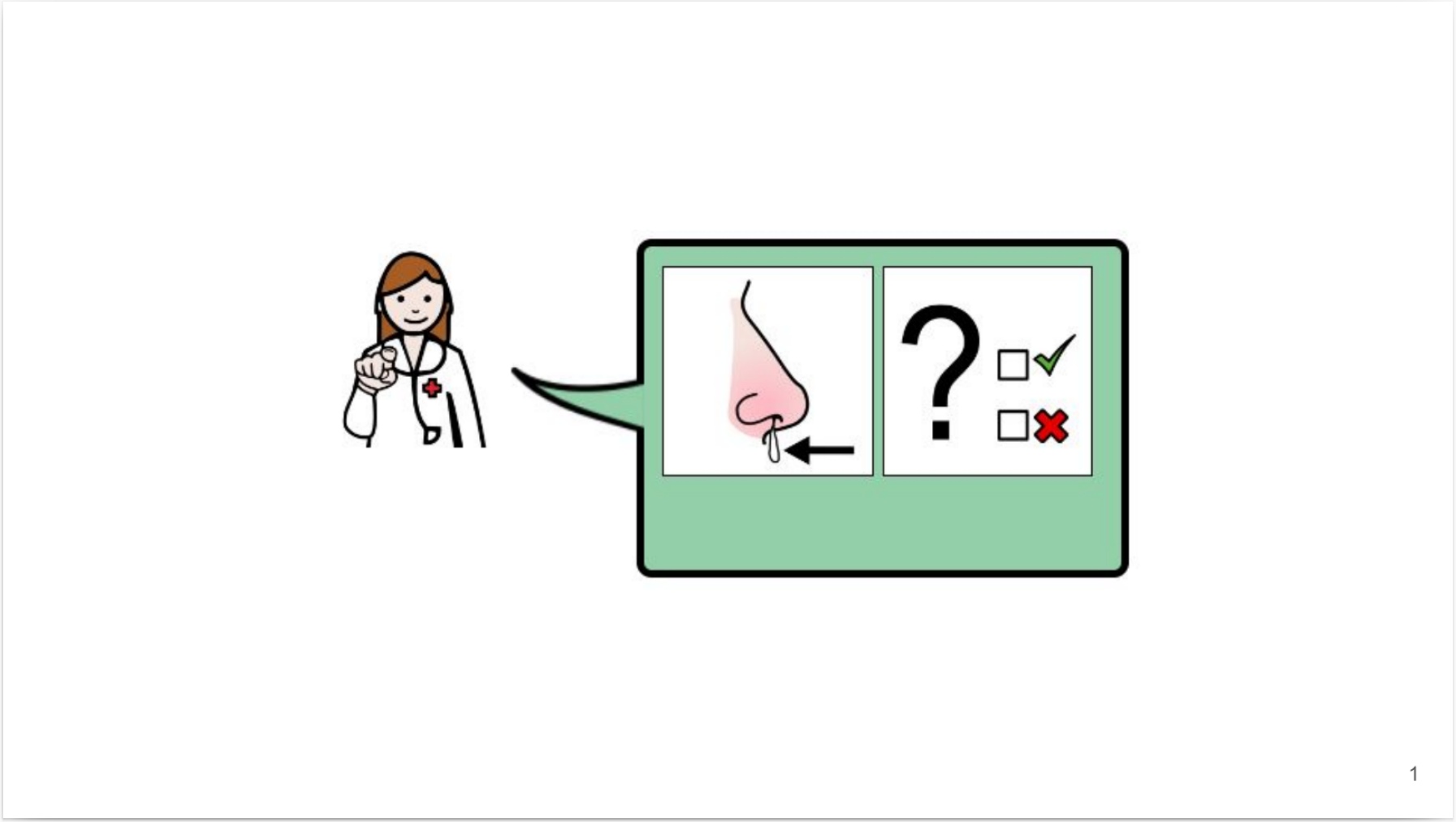
TABLE B.3 – Questionnaire : traductions de PictoDr, anamnèse du COVID-19

8.2 Liste de phrases et de pictogrammes

Questionnaire

N° Phrase	Sens de la phrase évaluée (sens des pictogrammes individuels)
0-ex.	bonjour je suis le médecin (bonjour/salut | je/moi suis | médecin/docteur)
1	avez-vous le nez qui coule ? (nez qui coule | ?)
2	avez-vous mal à la gorge ? (mal/douleur à la gorge | ?)
3	avez-vous des douleurs musculaires ? (mal/douleur | muscle | ?)
4	avez-vous des frissons ? (frisson | ?)
5	avez-vous une toux grasse ? (toux grasse | ?)
6	est-ce sanglant quand vous toussez ? (sang quand vous toussez | ?)
7	devez-vous cracher quand vous toussez ? (tousser/toussant | causer/en | cracher/crachant | ?)
8	avez-vous de la fièvre ? (fièvre | ?)
9	avez-vous des troubles de l'odorat ? (troubles de l'odorat | ?)
10	avez-vous des nausées ? (nausées | ?)
11	avez-vous la diarrhée ? (diarrhée | ?)
12	avez-vous mal au ventre ? (mal/douleur au ventre | ?)
13	êtes-vous fatigué ? (fatigué | ?)
14	avez-vous des difficultés à respirer ? (difficulté/difficile | respirer | ?)
15	avez-vous mal à la poitrine ? (mal/douleur à la poitrine | ?)
16	avez-vous mal à la tête ? (mal/douleur à la tête | ?)
17	avez-vous des allergies ? (allergies | ?)
18	à quelle date ont commencé les symptômes ? (à quelle date | symptômes | commencer)
19	respirez-vous trop rapidement ? (respirer rapidement | ?)
20	avez-vous pris votre température ? (prendre température | ?)
21	à combien était votre température la dernière fois que vous l'avez mesurée ? (température combien)
22	avez-vous eu un accident vasculaire cérébral ? (accident vasculaire cérébral | ?)
23	avez-vous eu une chimiothérapie ? (chimiothérapie | ?)
24	fumez-vous ? (fumer | ?)
25	avez-vous trop de tension artérielle ? (trop/beaucoup | tension artérielle | ?)
26	avez-vous été vacciné contre le covid ? (vacciné | covid/coronavirus | ?)
27	avez-vous voyagé récemment ? (récemment | voyager | ?)
28	avez-vous été en contact avec une personne confirmée positive au coronavirus dans les 14 derniers jours ? (être en contact avec une personne qui a le covid | 14 derniers jours | ?)
29	à quelle date avez-vous été en contact avec quelqu'un de positif au coronavirus ? (à quelle date | être en contact avec une personne qui a le covid | covid/coronavirus)
30	étiez-vous en quarantaine en tant que contact proche d'un cas confirmé de coronavirus ? (être en quarantaine | vous avez été dans le passé | être en contact avec une personne qui a la covid | ?)

8.3 Phrase en pictogrammes : exemple de « avez-vous le nez qui coule ? » avec PictoDr



9 Évaluation d'une anamnèse – Guide du déroulement

Guide pour la facilitatrice de l'étude

1. Au début de la séance, veuillez, s'il vous plaît :	Fait ?
• Présenter le contexte de l'étude : dire que le participant est le patient et que vous êtes le médecin, qu'il vient pour savoir s'il a le Covid-19, qu'il va devoir expliquer ce qu'il comprend de la phrase que vous allez lui donner sous la forme d'images.	
• Présenter la phrase exemple au participant.	
• Demander au participant son consentement (oral et écrit) pour participer à l'étude, être enregistré (audio seulement) et permettre la conservation de ses données (à demander déjà lors du recrutement).	
• Démarrer l'enregistrement audio.	

2. Pendant la séance, veuillez, s'il vous plaît :	Fait ?
• Poser la question sur ce qu'est le Covid-19.	
• Présenter la phrase en pictogrammes en faisant le lien entre lui/elle, vous et les pictogrammes (1. en vous désignant, 2. puis en montrant le pictogramme "médecin", 3. les pictogrammes dans la bulle, 4. en désignant le participant). Si besoin, vous pouvez agrandir les pictogrammes.	
• Poser les questions de compréhension préparées à l'avance après chaque phrase (voir document "Questions de compréhension/satisfaction") : 1) qu'est-ce que je veux savoir en tant que médecin (sens global de la phrase) ; 2) s'il ne comprend pas le sens (voir document "Questionnaire"), est-ce que le pictogramme aide à comprendre le mot et comment il le comprend avec le pictogramme.	
• Noter si les réponses aux questions sont correctes dans le document "Liste de contrôle".	
• Si nécessaire, poser d'autres questions que vous trouvez pertinentes pour vérifier la compréhension de certains pictogrammes et les noter dans la "Liste de contrôle".	
• Poser les questions de satisfaction.	

3. A la fin de la séance, veuillez, s'il vous plaît :	Fait ?
• Arrêter l'enregistrement audio.	
• Sauvegarder le fichier audio, si possible, avec le nom « seance-JJMMAAAA » (par exemple « seance-10052024 » si la séance a eu lieu le 10 mai 2024).	
• Compléter le document "Liste des participants".	
• Vérifier après chaque séance la qualité de l'enregistrement (voix des participants).	
• Scanner les documents d'évaluation de la séance ("Liste de contrôle" et "Liste des participants").	
• Envoyer les fichiers ci-dessous à Mme Magali Norré : ○ Documents d'évaluation ("Liste de contrôle" et "Liste des participants") ○ Formulaires de consentement ○ Fichiers audio	

10 Évaluation d'une anamnèse – Instructions

Questions de compréhension/satisfaction

Instructions générales

Au début de la séance, veuillez, s'il vous plaît, dire que le participant est un patient et que vous êtes un médecin qui doit lui poser quelques questions sur son état de santé afin de savoir s'il a le Covid-19 (anamnèse), qu'il va devoir expliquer ce qu'il comprend de la phrase que vous allez lui demander sous la forme d'images affichées sur un ordinateur portable, une à la fois.

Veuillez expliquer la tâche en montrant la phrase 0 à chacun des participants. Vous pouvez lancer l'enregistrement après.

Pendant la séance, veuillez d'abord lui demander une fois :

- Savez-vous ce que c'est la maladie du Covid-19, pouvez-vous la décrire ?

Pour chaque phrase en pictogramme, veuillez, s'il vous plaît, toujours :

1) dire "je suis le médecin" en vous montrant du doigt, puis en montrant le pictogramme ci-dessous :

dire "vous êtes le patient", le désigner du doigt et montrer du doigt la bulle verte.

2) vérifier la compréhension du sens global de la phrase : avec "qu'est-ce que je veux savoir en tant que médecin ? (= que veut dire cette phrase ?)"

- Si le participant répond correctement (c'est-à-dire que sa réponse correspond au sens dans le document "Questionnaire" ; les synonymes et formulations équivalentes sont acceptées), marquez la cellule du document "Liste de contrôle" avec un ✓
- Si la réponse du participant est fausse, s'il ne répond pas, qu'il ne sait pas ou que vous n'êtes pas certaine de sa réponse, marquez la cellule avec un ✖

  -> Donner la bonne réponse et poser les deux questions suivantes :
  - Est-ce que le(s) pictogramme(s) aide(nt) à comprendre le mot/la phrase ? ✓ oui | ✖ non
  - Comment comprenez-vous le mot/la phrase avec le(s) pictogramme(s) ?

Vous pouvez inviter le participant à :
- Demander d'agrandir un ou plusieurs pictogramme(s) en particulier (zoom)
- Signaler les pictogrammes qu'il ne comprend pas
- Expliquer avec ses mots la signification des pictogrammes (ex. : "hospitaliser" = rester à l'hôpital)

Avant la fin de la séance, veuillez, s'il vous plaît, poser les questions de satisfaction ci-dessous :

- Aimez-vous utiliser ces (traductions en) pictogrammes ? Et aimeriez-vous continuer à utiliser ces pictogrammes lors d'une consultation médicale avec le médecin ou à l'hôpital ? Pourquoi ? Pensez-vous que c'est utile ?
- Et que faut-il améliorer ? Qu'est-ce qui était (plus) difficile à comprendre ?

Annexe C

Code source pour l'analyse des données

Nous donnons le code R pour la construction des meilleurs modèles linéaires généralisés à effets mixtes utilisés pour la prédiction de la compréhension des pictogrammes et des phrases de l'évaluation de deux traductions (cf. chapitre 6, section 1.6.2). Il inclut les dépendances et les commandes des analyses statistiques.

```
library(readr, dplyr, psych) #readr==2.1.4 | dplyr==1.1.4 | psych==2.3.9
library(glmmTMB) #glmmTMB==1.1.8
library(DHARMa) #DHARMa==0.4.6
library(performance) #performance==0.10.8
library(car, caret) #car==3.1-2 | caret==6.0-94
df <- read_tsv(file = "results.tsv") #773 rows/pictographs
df_picto <- subset(df, select = c(annotator1_picto_all, A1_PictoParPhrase,
    translation, V2, V3, V4, V5, V6, V7, V8, sentence_type, pos, level,
    participant))
df_picto <- na.omit(df_picto) #683 rows/pictographs
df_sentence <- df %>% distinct(n_sentence, .keep_all = T) #180 sentences
res_picto <- glmmTMB(annotator1_picto_all ~ translation + V2 + V3 + V4 + V5 + V6
     + V7 + V8 + sentence_type + pos + level + (1|participant), data = df_picto,
    family = binomial(link="logit"))
res_sentence <- glmmTMB(annotator1_sentence ~ A1_PictoParPhrase + level + (1|
    participant), data = df_sentence, family = binomial(link="logit"))
#or res_sentence
plot(DHARMa::simulateResiduals(res_picto))
summary(res_picto)
confint(res_picto)
exp(confint(res_picto))
car::Anova(res_picto)
performance::check_collinearity(res_picto)
#or res_sentence / df_sentence$annotator1_sentence
predictions <- predict(res_picto, type = "response")
predictions <- ifelse(predictions>.5,1,0)
confusionMatrix(table(predictions, df_picto$annotator1_picto_all))
prop.table(table(predictions, df_picto$annotator1_picto_all))*100
```

Bibliographie

(Abeillé *et al.*, 2003) Anne ABEILLÉ, Lionel CLÉMENT et François TOUSSENEL, Building a Treebank for French. *In* Anne ABEILLÉ, éditeur, *Treebanks : Building and Using Parsed Corpora*, Text, Speech and Language Technology, pages 165–187. Springer, Dordrecht, 2003.

(Abeillé et Crabbé, 2013) Anne ABEILLÉ et Benoit CRABBÉ, Vers un treebank du français parlé. *In Proceedings of TALN 2013*, pages 174–187, 2013.

(Akaike, 1973) Hirotogu AKAIKE, Information Theory and an Extension of the Maximum Likelihood Principle. *In Second International Symposium on Information Theory*, pages 267–281, 1973.

(Al-Thanyyan et Azmi, 2021) Suha S. AL-THANYYAN et Aqil M. AZMI, Automated Text Simplification : A Survey. *ACM Computing Surveys (CSUR)*, 54(2):1–36, 2021.

(Alfons, 2024) Andreas ALFONS, *cvTools : Cross-Validation Tools for Regression Models*, 2024. R package version 0.3.3, https://cran.r-project.org/web/packages/cvTools/cvTools.pdf.

(Altschuler *et al.*, 2021) Tami ALTSCHULER, Rachel SANTIAGO et Jessica GORMLEY, Ensuring communication access for all during the COVID-19 pandemic and beyond : supporting patients, providers, and caregivers in hospitals. *Augmentative and Alternative Communication*, pages 1–13, 2021.

(Alvarez, 2011) Juliana ALVAREZ, Évaluation des systèmes symboliques des symptômes biomédicaux pour la création d'une interface graphique pluriculturelle de l'échelle de triage canadien. Mémoire de master, Université de Montréal, 2011.

(Alvarez, 2014) Juliana ALVAREZ, Visual design. A step towards multicultural health care. *Arch Argent Pediatr*, 112(1):33–40, 2014.

(Anand *et al.*, 2024) Bhavana ANAND, Themozhi J., Shreyas Sai R., Charumathi P. et Mirnalinee T. T., SSN-MLRG at Text to Picto 2024 : A BERT-Based Approach for Mapping French Sentences to Pictogram Terms. *In Working Notes of the Conference and*

Labs of the Evaluation Forum (CLEF 2024), volume 3740, pages 1474–1479, Grenoble, France, 2024.

(Antoine *et al.*, 2014) Jean-Yves ANTOINE, Anaïs LEFEUVRE-HALFTERMEYER et Willy ALLEGRE, Pour une réflexion éthique sur les conséquences de l'usage des NTIC : le cas des aides techniques (à composante langagière ou non) aux personnes handicapées. *Actes de la journée ATALA Éthique et TAL*, 2014.

(Antoine et Maurel, 2007) Jean-Yves ANTOINE et Denis MAUREL, Aide à la communication pour personnes handicapées et prédiction de texte : Problèmatique, état des lieux et retour sur trente ans de recherche en communication augmentée. *Traitement Automatique des Langues*, 48(2):9–46, 2007.

(Baker, 1982) Bruce BAKER, Minspeak : A semantic compaction system that makes self-expression easier for communicatively disabled individuals. *Byte*, 7(9):186–202, 1982.

(Banerjee et Lavie, 2005) Satanjeev BANERJEE et Alon LAVIE, METEOR : An automatic metric for MT evaluation with improved correlation with human judgments. *In Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72, 2005.

(Barbu *et al.*, 2013) Eduard BARBU, Maria Teresa MARTÍN-VALDIVIA et Luis Alfonso UREÑA-LÓPEZ, Open Book : a tool for helping ASD users' semantic comprehension. *In Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility*, pages 11–19, Atlanta, Georgia, 2013. Association for Computational Linguistics.

(Barlacchi et Tonelli, 2013) Gianni BARLACCHI et Sara TONELLI, ERNESTA : A Sentence Simplification Tool for Children's Stories in Italian. *In Computational Linguistics and Intelligent Text Processing*, pages 476–487. Springer, 2013.

(Baroni *et al.*, 2009) Marco BARONI, Silvia BERNARDINI, Adriano FERRARESI et Eros ZANCHETTA, The Wacky Wide Web : A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation*, 43(3):209–226, 2009.

(Barros *et al.*, 2014) Izadora M.C. BARROS, Thaciana S. ALCÂNTARA, Alessandra R. MESQUITA, Anne Caroline O. SANTOS, Felipe P. PAIXÃO et Divaldo P. LYRA JR., The use of pictograms in the health care : A literature review. *Research in Social and Administrative Pharmacy*, 10(5):704–719, 2014.

(Bates *et al.*, 2015) Douglas BATES, Martin MÄCHLER, Ben BOLKER et Steve WALKER, Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

(Bautista *et al.*, 2017) Susana BAUTISTA, Raquel HERVÁS, Agustín HERNÁNDEZ-GIL, Carlos MARTÍNEZ-DÍAZ, Sergio PASCUA et Pablo GERVÁS, AraTraductor : Text to Pictogram Translation using Natural Language Processing Techniques. *In Proceedings of the XVIII International Conference on Human Computer Interaction*, pages 1–8, 2017.

(Beail et Williams, 2014) Nigel BEAIL et Katie WILLIAMS, Using Qualitative Methods in Research with People Who Have Intellectual Disabilities. *Journal of Applied Research in Intellectual Disabilities*, 27(2):85–96, 2014.

(Beukelman et Mirenda, 2017) David R. BEUKELMAN et Pat MIRENDA, *Communication alternative et améliorée : Aider les enfants et les adultes avec des difficultés de communication*. De Boeck Superieur, 2017.

(Billami *et al.*, 2018) Mokhtar BILLAMI, Thomas FRANÇOIS et Núria GALA, ReSyf : A French lexicon with ranked synonyms. *In Proceedings of the 27th International Conference on Computational Linguistics*, pages 2570–2581, Santa Fe, New Mexico, USA, 2018. Association for Computational Linguistics.

(Biotteau *et al.*, 2013) Christian BIOTTEAU, Jean CARON, Jérôme COLONNA, Franck HAYART, Coryne HUSSE, Michel JAUSSERAN, Marie-Françoise JUNKE, Pierre LAGIER, Florence LOMBARD, Muriel POHER et Dominique PRÉVOT, Pour une santé accessible aux personnes handicapées mentales, 2013. Livre blanc de l'Unapei, https://www.unapei.org/wp-content/uploads/2019/10/LIVRE-BLANC-SANTE-HD.pdf.

(Bliss, 1965) Charles Kasiel BLISS, *Semantography*. Australia : Semantography Publications, 1965.

(Blissymbolics Communication International, 2004) BLISSYMBOLICS COMMUNICATION INTERNATIONAL, The fundamental rules of Blissymbolics : creating new Blissymbolics characters and vocabulary, 2004. https://www.blissymbolics.org/images/bliss-rules.pdf.

(Bloomberg *et al.*, 1990) Karen BLOOMBERG, George R. KARLAN et Lyle L. LLOYD, The Comparative Translucency of Initial Lexical Items Represented in Five Graphic Symbol Systems and Sets. *Journal of Speech, Language, and Hearing Research*, 33(4):717–725, 1990.

(Bodenreider, 2004) Olivier BODENREIDER, The Unified Medical Language System (UMLS) : integrating biomedical terminology. *Nucleic Acids Research*, 32:D267–D270, 2004.

(Bojanowski *et al.*, 2017) Piotr BOJANOWSKI, Edouard GRAVE, Armand JOULIN et Tomas MIKOLOV, Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017.

(Bondy et Frost, 1994) Andy BONDY et Lori FROST, The Picture Exchange Communication System. *Focus on Autistic Behavior*, 9(3):1–19, 1994.

(Bouillon *et al.*, 2021) Pierrette BOUILLON, Johanna GERLACH, Jonathan MUTAL, Nikolaos TSOURAKIS et Hervé SPECHBACH, A Speech-enabled Fixed-phrase Translator for Healthcare Accessibility. *In Proceedings of the 1st Workshop on NLP for Positive Impact*, pages 135–142. Association for Computational Linguistics, 2021.

(Bouillon *et al.*, 2023) Pierrette BOUILLON, Johanna GERLACH, Magali NORRÉ et Hervé SPECHBACH, Enhancing Speech Translation in Medical Emergencies with Pictographs : BabelDr. *In* Meng JI, Pierrette BOUILLON et Mark SELIGMAN, éditeurs, *Translation Technology in Accessible Health Communication*, Studies in Natural Language Processing, pages 129–151. Cambridge University Press, Cambridge, 2023.

(Bouillon *et al.*, 2017) Pierrette BOUILLON, Johanna GERLACH, Hervé SPECHBACH, Nikolaos TSOURAKIS et Sonia HALIMI, BabelDr vs Google Translate : A user study at Geneva University Hospitals (HUG). *In Proceedings of the 20th Annual Conference of the European Association for Machine Translation*, pages 47–52, Prague, Czech Republic, 2017.

(Bregeon *et al.*, 2019) Dany BREGEON, Jean-Yves ANTOINE, Jeanne VILLANEAU et Anaïs LEFEUVRE-HALFTERMEYER, Redonner du sens à l'accord interannotateurs : vers une interprétation des mesures d'accord en termes de reproductibilité de l'annotation. *Traitement Automatique des Langues*, 60(2):23, 2019.

(Brooks *et al.*, 2017) Mollie E. BROOKS, Kasper KRISTENSEN, Koen J. VAN BENTHEM, Arni MAGNUSSON, Casper W. BERG, Anders NIELSEN, Hans J. SKAUG, Martin MÄCHLER et Benjamin M. BOLKER, glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9(2):378–400, 2017.

(Brouwers *et al.*, 2014) Laetitia BROUWERS, Delphine BERNHARD, Anne-Laure LIGOZAT et Thomas FRANÇOIS, Syntactic Sentence Simplification for French. *In Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 47–56, Gothenburg, Sweden, 2014. Association for Computational Linguistics.

(Brown *et al.*, 2020) Tom BROWN, Benjamin MANN, Nick RYDER, Melanie SUBBIAH, Jared KAPLAN, Prafulla DHARIWAL, Arvind NEELAKANTAN, Pranav SHYAM, Girish SASTRY, Amanda ASKELL, Sandhini AGARWAL, Ariel HERBERT-VOSS, Gretchen KRUEGER, Tom HENIGHAN, Rewon CHILD, Aditya RAMESH, Daniel M. ZIEGLER, Jeffrey WU, Clemens WINTER, Christopher HESSE, Mark CHEN, Eric SIGLER, Mateusz LITWIN, Scott GRAY,

Benjamin CHESS, Jack CLARK, Christopher BERNER, Sam MCCANDLISH, Alec RADFORD, Ilya SUTSKEVER et Dario AMODEI, Language Models are Few-Shot Learners. *In Proceedings of the 34th International Conference on Neural Information Processing Systems*, volume 33, pages 1877–1901, 2020.

(Bui *et al.*, 2012) Duy Duc An BUI, Carlos NAKAMURA, Bruce E. BRAY et Qing ZENG-TREITLER, Automated Illustration of Patients Instructions. *In AMIA Annual Symposium Proceedings*, volume 2012, pages 1158–1167. American Medical Informatics Association, 2012.

(Bulté *et al.*, 2018) Bram BULTÉ, Leen SEVENS et Vincent VANDEGHINSTE, Automating lexical simplification in Dutch. *Computational Linguistics in the Netherlands Journal*, 8:24–48, 2018.

(Bulté *et al.*, 2021) Bram BULTÉ, Vincent VANDEGHINSTE, Leen SEVENS, Ineke SCHUURMAN et Frank VAN EYNDE, Can Pictograph Translation Technologies Facilitate Communication and Integration in Migration Settings? *Computational Linguistics in the Netherlands Journal*, 11:189–212, 2021.

(Burchardt, 2013) Aljoscha BURCHARDT, Multidimensional Quality Metrics : A Flexible System for Assessing Translation Quality. *In Proceedings of Translating and the Computer 35*, London, UK, 2013. Aslib.

(Cabello *et al.*, 2018) Laura CABELLO, Eduardo LLEIDA, Javier SIMÓN, Antonio MIGUEL et Alfonso ORTEGA, Text-to-Pictogram Summarization for Augmentative and Alternative Communication. *Procesamiento del Lenguaje Natural*, 61:15–22, 2018.

(Callison-Burch *et al.*, 2006) Chris CALLISON-BURCH, Miles OSBORNE et Philipp KOEHN, Re-evaluating the Role of BLEU in Machine Translation Research. *In 11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 249–256. Association for Computational Linguistics, 2006.

(Cantini, 2023) Nicole Christine CANTINI, L'efficacité des pictogrammes compositionnels et non-compositionnels dans la communication médicale : une étude comparative. Mémoire de master, Université de Genève, 2023.

(Cardon et Bibal, 2023) Rémi CARDON et Adrien BIBAL, On Operations in Automatic Text Simplification. *In Proceedings of the Second Workshop on Text Simplification, Accessibility and Readability*, pages 116–130, Varna, Bulgaria, 2023.

(Cardon, 2018) Rémi CARDON, Approche lexicale de la simplification automatique de textes médicaux. *In Actes de la 25e conférence sur le Traitement Automatique des Langues Naturelles*, volume 2, pages 159–174, Rennes, France, 2018. ATALA.

(Cardon, 2021) Rémi CARDON, *Simplification automatique de textes techniques et spécialisés*. Thèse de doctorat, Université de Lille, 2021.

(Cardon *et al.*, 2022) Rémi CARDON, Adrien BIBAL, Rodrigo WILKENS, David ALFTER, Magali NORRÉ, Adeline MÜLLER, Patrick WATRIN et Thomas FRANÇOIS, Linguistic Corpus Annotation for Automatic Text Simplification Evaluation. *In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1842–1866, Abu Dhabi, United Arab Emirates, 2022. Association for Computational Linguistics.

(Cardon *et al.*, 2023) Rémi CARDON, Adrien BIBAL, Rodrigo WILKENS, David ALFTER, Magali NORRÉ, Adeline MÜLLER, Patrick WATRIN et Thomas FRANÇOIS, Annotation linguistique pour l'évaluation de la simplification automatique de textes. *In Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, volume 4, pages 35–48, Paris, France, 2023. ATALA.

(Cataix-Nègre, 2011) Élisabeth CATAIX-NÈGRE, *Communiquer autrement. Accompagner les personnes avec des troubles de la parole ou du langage : les communications alternatives*. Solal, 2011.

(Chandrasekar *et al.*, 1996) Raman CHANDRASEKAR, Christine DORAN et Srinivas BANGALORE, Motivations and Methods for Text Simplification. *In COLING 1996 Volume 2 : The 16th International Conference on Computational Linguistics*, pages 1041–1044, 1996.

(Chehab *et al.*, 2019) Nael CHEHAB, Hadmut HOLKEN et Mathilde MALGRANGE, Étude Recueil des besoins FALC, 2019. Rapport final. Holken Consultants & Partners, http://51.91.138.70/simples/docs/SIMPLES_Etude_Recueil_desBesoins_FALC_HC.pdf.

(Chen *et al.*, 2017) Ping CHEN, John ROCHFORD, David N. KENNEDY, Soussan DJAMASBI, Peter FAY et Will SCOTT, Automatic Text Simplification for People with Intellectual Disabilities. *In Proceedings of the International Conference on Artificial Intelligence Science and Technology*, pages 725–731, 2017.

(Chiang *et al.*, 2024) Wei-Lin CHIANG, Lianmin ZHENG, Ying SHENG, Anastasios Nikolas ANGELOPOULOS, Tianle LI, Dacheng LI, Hao ZHANG, Banghua ZHU, Michael JORDAN, Joseph E. GONZALEZ et Ion STOICA, Chatbot Arena : An Open Platform for Evaluating LLMs by Human Preference. *In Proceedings of the 41st International Conference on Machine Learning*, pages 8359–8388. PMLR, 2024.

(Christensen *et al.*, 2017) Carrie CHRISTENSEN, Doug REDD, Erica LAKE, Jean P. SHIPMAN, Heather AIONO, Roger ALTIZER JR, Bruce E. BRAY et Qing ZENG-TREITLER, Doodle

Health : A Crowdsourcing Game for the Co-design and Testing of Pictographs to Reduce Disparities in Healthcare Communication. *In AMIA Annual Symposium Proceedings*, volume 2017, pages 585–594. American Medical Informatics Association, 2017.

(Cohen, 1960) Jacob COHEN, A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

(Collignon et Giusiano, 2001) Patrick COLLIGNON et Bernard GIUSIANO, Validation of a pain evaluation scale for patients with severe cerebral palsy. *European Journal of Pain*, 5 (4):433–442, 2001.

(Conneau *et al.*, 2020) Alexis CONNEAU, Kartikay KHANDELWAL, Naman GOYAL, Vishrav CHAUDHARY, Guillaume WENZEK, Francisco GUZMÁN, Edouard GRAVE, Myle OTT, Luke ZETTLEMOYER et Veselin STOYANOV, Unsupervised Cross-lingual Representation Learning at Scale. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451. Association for Computational Linguistics, 2020.

(Côté, 1996) Roger A. CÔTÉ, Répertoire d'anatomopathologie de la SNOMED internationale, v3. 4. *Université de Sherbrooke, Sherbrooke, Québec*, 1996.

(Crunelle, 2018) Dominique CRUNELLE, *Évaluer et faciliter la communication des personnes en situation de handicap complexe : Polyhandicap, syndrome d'Angelman, syndrome de Rett, autres anomalies génétiques, autisme déficitaire, AVC sévère, traumatisme crânien, démences, ...* De Boeck Supérieur, 2018.

(De Belder et Moens, 2010) Jan DE BELDER et Marie-Francine MOENS, Text Simplification for Children. *In Proceedings of the SIGIR workshop on accessible search systems*, pages 19–26, 2010.

(de Knegt *et al.*, 2016a) Nanda DE KNEGT, Frank LOBBEZOO, Carlo SCHUENGEL, Heleen M. EVENHUIS et Erik J. A. SCHERDER, Self-Reporting Tool On Pain in People with Intellectual Disabilities (STOP-ID!) : A Usability Study. *Augmentative and Alternative Communication*, 32(1):1–11, 2016a.

(de Knegt *et al.*, 2013) Nanda DE KNEGT, Marjoleine J. C. PIEPER, Frank LOBBEZOO, Carlo SCHUENGEL, Heleen M. EVENHUIS, Jan PASSCHIER et Erik J. A. SCHERDER, Behavioral Pain Indicators in People With Intellectual Disabilities : A Systematic Review. *The Journal of Pain*, 14(9):885–896, 2013.

(de Knegt *et al.*, 2016b) Nanda DE KNEGT, Carlo SCHUENGEL, Frank LOBBEZOO, Corine M. VISSCHER, Heleen M. EVENHUIS, Judith A. BOEL et Erik J. A. SCHERDER, Comprehen-

sion of pictograms for pain quality and pain affect in adults with Down syndrome. *Journal of Intellectual & Developmental Disability*, 41(3):222–232, 2016b.

(Deng *et al.*, 2009) Jia DENG, Wei DONG, Richard SOCHER, Li-Jia LI, Kai LI et Li FEI-FEI, ImageNet : A large-scale hierarchical image database. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009.

(Devlin et Unthank, 2006) Siobhan DEVLIN et Gary UNTHANK, Helping Aphasic People Process Online Information. *In Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 225–226, 2006.

(Diacquenod et Santi, 2018) Cindy DIACQUENOD et France SANTI, La mise en œuvre du langage facile à lire et à comprendre (FALC) : enjeux, défis et perspectives. *Revue suisse de pédagogie spécialisée*, 2:29–35, 2018.

(Doddington, 2002) George DODDINGTON, Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. *In Proceedings of the second international conference on Human Language Technology Research*, pages 138–145, 2002.

(Dowse et Ehlers, 2001) Ros DOWSE et Martina S. EHLERS, The evaluation of pharmaceutical pictograms in a low-literate South African population. *Patient education and counseling*, 45(2):87–99, 2001.

(Elliah *et al.*, 2024) Abbhinav ELLIAH, Ananth Narayanan P., Bhuvan S. et P. MIRUNALINI, Text-To-Picto using Lexical Simplification. *In Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024)*, volume 3740, pages 1579–1584, Grenoble, France, 2024.

(Fauconnier, 2016) Jean-Philippe FAUCONNIER, *Acquisition de liens sémantiques à partir d'éléments de mise en forme des textes*. Thèse de doctorat, Université de Toulouse, 2016.

(Faustino *et al.*, 2020) Sara FAUSTINO, Sofia OLIVEIRA-MARTINS et Ana Margarida ADVINHA, Pictograms - A Useful (Digital and/or Physical) Tool to Assist Elderly Patients in Understanding Medication Instructions? – A Systematic Review. *In International Workshop on Gerontechnology*, pages 184–201. Springer, 2020.

(Feng *et al.*, 2023) Yutao FENG, Jipeng QIANG, Yun LI, Yunhao YUAN et Yi ZHU, Sentence Simplification via Large Language Models. *arXiv preprint arXiv :2302.11957*, 2023.

(Fitzgerald, 1949) Edith FITZGERALD, *Straight Language for the Deaf : A System of Instruction for Deaf Children*. Volta Bureau, 1949.

(Fox et Weisberg, 2019) John FOX et Sanford WEISBERG, *An R Companion to Applied Regression*. SAGE Publications, 2019.

(François, 2011) Thomas FRANÇOIS, *Les apports du traitement automatique du langage à la lisibilité du français langue étrangère*. Thèse de doctorat, Université catholique de Louvain, 2011.

(François *et al.*, 2016) Thomas FRANÇOIS, Mokhtar BILLAMI, Núria GALA et Delphine BERNHARD, Bleu, contusion, ecchymose : tri automatique de synonymes en fonction de leur difficulté de lecture et compréhension. *In Actes de la conférence conjointe JEP-TALN-RECITAL*, volume 2, pages 15–28. ATALA, 2016.

(François *et al.*, 2014) Thomas FRANÇOIS, Núria GALA, Patrick WATRIN et Cédrick FAIRON, FLELex : a graded lexical resource for French foreign learners. *In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3766–3773, Reykjavik, Iceland, 2014. European Language Resources Association.

(François *et al.*, 2020) Thomas FRANÇOIS, Adeline MÜLLER, Eva ROLIN et Magali NORRÉ, AMesure : A Web Platform to Assist the Clear Writing of Administrative Texts. *In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing : System Demonstrations*, pages 1–7, Suzhou, China, 2020. Association for Computational Linguistics.

(Fujisawa *et al.*, 2011) Kazuko FUJISAWA, Tomoyoshi INOUE, Yuko YAMANA et Humirhiro HAYASHI, The Effect of Animation on Learning Action Symbols by Individuals with Intellectual Disabilities. *Augmentative and Alternative Communication*, 27(1):53–60, 2011.

(Fuller et Lloyd, 1991) Donald FULLER et Lyle LLOYD, Toward a common usage of iconicity terminology. *Augmentative and Alternative Communication*, 7(3):215–220, 1991.

(Gala *et al.*, 2018) Núria GALA, Thomas FRANÇOIS, Ludivine JAVOUREY-DREVET et Johannes Christoph ZIEGLER, La simplification de textes, une aide à l'apprentissage de la lecture. *Langue française*, 3(199):123–131, 2018.

(Gala *et al.*, 2020) Núria GALA, Anaïs TACK, Ludivine JAVOUREY-DREVET, Thomas FRANÇOIS et Johannes C. ZIEGLER, Alector : A Parallel Corpus of Simplified French Texts with Alignments of Misreadings by Poor and Dyslexic Readers. *In Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1353–1361, Marseille, France, 2020. European Language Resources Association.

(Gerlach *et al.*, 2024) Johanna GERLACH, Pierrette BOUILLON, Jonathan MUTAL et Hervé SPECHBACH, A Concept Based Approach for Translation of Medical Dialogues into

Pictographs. *In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 233–242, Torino, Italy, 2024. European Language Resources Association and International Committee on Computational Linguistics.

(Gerlach *et al.*, 2023) Johanna GERLACH, Pierrette BOUILLON, Magali NORRÉ et Hervé SPECHBACH, Translating Medical Dialogues into Pictographs : An Approach Using UMLS. *In Caring is Sharing – Exploiting the Value in Data for Health and Innovation. Proceedings of the 33rd Medical Informatics Europe Conference*, pages 823–824, Gothenburg, Sweden, 2023.

(Grabar et Cardon, 2018) Natalia GRABAR et Rémi CARDON, CLEAR – Simple Corpus for Medical French. *In Proceedings of the 1st Workshop on Automatic Text Adaptation*, pages 3–9, Tilburg, the Netherlands, 2018. Association for Computational Linguistics.

(Grabar et Hamon, 2016) Natalia GRABAR et Thierry HAMON, A Large Rated Lexicon with French Medical Words. *In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2643–2648, Portorož, Slovenia, 2016. European Language Resources Association.

(Grave *et al.*, 2018) Edouard GRAVE, Piotr BOJANOWSKI, Prakhar GUPTA, Armand JOULIN et Tomas MIKOLOV, Learning Word Vectors for 157 Languages. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 3483–3487, Miyazaki, Japan, 2018. European Language Resources Association.

(Gu *et al.*, 2021) Yu GU, Robert TINN, Hao CHENG, Michael LUCAS, Naoto USUYAMA, Xiaodong LIU, Tristan NAUMANN, Jianfeng GAO et Hoifung POON, Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23, 2021.

(Hämeen-Anttila *et al.*, 2004) Katri HÄMEEN-ANTTILA, Kati KEMPPAINEN, Hannes ENLUND, J. Bush PATRICIA et Airaksinen MARJA, Do pictograms improve children's understanding of medicine leaflet information? *Patient Education and Counseling*, 55(3):371–378, 2004.

(Hart *et al.*, 1968) Peter E. HART, Nils J. NILSSON et Bertram RAPHAEL, A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

(Hartig, 2022) Florian HARTIG, *DHARMa : Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*, 2022. R package version 0.4.6, https://cran.r-project.org/web/packages/DHARMa.

(Hendy *et al.*, 2023) Amr HENDY, Mohamed ABDELREHIM, Amr SHARAF, Vikas RAUNAK, Mohamed GABR, Hitokazu MATSUSHITA, Young Jin KIM, Mohamed AFIFY et Hany Hassan AWADALLA, How Good Are GPT Models at Machine Translation ? A Comprehensive Evaluation. *arXiv preprint arXiv :2302.09210*, 2023.

(Hessels et Hessels-Schlatter, 2010) Marco HESSELS et Christine HESSELS-SCHLATTER, *Évaluation et intervention auprès d'élèves en difficultés*. Peter Lang, Hochfeldstrasse, Berne, 2010.

(Higginbotham *et al.*, 2012) D. Jeffery HIGGINBOTHAM, Gregory W. LESHER, Bryan J. MOULTON et Brian ROARK, The Application of Natural Language Processing to Augmentative and Alternative Communication. *Assistive Technology*, 24(1):14–24, 2012.

(Hill *et al.*, 2016) Brent HILL, Seneca PERRI-MOORE, Jinqiu KUANG, Bruce E. BRAY, Long NGO, Alexa DOIG et Qing ZENG-TREITLER, Automated pictographic illustration of discharge instructions with Glyph : impact on patient recall and satisfaction. *Journal of the American Medical Informatics Association*, 23(6):1136–1142, 2016.

(Houts *et al.*, 2006) Peter S. HOUTS, Cecilia C. DOAK, Leonard G. DOAK et Matthew J. LOSCALZO, The role of pictures in improving health communication : A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling*, 61(2):173–190, 2006.

(Howell *et al.*, 2008) David HOWELL, Yves BESTGEN, Vincent YZERBYT et Marylène ROGIER, *Méthodes statistiques en sciences humaines*. De Boeck, 2008.

(Hurtig et Downey, 2008) Richard R. HURTIG et A. Debora DOWNEY, *Augmentative and Alternative Communication in Acute and Critical Care Settings*. Plural Publishing, 2008.

(Hutchins, 2023) W. John HUTCHINS, Machine translation : History of research and applications. *In Routledge encyclopedia of translation technology*, pages 128–144. Routledge, 2023.

(Iacono et Johnson, 2004) Teresa IACONO et Hilary JOHNSON, Patients with disabilities and complex communication needs : The GP consultation. *Australian family physician*, 33(8), 2004.

(Ibba *et al.*, 2020) Frederica Carla IBBA, Lucie PILET, Camille GENECAND, Philippe SCHALLER, Isabelle ROYANNEZ et Guy Serge Antoine HALLER, Réseau handicap pour adultes à genève. *Revue Médicale Suisse*, 16(708):1807–1810, 2020.

(Imam *et al.*, 2019) Mai Farag IMAM, Amal Elsayed ABOUTABL et Ensaf H. MOHAMED, Automating Text Simplification Using Pictographs for People with Language Deficits. *I.J. Information Technology and Computer Science*, 9(1):26–34, 2019.

(Inclusion Europe, 2009) INCLUSION EUROPE, L'Information pour tous. Règles européennes pour une information facile à lire et à comprendre, 2009. https://www.unapei.org/wp-content/uploads/2018/11/L%E2%80%99information-pour-tous-Re%CC%80gles-europe%CC%81ennes-pour-une-information-facile-a%CC%80-lire-et-a%CC%80-comprendre.pdf.

(Ionescu *et al.*, 2024) Bogdan IONESCU, Henning MÜLLER, Ana Maria DRĂGULINESCU, Ahmad IDRISSI-YAGHIR, Ahmedkhan RADZHABOV, Alba Garcia Seco de HERRERA, Alexandra ANDREI, Alexandru STAN, Andrea M. STORÅS, Asma Ben ABACHA, Benjamin LECOUTEUX, Benno STEIN, Cécile MACAIRE, Christoph M. FRIEDRICH, Cynthia Sabrina SCHMIDT, Didier SCHWAB, Emmanuelle ESPERANÇA-RODIER, George IOANNIDIS, Griffin ADAMS, Henning SCHÄFER, Hugo MANGUINHAS, Ioan COMAN, Johanna SCHÖLER, Johannes KIESEL, Johannes RÜCKERT, Louise BLOCH, Martin POTTHAST, Maximilian HEINRICH, Meliha YETISGEN, A. Michael RIEGLER, Neal SNIDER, Pål HALVORSEN, Raphael BRÜNGEL, Steven A. HICKS, Vajira THAMBAWITA, Vassili KOVALEV, Yuri PROKOPCHUK et Wen-Wai YIM, Advancing Multimedia Retrieval in Medical, Social Media and Content Recommendation Applications with ImageCLEF 2024. *In Advances in Information Retrieval*, pages 44–52. Springer Nature, 2024.

(Janakiram *et al.*, 2020) Antony A. JANAKIRAM, Johanna GERLACH, Alyssa VUADENS-LEHMANN, Pierrette BOUILLON et Hervé SPECHBACH, *User Satisfaction with a Speech-Enabled Translator in Emergency Settings*, pages 1421–1422. Digital Personalized Health and Medicine. IOS, 2020.

(Jiao *et al.*, 2023) Wenxiang JIAO, Wenxuan WANG, Jen-tse HUANG, Xing WANG et Zhaopeng TU, Is ChatGPT a good translator? A preliminary study. *arXiv preprint arXiv :2301.08745*, 2023.

(Junczys-Dowmunt *et al.*, 2018) Marcin JUNCZYS-DOWMUNT, Roman GRUNDKIEWICZ, Tomasz DWOJAK, Hieu HOANG, Kenneth HEAFIELD, Tom NECKERMANN, Frank SEIDE, Ulrich GERMANN, Alham Fikri AJI, Nikolay BOGOYCHEV, André F. T. MARTINS et Alexandra BIRCH, Marian : Fast Neural Machine Translation in C++. *In Proceedings of ACL 2018, System Demonstrations*, pages 116–121, Melbourne, Australia, 2018. Association for Computational Linguistics.

(Jurafsky et Martin, 2024) Daniel JURAFSKY et James H. MARTIN, *Speech and Language Processing : An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. [Online], 2024. URL https://web.stanford.edu/~jurafsky/slp3. Online manuscript released August 20, 2024.

(Katz *et al.*, 2006) Marra G. KATZ, Sunil KRIPALANI et Barry D. WEISS, Use of pictorial aids

in medication instructions : A review of the literature. *American journal of health-system pharmacy*, 63(23):2391–2397, 2006.

(Keskinen *et al.*, 2012) Tuuli KESKINEN, Tomi HEIMONEN, Markku TURUNEN, Juha-Pekka RAJANIEMI et Sami KAUPPINEN, SymbolChat : A flexible picture-based communication platform for users with intellectual disabilities. *Interacting with Computers*, 24(5):374–386, 2012.

(Kew *et al.*, 2023) Tannon KEW, Alison CHI, Laura VÁSQUEZ-RODRÍGUEZ, Sweta AGRAWAL, Dennis AUMILLER, Fernando ALVA-MANCHEGO et Matthew SHARDLOW, BLESS : Benchmarking Large Language Models on Sentence Simplification. *In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13291–13309, Singapore, 2023. Association for Computational Linguistics.

(Kim *et al.*, 2009) Hyeoneui KIM, Carlos NAKAMURA et Qing ZENG, Assessment of Pictographs Developed Through a Participatory Design Process Using an Online Survey Tool. *Journal of Medical Internet Research*, 11(1):e4, 2009.

(Kitaev et Klein, 2018) Nikita KITAEV et Dan KLEIN, Constituency Parsing with a Self-Attentive Encoder. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 2676–2686, Melbourne, Australia, 2018. Association for Computational Linguistics.

(Klein *et al.*, 2017) Guillaume KLEIN, Yoon KIM, Yuntian DENG, Jean SENELLART et Alexander RUSH, OpenNMT : Open-Source Toolkit for Neural Machine Translation. *In Proceedings of ACL 2017, System Demonstrations*, pages 67–72, Vancouver, Canada, 2017. Association for Computational Linguistics.

(Koehn, 2009) Philipp KOEHN, *Statistical Machine Translation*. Cambridge University Press, Cambridge, 2009.

(Koehn *et al.*, 2007) Philipp KOEHN, Hieu HOANG, Alexandra BIRCH, Chris CALLISON-BURCH, Marcello FEDERICO, Nicola BERTOLDI, Brooke COWAN, Wade SHEN, Christine MORAN, Richard ZENS, Chris DYER, Ondřej BOJAR, Alexandra CONSTANTIN et Evan HERBST, Moses : Open Source Toolkit for Statistical Machine Translation. *In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 177–180, Prague, Czech Republic, 2007. Association for Computational Linguistics.

(Koptient *et al.*, 2019) Anaïs KOPTIENT, Rémi CARDON et Natalia GRABAR, Simplification-induced transformations : typology and some characteristics. *In Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 309–318, Florence, Italy, 2019. Association for Computational Linguistics.

(Koptient et Grabar, 2020) Anaïs KOPTIENT et Natalia GRABAR, Rated Lexicon for the Simplification of Medical Texts. *In The Fifth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing HEALTHINFO 2020*, 2020.

(Koushik *et al.*, 2024) Avaneesh KOUSHIK, Morrison Jithu S, Mirunalini P. et Jothir Aditya R. K., A Transformer Based Approach for Text-to-Picto Generation. *In Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024)*, volume 3740, pages 1656–1661, Grenoble, France, 2024.

(Krippendorff, 2004) Klaus KRIPPENDORFF, Reliability in Content Analysis : Some Common Misconceptions and Recommendations. *Human Communication Research*, 30(3): 411–433, 2004.

(Kuhn, 2008) Max KUHN, Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5):1–26, 2008.

(Kultsova *et al.*, 2017) Marina KULTSOVA, Dmitry MATYUSHECHKIN, Andrey USOV, Svetlana KARPOVA et Roman ROMANENKO, Generation of pictograph sequences from the Russian text in the assistive mobile application for people with intellectual and developmental disabilities. *In The 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 294–297. IEEE, 2017.

(Labrak *et al.*, 2023a) Yanis LABRAK, Adrien BAZOGE, Richard DUFOUR, Mickael ROUVIER, Emmanuel MORIN, Béatrice DAILLE et Pierre-Antoine GOURRAUD, DrBERT : Un modèle robuste pré-entraîné en français pour les domaines biomédical et clinique. *In Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, volume 4, pages 109–120, Paris, France, 2023a. ATALA.

(Labrak *et al.*, 2023b) Yanis LABRAK, Adrien BAZOGE, Richard DUFOUR, Mickael ROUVIER, Emmanuel MORIN, Béatrice DAILLE et Pierre-Antoine GOURRAUD, DrBERT : A Robust Pre-trained Model in French for Biomedical and Clinical domains. *In Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics (ACL'23)*, pages 16207–16221, Toronto, Canada, 2023b. Association for Computational Linguistics.

(Lafourcade, 2007) Mathieu LAFOURCADE, Making people play for Lexical Acquisition with the JeuxDeMots prototype. *In SNLP'07 : 7th International Symposium on Natural Language Processing*, 2007.

(Landis et Koch, 1977) J. Richard LANDIS et Gary G. KOCH, The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, 1977.

(Le *et al.*, 2020) Hang LE, Loïc VIAL, Jibril FREJ, Vincent SEGONNE, Maximin COAVOUX, Benjamin LECOUTEUX, Alexandre ALLAUZEN, Benoît CRABBÉ, Laurent BESACIER et Didier SCHWAB, FlauBERT : Unsupervised Language Model Pre-training for French. *In Proceedings of the 12th Language Resources and Evaluation Conference*, pages 2479–2490, Marseille, France, 2020. European Language Resources Association.

(Lefeuvre-Halftermeyer *et al.*, 2016) Anaïs LEFEUVRE-HALFTERMEYER, Virginie GOVAERE, Jean-Yves ANTOINE, Willy ALLEGRE, Samuel POUPLIN, Jean-Paul DEPARTE, Samia SLIMANI et Aurore SPAGNULO, Typologie des risques pour une analyse éthique de l'impact des technologies du TAL. *Traitement Automatique des Langues*, 57(2):47–71, 2016.

(Levenshtein, 1966) Vladimir LEVENSHTEIN, Binary codes capable of correcting deletions, insertions, and reversals. *In Soviet physics doklady*, volume 10, pages 707–710. Soviet Union, 1966.

(Lüdecke *et al.*, 2021) Daniel LÜDECKE, Mattan S. BEN-SHACHAR, Indrajeet PATIL, Philip WAGGONER et Dominique MAKOWSKI, performance : An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, 6 (60):3139, 2021.

(Macaire *et al.*, 2023a) Cécile MACAIRE, Chloé DION, Jordan ARRIGO, Claire LEMAIRE, Emmanuelle ESPERANÇA-RODIER, Benjamin LECOUTEUX et Didier SCHWAB, Propicto, 2023a. ORTOLANG (Open Resources and TOols for LANGuage), https://hdl.handle.net/11403/propicto.

(Macaire *et al.*, 2024a) Cécile MACAIRE, Chloé DION, Jordan ARRIGO, Claire LEMAIRE, Emmanuelle ESPERANÇA-RODIER, Benjamin LECOUTEUX et Didier SCHWAB, A Multimodal French Corpus of Aligned Speech, Text, and Pictogram Sequences for Speech-to-Pictogram Machine Translation. *In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 839–849, Torino, Italy, 2024a. European Language Resources Association and International Committee on Computational Linguistics.

(Macaire *et al.*, 2024b) Cécile MACAIRE, Chloé DION, Didier SCHWAB, Benjamin LECOUTEUX et Emmanuelle ESPERANÇA-RODIER, Approches cascade et de bout-en-bout pour la traduction automatique de la parole en pictogrammes. *In Actes de JEP-TALN-RECITAL 2024. Actes de la 31e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, volume 1, pages 22–35, Toulouse, France, 2024b. ATALA & AFPC.

(Macaire *et al.*, 2024c) Cécile MACAIRE, Chloé DION, Didier SCHWAB, Benjamin LECOUTEUX et Emmanuelle ESPERANÇA-RODIER, Un corpus multimodal alignant parole, trans-

cription et séquences de pictogrammes dédié à la traduction automatique de la parole vers des pictogrammes. *In Actes de JEP-TALN-RECITAL 2024. Actes de la 31e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, volume 2, pages 20–21, Toulouse, France, 2024c. ATALA & AFPC.

(Macaire *et al.*, 2023b) Cécile MACAIRE, Emmanuelle ESPERANÇA-RODIER, Benjamin LECOUTEUX et Didier SCHWAB, Voice2Picto : un système de traduction automatique de la parole vers des pictogrammes. *In Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, volume 5, pages 10–13, Paris, France, 2023b. ATALA.

(Macaire *et al.*, 2022) Cécile MACAIRE, Lucía ORMAECHEA GRIJALBA et Adrien PUPIER, Une chaîne de traitements pour la simplification automatique de la parole et sa traduction automatique vers des pictogrammes. *In Actes de la 29e conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, volume 2, pages 111–123, Avignon, France, 2022. ATALA.

(Mali, 2023) Rebeka Karola MALI, BabelDr vs. Google Translate : translation in a pharmacy setting. Mémoire de master, Université de Genève, 2023.

(Martin *et al.*, 2022) Louis MARTIN, Angela FAN, Eric DE LA CLERGERIE, Antoine BORDES et Benoît SAGOT, MUSS : Multilingual Unsupervised Sentence Simplification by Mining Paraphrases. *In Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1651–1664, Marseille, France, 2022. European Language Resources Association.

(Martin *et al.*, 2020) Louis MARTIN, Benjamin MULLER, Pedro Javier ORTIZ SUÁREZ, Yoann DUPONT, Laurent ROMARY, Eric DE LA CLERGERIE, Djamé SEDDAH et Benoît SAGOT, CamemBERT : a Tasty French Language Model. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219. Association for Computational Linguistics, 2020.

(McCrae *et al.*, 2020) John Philip MCCRAE, Alexandre RADEMAKER, Ewa RUDNICKA et Francis BOND, English WordNet 2020 : Improving and Extending a WordNet for English using an Open-Source Methodology. *In Proceedings of the LREC 2020 : Workshop on Multimodal Wordnets (MMW2020)*, pages 14–19, 2020.

(McNemar, 1947) Quinn MCNEMAR, Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157, 1947.

(Mesko *et al.*, 2011) Phyllis J. MESKO, Aris Beoglos ELIADES, Cheryl CHRIST-LIBERTIN et Debra SHELESTAK, Use of picture communication aids to assess pain location in pediatric postoperative patients. *Journal of Perianesthesia Nursing*, 26(6):395–404, 2011.

(Mihalcea et Leong, 2008) Rada MIHALCEA et Chee Wee LEONG, Toward Communicating Simple Sentences Using Pictorial Representations. *Machine Translation*, 22(3):153–173, 2008.

(Mikolov *et al.*, 2013) Tomas MIKOLOV, Ilya SUTSKEVER, Kai CHEN, Greg CORRADO et Jeff DEAN, Distributed Representations of Words and Phrases and their Compositionality. *In Proceedings of the 26th International Conference on Neural Information Processing Systems*, volume 26, pages 3111–3119. Curran Associates Inc., 2013.

(Miller, 1995) George A. MILLER, WordNet : A Lexical Database for English. *Communications of the ACM*, 38(11):39–41, 1995.

(Mouton et de Chalendar, 2010) Claire MOUTON et Gaël de CHALENDAR, JAWS : Just Another WordNet Subset. *In Actes de la 17e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, pages 111–116, Montréal, Canada, 2010. ATALA.

(Mutal *et al.*, 2022) Jonathan MUTAL, Pierrette BOUILLON, Magali NORRÉ, Johanna GERLACH et Lucía ORMAECHEA GRIJALBA, A Neural Machine Translation Approach to Translate Text to Pictographs in a Medical Speech Translation System – The BabelDr Use Case. *In Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas*, pages 252–263, Orlando, USA, 2022. Association for Machine Translation in the Americas.

(Nakamura et Zeng-Treitler, 2012) Carlos NAKAMURA et Qing ZENG-TREITLER, A taxonomy of representation strategies in iconic communication. *International journal of human-computer studies*, 70(8):535–551, 2012.

(Nandy, 2019) Ankita NANDY, Beyond words : Pictograms for Indian languages. *International Journal of Research in Science and Technology*, 9(1):19–25, 2019.

(Nasiruddin *et al.*, 2015) Mohammad NASIRUDDIN, Andon TCHECHMEDJIEV, Hervé BLANCHON et Didier SCHWAB, Création rapide et efficace d'un système de désambiguïsation lexicale pour une langue peu dotée. *In Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles*, pages 83–94, Caen, France, 2015. ATALA.

(Nisioi *et al.*, 2017) Sergiu NISIOI, Sanja ŠTAJNER, Simone Paolo PONZETTO et Liviu P. DINU, Exploring Neural Text Simplification Models. *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 85–91, Vancouver, Canada, 2017. Association for Computational Linguistics.

(Norman, 2013) Don NORMAN, *The Design of Everyday Things*. Basic Books, 2013.

(Norré, 2020a) Magali NORRÉ, Evaluation of a Word Prediction System in an Augmentative and Alternative Communication for Disabled People. *Modelling, Measurement and Control C*, 81(1-4):49–54, 2020a.

(Norré, 2020b) Magali NORRÉ, Évaluation de la prédiction de mots dans un système de communication améliorée et alternative pour les personnes en situation de handicap : Presage et ACAT. *In Actes de la 11e conférence de l'IFRATH sur les technologies d'assistance. Handicap 2020 : Technologies pour l'autonomie et l'inclusion*, pages 183–188. IFRATH, 2020b.

(Norré *et al.*, 2020) Magali NORRÉ, Pierrette BOUILLON, Johanna GERLACH et Hervé SPECHBACH, Évaluation de la compréhension de pictogrammes Arasaac et Sclera pour améliorer l'accessibilité du système de traduction médicale BabelDr. *In Actes de la 11e conférence de l'IFRATH sur les technologies d'assistance. Handicap 2020 : Technologies pour l'autonomie et l'inclusion*, pages 177–182. IFRATH, 2020.

(Norré *et al.*, 2021a) Magali NORRÉ, Pierrette BOUILLON, Johanna GERLACH et Hervé SPECHBACH, BabelDr : un système de traduction médicale avec des pictogrammes pour les patients allophones aux urgences et dans un secteur de dépistage COVID-19. *In Journée Technologies du Langage Humain (TLH) et Santé*, pages 6–7. AfIA & ATALA, 2021a.

(Norré *et al.*, 2021b) Magali NORRÉ, Pierrette BOUILLON, Johanna GERLACH et Hervé SPECHBACH, Evaluating the comprehension of Arasaac and Sclera pictographs for the BabelDr patient response interface. *In Proceedings of the 3rd Swiss Conference on Barrier-free Communication*, pages 55–63. Zurich University of Applied Sciences (ZHAW), 2021b.

(Norré *et al.*, 2023) Magali NORRÉ, Rémi CARDON, Vincent VANDEGHINSTE et Thomas FRANÇOIS, Word Sense Disambiguation for Automatic Translation of Medical Dialogues into Pictographs. *In Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing (RANLP 2023)*, pages 803–812, Varna, Bulgaria, 2023. INCOMA Ltd.

(Norré et De Wilde, 2018) Magali NORRÉ et Max DE WILDE, Configuration et évaluation d'un système de prédiction de mots au sein d'un logiciel de Communication Améliorée et Alternative (CAA). *In Proceedings of the 2nd Swiss Conference on Barrier-free Communication*, pages 71–74, Geneva, Switzerland, 2018.

(Norré *et al.*, 2024) Magali NORRÉ, Trang Trần Hạnh PHẠM, Pierrette BOUILLON, Vincent VANDEGHINSTE et Thomas FRANÇOIS, Évaluation de la traduction en pictogrammes pour la communication médecin-patient par des adultes avec une déficience intellectuelle. *In Actes de la 13e conférence de l'IFRATH sur les technologies d'assistance.*

Handicap 2024 : Des solutions personnalisées pour des besoins spécifiques, pages 181–186, Paris, France, 2024. IFRATH.

(Norré *et al.*, 2021c) Magali NORRÉ, Vincent VANDEGHINSTE, Pierrette BOUILLON et Thomas FRANÇOIS, Experiments for the Adaptation of Text2Picto to French. *In Meeting of Computational Linguistics in The Netherlands (CLIN31)*, 2021c.

(Norré *et al.*, 2021d) Magali NORRÉ, Vincent VANDEGHINSTE, Pierrette BOUILLON et Thomas FRANÇOIS, Extending a Text-to-Pictograph System to French and to Arasaac. *In Proceedings of the 13th International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1050–1059. INCOMA Ltd, 2021d.

(Norré *et al.*, 2022) Magali NORRÉ, Vincent VANDEGHINSTE, Thomas FRANÇOIS et Pierrette BOUILLON, Investigating the Medical Coverage of a Translation System into Pictographs for Patients with an Intellectual Disability. *In Proceedings of the Ninth Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2022)*, pages 44–49, Dublin, Ireland, 2022. Association for Computational Linguistics.

(Orkphol et Yang, 2019) Korawit ORKPHOL et Wu YANG, Word Sense Disambiguation Using Cosine Similarity Collaborates with Word2vec and WordNet. *Future Internet*, 11(5): 114, 2019.

(Ormaechea Grijalba *et al.*, 2023a) Lucía ORMAECHEA GRIJALBA, Pierrette BOUILLON, Maximin COAVOUX, Emmanuelle ESPERANÇA-RODIER, Johanna GERLACH, Jerôme GOULIAN, Benjamin LECOUTEUX, Cécile MACAIRE, Jonathan MUTAL, Magali NORRÉ, Adrien PUPIER et Didier SCHWAB, PROPICTO : Developing Speech-to-Pictograph Translation Systems to Enhance Communication Accessibility. *In Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 515–516, Tampere, Finland, 2023a. European Association for Machine Translation.

(Ormaechea Grijalba *et al.*, 2023b) Lucía ORMAECHEA GRIJALBA, Pierrette BOUILLON, Maximin COAVOUX, Emmanuelle ESPERANÇA-RODIER, Johanna GERLACH, Jerôme GOULIAN, Benjamin LECOUTEUX, Cécile MACAIRE, Jonathan MUTAL, Magali NORRÉ, Adrien PUPIER, Didier SCHWAB et Hervé SPECHBACH, PROPICTO : Développer des systèmes de traduction de la parole vers des séquences de pictogrammes pour améliorer l'accessibilité de la communication. *In Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, volume 6, pages 32–35, Paris, France, 2023b. ATALA.

(Ormaechea Grijalba et Tsourakis, 2023) Lucía ORMAECHEA GRIJALBA et Nikos TSOURAKIS, Extracting Sentence Simplification Pairs from French Comparable Corpora Using a

Two-Step Filtering Method. *In Proceedings of the 8th edition of the Swiss Text Analytics Conference*, pages 30–40, Neuchatel, Switzerland, 2023. Association for Computational Linguistics.

(Ormaechea Grijalba *et al.*, 2024) Lucía ORMAECHEA GRIJALBA, Nikos TSOURAKIS, Didier SCHWAB, Pierrette BOUILLON et Benjamin LECOUTEUX, Simplification Strategies in French Spontaneous Speech. *In Proceedings of the Workshop on DeTermIt! Evaluating Text Difficulty in a Multilingual Context @ LREC-COLING 2024*, pages 90–102, Torino, Italy, 2024. European Language Resources Association and International Committee on Computational Linguistics.

(Ortiz Suárez *et al.*, 2019) Pedro Javier ORTIZ SUÁREZ, Benoît SAGOT et Laurent ROMARY, Asynchronous Pipeline for Processing Huge Corpora on Medium to Low Resource Infrastructures. *In 7th Workshop on the Challenges in the Management of Large Corpora (CMLC-7)*, 2019.

(Othman et Al-Sinani, 2021) Achraf OTHMAN et Aisha AL-SINANI, Tawasol Symbols : Alternative Augmented Communication Pictograms to Support the Inclusion During Pandemics. *In* Daniel BURGOS, Ahmed TLILI et Anita TABACCO, éditeurs, *Radical Solutions for Education in a Crisis Context : COVID-19 as an Opportunity for Global Learning*, pages 225–239. Springer, Singapore, 2021.

(Paetzold et Specia, 2017) Gustavo H. PAETZOLD et Lucia SPECIA, A Survey on Lexical Simplification. *Journal of Artificial Intelligence Research*, 60:549–593, 2017.

(Paolieri et Marful, 2018) Daniela PAOLIERI et Alejandra MARFUL, Norms for a Pictographic System : the Aragonese Portal of Augmentative/Alternative Communication (ARASAAC) System. *Frontiers in psychology*, 9:1–9, 2018.

(Papineni *et al.*, 2002) Kishore PAPINENI, Salim ROUKOS, Todd WARD et Wei-Jing ZHU, BLEU : A method for automatic evaluation of machine translation. *In Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318, 2002.

(Park *et al.*, 2022) Chanjun PARK, Yoonna JANG, Seolhwa LEE, Jaehyung SEO, Kisu YANG et Heuiseok LIM, PicTalky : Augmentative and Alternative Communication for Language Developmental Disabilities. *In Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing : System Demonstrations*, pages 17–27, Taipei, Taiwan, 2022. Association for Computational Linguistics.

(Patil et Davies, 2014) Sumant PATIL et Patrick DAVIES, Use of Google Translate in medical communication : evaluation of accuracy. *British Medical Journal*, 349, 2014.

(Peng *et al.*, 2023) Keqin PENG, Liang DING, Qihuang ZHONG, Li SHEN, Xuebo LIU, Min ZHANG, Yuanxin OUYANG et Dacheng TAO, Towards Making the Most of ChatGPT for Machine Translation. *In Findings of the Association for Computational Linguistics : EMNLP 2023*, pages 5622–5633, Singapore, 2023. Association for Computational Linguistics.

(Pereira *et al.*, 2022) Jayr Alencar PEREIRA, David MACÊDO, Cleber ZANCHETTIN, Adriano Lorena Inácio DE OLIVEIRA et Robson do NASCIMENTO FIDALGO, PictoBERT : Transformers for next pictogram prediction. *Expert Systems with Applications*, 202: 117231, 2022.

(Post, 2018) Matt POST, A Call for Clarity in Reporting BLEU Scores. *In Proceedings of the Third Conference on Machine Translation*, pages 186–191, Brussels, Belgium, 2018. Association for Computational Linguistics.

(Pradet *et al.*, 2014) Quentin PRADET, Gaël DE CHALENDAR et Jeanne BAGUENIER DESORMEAUX, WoNeF, an improved, expanded and evaluated automatic French translation of WordNet. *In Proceedings of the Seventh Global Wordnet Conference*, pages 32–39, Tartu, Estonia, 2014. University of Tartu Press.

(Radford *et al.*, 2023) Alec RADFORD, Jong Wook KIM, Tao XU, Greg BROCKMAN, Christine MCLEAVEY et Ilya SUTSKEVER, Robust Speech Recognition via Large-Scale Weak Supervision. *In International Conference on Machine Learning*, pages 28492–28518. PMLR, 2023.

(Rayner *et al.*, 2018) Manny RAYNER, Johanna GERLACH, Pierrette BOUILLON, Nikos TSOURAKIS et Hervé SPECHBACH, Handling Ellipsis in a Spoken Medical Phraselator. *In Statistical Language and Speech Processing*, pages 140–152. Springer, 2018.

(Rello *et al.*, 2013) Luz RELLO, Ricardo BAEZA-YATES, Stefan BOTT et Horacio SAGGION, Simplify or Help ? Text Simplification Strategies for People with Dyslexia. *In Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, pages 1–10, New York, NY, USA, 2013. Association for Computing Machinery.

(Rodríguez Vázquez et Bouillon, 2021) Silvia RODRÍGUEZ VÁZQUEZ et Pierrette BOUILLON, A Reception Study of Machine-Translated Easy Language Text by Individuals with Reading Difficulties. *In 3rd International Conference on Translation, Interpreting and Cognition (ICTIC3)*, Forlì, Italy, 2021.

(Rolin *et al.*, 2021) Eva ROLIN, Quentin LANGLOIS, Patrick WATRIN et Thomas FRANÇOIS, FrenLyS : A tool for the automatic simplification of French general language texts. *In Proceedings of the 13th International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1196–1205. INCOMA Ltd, 2021.

(Romski et Sevcik, 2005) MaryAnn ROMSKI et Rose A. SEVCIK, Augmentative Communication and Early Intervention : Myths and Realities. *Infants & Young Children*, 18 (3):174–185, 2005.

(Saggion, 2017) Horacio SAGGION, Automatic Text Simplification. *Synthesis Lectures on Human Language Technologies*, 10(1):1–137, 2017.

(Saggion *et al.*, 2015) Horacio SAGGION, Sanja ŠTAJNER, Stefan BOTT, Simon MILLE, Luz RELLO et Biljana DRNDAREVIC, Making It Simplext : Implementation and Evaluation of a Text Simplification System for Spanish. *ACM Transactions on Accessible Computing (TACCESS)*, 6(4):1–36, 2015.

(Sagot et Fišer, 2008) Benoît SAGOT et Darja FIŠER, Building a free French WordNet from multilingual resources. *In OntoLex*, Marrakech, Morocco, 2008.

(Saldanha et O'Brien, 2014) Gabriela SALDANHA et Sharon O'BRIEN, *Research Methodologies in Translation Studies*. Routledge, 2014.

(Schmid, 1994) Helmut SCHMID, Probabilistic Part-of-Speech Tagging Using Decision Trees. *In Proceedings of the International Conference on New Methods in Language Processing*, 1994.

(Schwab, 2021) Didier SCHWAB, Contributions au Traitement Automatique des Langues et à un domaine d'application, la Communication Alternative et Augmentée, 2021. Habilitation à diriger des recherches. Université Grenoble Alpes.

(Schwab *et al.*, 2018) Didier SCHWAB, Amela FEJZA, Loïc VIAL et Yann ROBERT, The GazePlay Project : Open and Free Eye-Trackers Games and a Community for People with Multiple Disabilities. *In International Conference on Computers Helping People with Special Needs*, pages 254–261. Springer, 2018.

(Schwab *et al.*, 2020) Didier SCHWAB, Pauline TRIAL, Céline VASCHALDE, Loïc VIAL, Emmanuelle ESPERANÇA-RODIER et Benjamin LECOUTEUX, Providing semantic knowledge to a set of pictograms for people with disabilities : A set of links between WordNet and Arasaac : Arasaac-WN. *In Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 166–171, Marseille, France, 2020. European Language Resources Association.

(Schwarz, 1978) Gideon SCHWARZ, Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978.

(Segonne *et al.*, 2024a) Vincent SEGONNE, Aidan MANNION, Laura ALONZO-CANUL, Alexandre AUDIBERT, Xingyu LIU, Cécile MACAIRE, Adrien PUPIER, Yongxin ZHOU, Mathilde AGUIAR, Felix HERRON, Magali NORRÉ, Massih-Reza AMINI, Pierrette

BOUILLON, Iris ESHKOL-TARAVELLA, Emmanuelle ESPERANÇA-RODIER, Thomas FRANÇOIS, Lorraine GOEURIOT, Jérôme GOULIAN, Mathieu LAFOURCADE, Benjamin LECOUTEUX, François PORTET, Fabien RINGEVAL, Vincent VANDEGHINSTE, Maximin COAVOUX, Marco DINARELLI et Didier SCHWAB, Jargon : A Suite of Language Models and Evaluation Tasks for French Specialized Domains. *In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9463–9476, Torino, Italy, 2024a. European Language Resources Association and International Committee on Computational Linguistics.

(Segonne *et al.*, 2024b) Vincent SEGONNE, Aidan MANNION, Alexandre AUDIBERT, Xingyu LIU, Cécile MACAIRE, Adrien PUPIER, Yongxin ZHOU, Mathilde AGUIAR, Magali NORRÉ, Felix HERRON, Iris ESHKOL-TARAVELLA, Thomas FRANÇOIS, Goeuriot LORRAINE, Jérôme GOULIAN, Mathieu LAFOURCADE, Benjamin LECOUTEUX, François PORTET, Fabien RINGEVAL, Maximin COAVOUX, Vincent VANDEGHINSTE, Marco DINARELLI, Laura ALONZO-CANUL, Massih-Reza AMINI, Pierrette BOUILLON, Didier SCHWAB et Emmanuelle ESPERANÇA-RODIER, Jargon : Une suite de modèles de langues et de référentiels d'évaluation pour les domaines spécialisés du français. *In Actes de JEP-TALN-RECITAL 2024. Actes de la 31e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, volume 2, pages 9–10, Toulouse, France, 2024b. ATALA & AFPC.

(Seidl et Vandeghinste, 2024) Theresa SEIDL et Vincent VANDEGHINSTE, Controllable Sentence Simplification in Dutch. *Computational Linguistics in the Netherlands Journal*, 13:31–61, 2024.

(Seligman et Dillinger, 2013) Mark SELIGMAN et Mike DILLINGER, Automatic Speech Translation for Healthcare : Some Internet and Interface Aspects. *In Proceedings of the 10th International Conference on Terminology and Artificial Intelligence (TIA-13)*, pages 28–30, 2013.

(Sennott *et al.*, 2016) Samuel C. SENNOTT, Janice C. LIGHT et David MCNAUGHTON, AAC Modeling Intervention Research Review. *Research and Practice for Persons with Severe Disabilities*, 41(2):101–115, 2016.

(Seretan, 2012) Violeta SERETAN, Acquisition of Syntactic Simplification Rules for French. *In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 4019–4026, Istanbul, Turkey, 2012. European Language Resources Association.

(Sevens, 2018) Leen SEVENS, *Words Divide, Pictographs Unite : Pictograph Communication*

Technologies for People with an Intellectual Disability. LOT, Utrecht, The Netherlands, 2018.

(Sevens *et al.*, 2016a) Leen SEVENS, Gilles JACOBS, Vincent VANDEGHINSTE, Ineke SCHUURMAN et Frank VAN EYNDE, Improving Text-To-Pictograph Translation Through Word Sense Disambiguation. *In Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 131–135, Berlin, Germany, 2016a. Association for Computational Linguistics.

(Sevens *et al.*, 2016b) Leen SEVENS, Tom VANALLEMEERSCH, Ineke SCHUURMAN, Vincent VANDEGHINSTE et Frank VAN EYNDE, Automated Spelling Correction for Dutch Internet Users with Intellectual Disabilities. *In Proceedings of 1st LREC Workshop on Improving Social Inclusion using NLP : Tools and Resources (ISI-NLP)*, pages 11–19, Portorož, Slovenia, 2016b.

(Sevens *et al.*, 2015a) Leen SEVENS, Vincent VANDEGHINSTE, Ineke SCHUURMAN et Frank VAN EYNDE, Extending a Dutch Text-to-Pictograph Converter to English and Spanish. *In Proceedings of SLPAT 2015 : 6th Workshop on Speech and Language Processing for Assistive Technologies*, pages 110–117, Dresden, Germany, 2015a. Association for Computational Linguistics.

(Sevens *et al.*, 2015b) Leen SEVENS, Vincent VANDEGHINSTE, Ineke SCHUURMAN et Frank VAN EYNDE, Natural Language Generation from Pictographs. *In Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*, pages 71–75, Brighton, UK, 2015b. Association for Computational Linguistics.

(Sevens *et al.*, 2017) Leen SEVENS, Vincent VANDEGHINSTE, Ineke SCHUURMAN et Frank VAN EYNDE, Simplified Text-to-Pictograph Translation for People with Intellectual Disabilities. *In International Conference on Applications of Natural Language to Information Systems*, pages 185–196. Springer, 2017.

(Shiojiri *et al.*, 2013) Misato SHIOJIRI, Yukari NAKATANI et Tomoko YONEZAWA, Visual language communication system with multiple pictograms converted from weblog texts. *IASDR2013*, 2013.

(Shiwen, 1993) Yu SHIWEN, Automatic Evaluation of Output Quality for Machine Translation Systems. *Machine Translation*, 8(1):117–126, 1993.

(Smith et Fellbaum, 2004) Barry SMITH et Christiane FELLBAUM, Medical WordNet : A New Methodology for the Construction and Validation of Information Resources for Consumer Health. *In Proceedings of the 20th International Conference on Computational Linguistics*, pages 371–382, Geneva, Switzerland, 2004. COLING.

(Sohail et Traum, 2019) Usman SOHAIL et David TRAUM, A Blissymbolics Translation System. *In Proceedings of SLPAT 2019 : 8th Workshop on Speech and Language Processing for Assistive Technologies*, pages 32–36. Association for Computational Linguistics, 2019.

(Spechbach *et al.*, 2019) Hervé SPECHBACH, Johanna GERLACH, Sanae Mazouri KARKER, Nikos TSOURAKIS, Christophe COMBESCURE et Pierrette BOUILLON, A Speech-Enabled Fixed-Phrase Translator for Emergency Settings : Crossover Study. *JMIR medical informatics*, 7(2):e13167, 2019.

(Štajner et Popović, 2016) Sanja ŠTAJNER et Maja POPOVIĆ, Can Text Simplification Help Machine Translation ? *In Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, pages 230–242, Riga, Latvia, 2016.

(Strasly *et al.*, 2018) Irene STRASLY, Tanya SEBAÏ, Evelyne RIGOT, Valentin MARTI, Jesus Manuel GONZALEZ, Johanna GERLACH, Hervé SPECHBACH et Pierrette BOUILLON, Le projet BabelDr : rendre les informations médicales accessibles en Langue des Signes de Suisse Romande (LSF-SR). *In Proceedings of the 2nd Swiss Conference on Barrierfree Communication*, pages 92–96, Geneva, Switzerland, 2018.

(Tack, 2021) Anaïs TACK, *Mark My Words ! On the Automated Prediction of Lexical Difficulty for Foreign Language Readers*. Thèse de doctorat, Université catholique de Louvain & KU Leuven, 2021.

(Tenny, 2016) Carol TENNY, A linguist looks at AAC : Language representation systems for augmentative and alternative communication, compared with writing systems and natural language. *Writing Systems Research*, 8(1):84–119, 2016.

(Tiedemann et Thottingal, 2020) Jörg TIEDEMANN et Santhosh THOTTINGAL, OPUS-MT – Building open translation services for the World. *In Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480, Lisboa, Portugal, 2020.

(Todirascu *et al.*, 2022) Amalia TODIRASCU, Rodrigo WILKENS, Eva ROLIN, Thomas FRANÇOIS, Delphine BERNHARD et Núria GALA, HECTOR : A Hybrid TExt SimplifiCation TOol for Raw Texts in French. *In Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4620–4630, Marseille, France, 2022. European Language Resources Association.

(Touchent *et al.*, 2023) Rian TOUCHENT, Laurent ROMARY et Eric DE LA CLERGERIE, CamemBERT-bio : Un modèle de langue français savoureux et meilleur pour la santé. *In Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement*

Automatique des Langues Naturelles (TALN), volume 1, pages 323–334, Paris, France, 2023. ATALA.

(Tuffrey-Wijne et McEnhill, 2008) Irene TUFFREY-WIJNE et Linda MCENHILL, Communication difficulties and intellectual disability in end-of-life care. *International Journal of Palliative Nursing*, 14(4):189–194, 2008.

(Unapei, 2020) UNAPEI, Explications de l'évolution de la liste de vérification des documents FALC, 2020. https://www.unapei.org/wp-content/uploads/2020/01/explication_liste_verification_FALC-v2020-01-14.pdf.

(Unia, 2020) UNIA, COVID et droits humains : Impact sur les personnes handicapées et leurs proches, 2020. Résultats de la consultation, https://www.unia.be/files/Resultats_consultation_impact_COVID_sur_les_personnes_handicapees_et_leurs_proches.pdf.

(van Tilborg et Deckers, 2016) Arjan VAN TILBORG et Stijn R.J.M. DECKERS, Vocabulary Selection in AAC : Application of Core Vocabulary in Atypical Populations. *Perspectives of the ASHA Special Interest Groups*, 1(12):125–138, 2016.

(Vandeghinste, 2008) Vincent VANDEGHINSTE, *A Hybrid Modular Machine Translation System. LoRe-MT : Low Resources Machine Translation.* LOT, Utrecht, The Netherlands, 2008.

(Vandeghinste *et al.*, 2021) Vincent VANDEGHINSTE, Adeline MÜLLER, Thomas FRANÇOIS et Orphée DE CLERCQ, Easy Language in Belgium. *In* Camilla LINDHOLM et Ulla VANHATALO, éditeurs, *Handbook of Easy Languages in Europe*, pages 53–90. Frank & Timme, Berlin, 2021.

(Vandeghinste et Schuurman, 2014) Vincent VANDEGHINSTE et Ineke SCHUURMAN, Linking pictographs to synsets : Sclera2Cornetto. *In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, volume 9, pages 3404–3410. European Language Resources Association, 2014.

(Vandeghinste *et al.*, 2017) Vincent VANDEGHINSTE, Ineke SCHUURMAN, Leen SEVENS et Frank VAN EYNDE, Translating text into pictographs. *Natural Language Engineering*, 23(2):217–244, 2017.

(Vaschalde, 2018) Céline VASCHALDE, Génération automatique de pictogrammes à partir de la parole pour faciliter la mise en place d'une communication médiée. Mémoire de master, Université d'Orléans, 2018.

(Vaschalde *et al.*, 2018) Céline VASCHALDE, Pauline TRIAL, Emmanuelle ESPERANÇA-RODIER, Didier SCHWAB et Benjamin LECOUTEUX, Automatic pictogram generation

from speech to help the implementation of a mediated communication. *In Proceedings of the 2nd Swiss Conference on Barrier-free Communication*, pages 97–101, Geneva, Switzerland, 2018.

(Vaswani *et al.*, 2017) Ashish VASWANI, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N. GOMEZ, Łukasz KAISER et Illia POLOSUKHIN, Attention Is All You Need. *In Proceedings of the 31st Conference on Neural Information Processing Systems*, pages 6000–6010, 2017.

(Vauquois, 1968) Bernard VAUQUOIS, A Survey of Formal Grammars and Algorithms for Recognition and Transformation in Machine Translation. *In IFIP Congress*, pages 1114–1122, 1968.

(Vaz, 2013) Irene VAZ, Visual symbols in healthcare settings for children with learning disabilities and autism spectrum disorder. *British Journal of Nursing*, 22(3):156–159, 2013.

(Verborgh et De Wilde, 2013) Ruben VERBORGH et Max DE WILDE, *Using OpenRefine*. Packt Publishing, 2013.

(Vial *et al.*, 2019) Loïc VIAL, Benjamin LECOUTEUX et Didier SCHWAB, Sense Vocabulary Compression through the Semantic Knowledge of WordNet for Neural Word Sense Disambiguation. *In Proceedings of the 10th Global Wordnet Conference*, pages 108–117, Wroclaw, Poland, 2019. Global Wordnet Association.

(Von Tetzchner et Martinsen, 2000) Stephen VON TETZCHNER et Harald MARTINSEN, *Introduction to Augmentative and Alternative Communication*. Whurr Publishers, London, 2000.

(Vossen *et al.*, 2012) Piek VOSSEN, Attila GÖRÖG, Rubén IZQUIERDO et Antal VAN DEN BOSCH, DutchSemCor : Targeting the ideal sense-tagged corpus. *In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 584–589, Istanbul, Turkey, 2012. European Language Resources Association.

(Vossen *et al.*, 2008) Piek VOSSEN, Isa MAKS, Roxane SEGERS et Hennie VAN DER VLIET, Integrating lexical units, synsets and ontology in the Cornetto Database. *In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 1006–1013, Marrakech, Morocco, 2008. European Language Resources Association.

(Wilkens *et al.*, 2022) Rodrigo WILKENS, David ALFTER, Rémi CARDON, Isabelle GRIBOMONT, Adrien BIBAL, Patrick WATRIN, Marie-Catherine DE MARNEFFE et Thomas

FRANÇOIS, CENTAL at TSAR-2022 Shared Task : How Does Context Impact BERT-Generated Substitutions for Lexical Simplification ? *In Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022)*, pages 231–238, Abu Dhabi, United Arab Emirates, 2022.

(Wołk *et al.*, 2017) Krzysztof WOŁK, Agnieszka WOŁK et Wojciech GLINKOWSKI, A Cross-Lingual Mobile Medical Communication System Prototype for Foreigners and Subjects with Speech, Hearing, and Mental Disabilities Based on Pictograms. *Computational and Mathematical Methods in Medicine*, 2017, 2017.

(Woodard et Nelson, 1982) J. P. WOODARD et J. T. NELSON, An information theoretic measure of speech recognition performance. *In Workshop on standardisation for speech I/O technology, Naval Air Development Center, Warminster, PA*, 1982.

(Xu *et al.*, 2016) Wei XU, Courtney NAPOLES, Ellie PAVLICK, Quanze CHEN et Chris CALLISON-BURCH, Optimizing Statistical Machine Translation for Text Simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415, 2016.

(Yaneva *et al.*, 2015) Victoria YANEVA, Irina TEMNIKOVA et Ruslan MITKOV, Accessible Texts for Autism : An Eye-Tracking Study. *In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 49–57, 2015.

(Yaneva *et al.*, 2016) Victoria YANEVA, Irina TEMNIKOVA et Ruslan MITKOV, Evaluating the Readability of Text Simplification Output for Readers with Cognitive Disabilities. *In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 293–299, Portorož, Slovenia, 2016. European Language Resources Association.

(Yusufali *et al.*, 2023) Hussein YUSUFALI, Stefan GOETZE et Roger K. MOORE, Bridging the Communication Rate Gap : Enhancing Text Input for Augmentative and Alternative Communication (AAC). *In International Conference on Human-Computer Interaction*, pages 434–452. Springer, 2023.

(Zabalia *et al.*, 2011) Marc ZABALIA, Lynn M BREAU, Chantal WOOD, Christine LÉVÊQUE, Martine HENNEQUIN, Edith VILLENEUVE, Estelle FALL, Laurent VALLET, Marie-Claude GRÉGOIRE et Genevieve BREAU, Validation francophone de la grille d'évaluation de la douleur-déficience intellectuelle – version postopératoire. *Canadian Journal of Anesthesia / Journal canadien d'anesthésie*, 58(11):1016–1023, 2011.

(Zeng-Treitler *et al.*, 2014) Qing ZENG-TREITLER, Seneca PERRI, Carlos NAKAMURA, Jinqiu KUANG, Brent HILL, Duy Duc An BUI, Gregory J. STODDARD et Bruce E. BRAY, Evaluation of a pictograph enhancement system for patient instruction : A recall study. *Journal of the American Medical Informatics Association*, 21(6):1026–1031, 2014.

(Zhang et Lapata, 2017) Xingxing ZHANG et Mirella LAPATA, Sentence Simplification with Deep Reinforcement Learning. *In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594, Copenhagen, Denmark, 2017. Association for Computational Linguistics.

(Zhao *et al.*, 2023) Wayne Xin ZHAO, Kun ZHOU, Junyi LI, Tianyi TANG, Xiaolei WANG, Yupeng HOU, Yingqian MIN, Beichen ZHANG, Junjie ZHANG, Zican DONG, Yifan DU, Chen YANG, Yushuo CHEN, Zhipeng CHEN, Jinhao JIANG, Ruiyang REN, Yifan LI, Xinyu TANG, Liu ZIKANG, Peiyu LIU, Jian-Yun NIE et Ji-Rong WEN, A Survey of Large Language Models. *arXiv preprint arXiv :2303.18223*, 2023.

(Zheng *et al.*, 2024) Lianmin ZHENG, Wei-Lin CHIANG, Ying SHENG, Siyuan ZHUANG, Zhanghao WU, Yonghao ZHUANG, Zi LIN, Zhuohan LI, Dacheng LI, Eric XING, Hao ZHANG, Joseph GONZALEZ et Ion STOICA, Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. *In Advances in Neural Information Processing Systems*, volume 36, pages 46595–46623. Curran Associates Inc., 2024.

Déclaration de l'auteur

Déclaration attestant le caractère original du travail effectué

J'affirme avoir pris connaissance des documents d'information et de prévention du plagiat émis par l'Université de Genève et la Faculté de traduction et d'interprétation (notamment la Directive en matière de plagiat des étudiant-es, le Règlement d'études de la Faculté de traduction et d'interprétation, ainsi que l'Aide-mémoire à l'intention des étudiants préparant un mémoire de Ma en traduction). J'atteste que ce travail est le fruit d'un travail personnel et a été rédigé de manière autonome. Je déclare que toutes les sources d'information utilisées sont citées de manière complète et précise, y compris les sources sur Internet. Je suis consciente que le fait de ne pas citer une source ou de ne pas la citer correctement est constitutif de plagiat et que le plagiat est considéré comme une faute grave au sein de l'Université, passible de sanctions.

Le travail présenté dans cette thèse a été réalisé dans le cadre d'un doctorat en cotutelle entre l'Université catholique de Louvain (UCLouvain, Belgique) et l'Université de Genève (UNIGE, Suisse). Cette thèse a été réalisée d'octobre 2020 à octobre 2024, sous la direction du Prof. Dr. Thomas François de la Faculté de philosophie, arts et lettres de l'UCLouvain, de la Prof. Dr. Pierrette Bouillon de la Faculté de traduction et d'interprétation de l'UNIGE, ainsi que du Dr. Vincent Vandeghinste de l'Instituut voor de Nederlandse Taal (Pays-Bas) et de la KU Leuven (Belgique). Elle a bénéficié du soutien financier de l'UNIGE et de l'UCLouvain, incluant deux bourses UCL-FSR.

Au vu de ce qui précède, je déclare sur l'honneur que le présent travail est original.

Bruxelles, le 4 novembre 2024

[Magali Norré]