

# Évaluation de la traduction en pictogrammes pour la communication médecin-patient par des adultes avec une déficience intellectuelle

Magali Norré,<sup>1,2</sup> Trang Pham Tran Hanh,<sup>1</sup> Pierrette Bouillon,<sup>2</sup> Vincent Vandeghinste,<sup>3</sup> Thomas François<sup>1</sup>

<sup>1</sup> Centre de Traitement Automatique du Langage, ILC, Université catholique de Louvain, Belgique

<sup>2</sup> Département de Traitement Informatique Multilingue, FTI, Université de Genève, Suisse

<sup>3</sup> Instituut voor de Nederlandse Taal, Pays-Bas & CCL/Leuven.AI, KU Leuven, Belgique

**Résumé**—Cette étude a pour but de déterminer les facteurs qui influencent la compréhension par des personnes avec une Déficience Intellectuelle (DI) d'interactions médicales traduites automatiquement en pictogrammes via deux systèmes : Text-to-Picto et PictoDr. Nous présentons une méthodologie d'évaluation originale. Les personnes avec une DI avaient pour tâches de dire ce qu'ils comprenaient des phrases et des pictogrammes évalués. Nous montrons que plusieurs facteurs influencent la compréhension des – phrases traduites en – pictogrammes. Bien que ces personnes aient souvent des difficultés à communiquer avec un médecin, notre étude montre aussi que recourir uniquement à des phrases en pictogrammes n'est pas une solution pour toutes les personnes avec une DI. Les pictogrammes restent toutefois un moyen utile pour améliorer la communication, notamment dans une approche multimodale.

**Mots clés**—Communication Alternative et Améliorée, CAA, évaluation, traduction médicale, pictogramme, Arasaac, déficience intellectuelle

## I. INTRODUCTION

Pour permettre aux personnes avec une Déficience Intellectuelle (DI) de communiquer, il existe plusieurs formes de Communication Alternative et Améliorée (CAA) [1], [2]. Ces méthodes exploitent divers systèmes de représentation, dont les symboles et les images [3]. Les pictogrammes sont utilisés pour exprimer des besoins basiques de la vie quotidienne, mais aussi dans le cadre des consultations médicales [4], [5]. Différents ensembles d'images sont conçus pour les personnes avec une DI dans le domaine de la santé, par exemple, les pictogrammes Widgit Health<sup>1</sup> ou les images de SantéBD.<sup>2</sup>

Afin d'améliorer la communication entre les médecins et les patients avec une DI, une possibilité est de recourir à des systèmes de traduction automatique vers des pictogrammes, tels qu'Arasaac.<sup>3</sup> Nous proposons donc de comparer la sortie de nos deux systèmes : Text-to-Picto [6], [7] et PictoDr [8], qui diffèrent par leur approche, afin de voir ce qui affectent la compréhension des sorties. Text-to-Picto est une adaptation d'un système basé sur les mots, développé initialement pour les personnes avec une DI et la communication sur Internet [9].

PictoDr est fondé sur des concepts et a été conçu spécifiquement pour la communication médicale, mais pour les patients allophones. Les deux systèmes génèrent donc des traductions différentes (ordre et nombre de pictogrammes différents, etc.). La compréhension de phrases en pictogrammes générées par Text-to-Picto a été évaluée par des migrants [10], ainsi que via un focus group auprès de personnes avec une DI [9], mais il ne s'agissait pas de la version française. La compréhension de pictogrammes Arasaac médicaux a aussi été évaluée, mais avec des personnes allophones, soit via un questionnaire en ligne [11], soit avec des tests utilisateurs et un prototype [12].

Dans cet article, nous présentons une façon originale d'évaluer des phrases médicales traduites automatiquement en pictogrammes Arasaac par Text-to-Picto et PictoDr, deux systèmes que nous avons développés. L'étude a été menée auprès d'adultes avec une DI en Suisse, avec la collaboration d'une facilitatrice, qui travaille au sein d'une association.<sup>4</sup> Dans la suite, nous décrivons la méthodologie (section II) et nos résultats (section III), que nous discutons à la section IV.

## II. MÉTHODOLOGIE

Nous présentons ici une étude pilote destinée à investiguer comment évaluer la compréhension de phrases médicales traduites en pictogrammes 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), 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 [13].

### A. Hypothèses

Les hypothèses peuvent être formulées comme suit :

4. Il s'agit d'une grande organisation d'aide aux personnes en situation de handicap en Suisse. Elle propose un service de traduction de textes en langage simplifié, qui respecte les règles du Facile à Lire et à Comprendre (FALC).

1. [www.widgit-health.com](http://www.widgit-health.com)

2. [www.santebd.org](http://www.santebd.org)

3. [www.arasaac.org](http://www.arasaac.org)

- 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 participants avec un niveau de langue plus élevé.
- H3 : La compréhension des pictogrammes et la compréhension des phrases en pictogrammes est différente avec Text-to-Picto et PictoDr.
- H4 : La compréhension des instructions est meilleure que la compréhension des questions en pictogrammes.
- H5 : Certaines catégories morphosyntaxiques des mots traduits en pictogrammes sont plus compréhensibles que d'autres.

**B. Participants**

Nous avons collaboré avec une facilitatrice suisse qui était chargée d'interroger les participants avec une DI. Elle a l'habitude de travailler avec des personnes en situation de handicap afin de s'assurer que des textes sont faciles à lire et à comprendre. Elle a également déjà participé à une étude de réception portant sur des textes traduits humainement ou automatiquement en langue facile auprès de ce public [14].

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 l'association. Le public cible était constitué de 9 adultes, 5 hommes et 4 femmes, avec une DI (Table I). Nous n'avons pas pu accéder à plus d'informations sur le type de handicap. Le QI n'était pas un critère de sélection des participants. 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 : A1, A2 et B1.<sup>5</sup>

TABLE I  
DESCRIPTION DES PARTICIPANTS

P	Genre	Tranche d'âge	Niveau de français	Déjà utilisé les pictogrammes ?	Temps Min : sec.
P1	H	36-45	A2	Non	22 : 40
P2	F	26-35	B1	Non	18 : 00
P3	F	36-45	B1	Non	24 : 02
P4	H	36-45	B1	Oui	25 : 37
P5	H	26-35	A2	Non	18 : 06
P6	F	26-35	A1	Oui	19 : 55
P7	H	26-35	A1	Oui	19 : 36
P8	H	36-45	A1	Oui	23 : 39
P9	F	26-35	A2	Non	19 : 31

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

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 les images utilisées dans l'étude. Bien que la

5. Le CECR n'est pas un indicateur de la DI. Il est cependant utilisé par l'association pour catégoriser ses groupes de relecteurs FALC.

facilitatrice ne l'ait pas indiqué, nous avons noté lors des entretiens que plusieurs participants semblaient avoir une trisomie. Certains avaient aussi des problèmes de vue et préféraient être plus proches de l'écran ou agrandir les pictogrammes.

Avant chaque test, nous avons demandé le consentement oral puis écrit des participants, via un formulaire adapté en langage simplifié avec quelques pictogrammes. Les tests ont eu lieu dans des locaux de l'association à Fribourg et ont duré environ 30 minutes par participant, incluant l'installation, la contextualisation de l'étude et l'explication des instructions. Pour 7 des tests, le premier auteur a observé et pris des notes supplémentaires, sans intervenir.

**C. Questionnaire**

Pour cette étude, nous avons sélectionné 20 phrases extraites de l'échelle de tri des urgences hospitalières et provenant du système BabelDr [15]. Nous n'avons retenu que des phrases dont les traductions automatiques en pictogrammes Arasaac étaient correctes, afin d'éviter la post-édition. Nous avons utilisé plusieurs critères de sélection en fonction de nos variables d'intérêt (Table II). 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 1) : la traduction automatique de Text-to-Picto et celle de PictoDr. Les phrases doivent avoir une traduction différente dans les deux versions. Le nombre de pictogrammes de la traduction de Text-to-Picto doit 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. Table II).

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 (S) + Verbe (V) + Objet (O) en O ou V + concept « causer » (pictogramme →) + O ou V (variable 2, cf. Figure 1). 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 ». L'auxiliaire « être » ainsi que les semi-auxiliaires modaux « devoir » et « pouvoir » sont 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 pronoms et un pictogramme « médecin » d'Arasaac. Contrairement à Text-to-Picto, PictoDr traduit les expressions multi-mots par un seul pictogramme (variable 6). Quant aux autres mots, soit ils sont non traduits dans PictoDr ou traduits par un pictogramme différent dans les deux versions (variable 7), soit ils sont identiques (variable 8).

La facilitatrice a 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

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

FIGURE 1. Différences de traduction entre Text-to-Picto et PictoDr : exemples, le pictogramme à évaluer obligatoirement par phrase est indiqué en gras

TABLE II  
DIFFÉRENCES DE TRADUCTION ENTRE TEXT-TO-PICTO ET PICTODR

# Variable	Text-to-Picto	PictoDr
1 Nombre de picto	> PictoDr	< Text-to-Picto
2 Ordre des picto	SVO	O   V → O   V
3 Temps	⊕ ⊙   ⊕ ⊙	[non traduit]
4 (Semi-)auxiliaires	picto Arasaac	[non traduit]
5 Pronoms	picto Arasaac	picto Arasaac adapté
6 Expressions multi-mots	2-3 picto	1 picto
7 Autres mots (1)	différent	différent   [non traduit]
8 Autres mots (2)	identique	identique

centre de l'écran. 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), et 5 questions vs. 5 instructions. Les 5 premiers participants ont répondu oralement au premier questionnaire et les 4 derniers, au deuxième questionnaire.

L'ordre des phrases a chaque fois été randomisé 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 même phrase qui ne comptait pas dans les résultats et qui servait d'exemple pour expliquer la tâche aux participants. Nous avons fourni à la facilitatrice les documents pour l'évaluation.

Pour chaque phrase traduite en pictogrammes, 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 l' image-ci »). Au total, la facilitatrice a fait évaluer 180 phrases et 180 pictogrammes (un échantillon aléatoire). En fonction de la réponse au premier point, elle pouvait poser des questions supplémentaires sur les pictogrammes de son choix. Elle devait noter si les réponses étaient correctes ou non. Les paraphrases, reformulations et synonymes étaient acceptés. Des questions sur la communication avec le médecin en général, l'utilité, ainsi que 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 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 faudrait-il améliorer ?

#### D. Annotation

Deux des auteurs, chercheurs en linguistique, 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 [16] 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, soit nous en avons directement post-éditées.

En nous basant sur nos retranscriptions, nous avons déterminé pour les 180 phrases et les 683 pictogrammes si le sens donné par le participant était correct ou non. Nous nous sommes référés autant que possible aux mots de chaque participant ou aux reformulations de la facilitatrice approuvées par le participant. À noter que nous avons préalablement annoté chaque phrase avec nos variables (cf. Table II) et chaque pictogramme avec la catégorie morphosyntaxique du mot source (nom, verbe, adjectif, etc.). L'étiquette « temps » a été ajoutée pour des pictogrammes de Text-to-Picto (variable 3).

Exactitude des pictogrammes	A2			Exactitude des phrases	A2			
	Correct	Incorrect	Total		Correct	Incorrect	Total	
A1	Correct	36 20,0 %	26 14,4 %	62 34,4 %	Correct	36 20,0 %	10 5,5 %	46 25,5 %
	Incorrect	9 5,0 %	109 60,6 %	118 65,6 %	Incorrect	9 5,0 %	125 69,5 %	134 74,5 %
	Total	45 25,0 %	135 75,0 %	180 100 %	Total	45 25,0 %	135 75,0 %	180 100 %

FIGURE 2. Matrices de confusion avec les catégories « Correct » et « Incorrect » (A1) et les catégories « Correct » et « Incorrect/Ne sait pas » (A2) d'exactitude au niveau des pictogrammes (#180) et des phrases (#180)

Nous avons évalué la compréhensibilité à deux niveaux : pour chaque pictogramme et pour chaque phrase. L'évaluation a été réalisée par deux annotateurs : l'auteure principale 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 ». A2 n'avait pas d'expérience avec les pictogrammes. Un accord inter-annotateurs a été calculé avec nos réponses et celles de la facilitatrice par pictogramme (#180) et par phrase (#180). En adaptant son échelle d'annotation à la nôtre (Figure 2), le coefficient  $\kappa$  [17] est de 0,54 pour les pictogrammes et de 0,72 pour les phrases, soit respectivement un accord modéré et un accord substantiel (échelle [18]). Toutes nos analyses statistiques ont été réalisées avec le logiciel R. Pour vérifier nos hypothèses, nous avons utilisé des modèles linéaires généralisés à effets mixtes du package glmmTMB (v.1.1.8) [19]. Dans nos modèles, nous avons toujours inclus un effet aléatoire sur les participants et un effet fixe sur le niveau de langue.

### III. RÉSULTATS

Nous détaillons les résultats quantitatifs sur la compréhension des pictogrammes et la compréhension des phrases (section III-A). Les résultats qualitatifs des questions liées à la communication, la potentielle utilité des pictogrammes et les améliorations seront présentés par après (section III-B).

#### A. Compréhension des pictogrammes et des phrases

En ce qui concerne H1, la Table III 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 sur les échantillons aléatoires de 180 pictogrammes (34,44 % et 25 %), un obligatoire par phrase. Le pourcentage moyen de réponses correctes est plus élevé pour les 180 pictogrammes (29,72 %) que pour les 180 phrases (25,27 %). La différence entre les pourcentages A1A2-Picto (#180) et A1A2-Phrase (#180) – 29,72 % vs. 25,27 % – n'est pas significative en utilisant un t-test païré. Avec les 683 pictogrammes annotés par A1 (42,92 %), la différence entre les pictogrammes et les phrases est très significative ( $p = 0,001011$ ).

TABLE III  
H1 : COMPRÉHENSION DES PICTOGRAMMES ET COMPRÉHENSION DES PHRASES PAR PARTICIPANT SELON DEUX ANNOTATEURS (A1 ET A2)

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 %
#	42,92 %	34,44 %	25 %	25,55 %	25 %
		29,72 %		25,27 %	

TABLE IV  
H2 : COMPRÉHENSION DES PICTOGRAMMES ET COMPRÉHENSION DES PHRASES PAR NIVEAU DE LANGUE SELON UN ANNOTATEUR (A1)

Niveau de langue – P	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 %

Pour H2, la Table IV montre les pourcentages de réponses correctes des pictogrammes et des phrases par niveau de langue. Ceux du niveau B1 sont les plus élevés pour les pictogrammes (55,42 %) et pour les phrases (31,66 %). Nous avons introduit la variable « niveau » dans le modèle mixte. Par rapport à A1, le niveau B1 a un impact positif significatif sur les réponses annotées pour les pictogrammes (coefficient de régression de B1 = 0,99\* ; Odds Ratio (OR) = 2,68 ;  $\chi^2 = (\text{degré de liberté (ddl)} = 2, N = 683) = 6,3^*)^6$ , mais pas significatif pour les phrases.

TABLE V  
H3 : COMPRÉHENSION DES PICTOGRAMMES ET COMPRÉHENSION DES PHRASES PAR TYPE DE TRADUCTION SELON UN ANNOTATEUR (A1)

Type de traduction	A1-Picto (#683)	A1-Phrase (#180)
Text-to-Picto	39,36 %	20,00 %
PictoDr	49,63 %	31,11 %

Pour H3, la Table V présente les pourcentages de réponses correctes des pictogrammes et des phrases par type de traduction. PictoDr a obtenu un score plus élevé pour les pictogrammes (49,63 %) et pour les phrases (31,11 %). Nous avons ajouté la variable « type de traduction » dans le modèle mixte. 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 ;  $\chi^2 = (\text{ddl} = 1, N = 683) = 7,14^{**}$ ), mais pas pour les phrases.

TABLE VI  
H4 : COMPRÉHENSION DES PICTOGRAMMES ET COMPRÉHENSION DES PHRASES PAR TYPE DE PHRASE SELON UN ANNOTATEUR (A1)

Type de phrase	A1-Picto (#683)	A1-Phrase (#180)
Instructions	52,49 %	36,66 %
Questions	34,50 %	14,44 %

Pour H4, la Table VI montre les pourcentages de réponses correctes des pictogrammes et des phrases par type de phrase source pour tous les participants. Les instructions ont un pourcentage plus élevé pour les pictogrammes (52,49 %) et pour les phrases (36,66 %). Nous avons remplacé la variable « type de traduction » par « type de phrase » dans le modèle mixte. Les questions ont un impact négatif extrêmement significatif, par rapport aux instructions, sur la compréhension des pictogrammes (coefficient de régression des questions de

6. Les taux de significativité sont indiqués comme suit : \*  $p < 0,05$  ; \*\*  $p < 0,01$  ; \*\*\*  $p < 0,001$ . Pour les effets fixes, nous avons utilisé le test par défaut du  $\chi^2$  de Wald de type II.

-0,80\*\*\* ; OR = 0,44 ;  $\chi^2 = (\text{ddl} = 1, N = 683) = 24,06***$ ) et des phrases (coefficient des questions de -1,28\*\*\* ; OR = 0,27 ;  $\chi^2 = (\text{ddl} = 1, N = 180) = 11,33***$ ).

TABLE VII  
H5 : COMPRÉHENSION DES PICTOGRAMMES PAR CATÉGORIE MORPHOSYNTAXIQUE SELON UN ANNOTATEUR (A1)

Catégorie	A1-Picto (#683)	Text-to-Picto	PictoDr
Adjectifs	34,28 %	36,36 % †	30,76 %
Adverbes	33,75 %	31,91 %	36,36 % †
Interjections	27,77 %	33,33 % †	22,22 %
Noms	<b>53,84 %</b>	<b>51,35 %</b>	<b>57,14 % †</b>
Pronoms	42,07 %	43,68 % †	40,00 %
Verbes	51,45 %	42,27 %	65,06 % †
(Temps)	00,00 %	00,00 %	-

† = Meilleur résultat entre Text-to-Picto et PictoDr

Pour H5, la Table VII présente les pourcentages de réponses correctes des pictogrammes par catégorie morphosyntaxique. Les noms sont les mieux compris, suivis par les verbes et les pronoms. L'ordre des catégories les mieux reconnues diffère 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 « passé » et « futur » dans Text-to-Picto ne sont pas compris du tout. Nous avons employé la variable « catégorie » dans le modèle mixte. Cette variable est significative au niveau des pictogrammes ( $\chi^2 = (\text{ddl} = 6, N = 683) = 15,79^*$ ).

*B. Communication, utilité des pictogrammes et amélioration*

Nous détaillons ici les réponses aux trois dernières questions des tests. 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 participants ont répondu ne pas avoir de difficultés ou que cela allait (P2 : « non [...] ça va » ; P4 : « non »).

La deuxième question concernait leur ressenti 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... [mais] 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 pistes d'amélioration. Certains participants ont parlé des limites des picto-

grammes 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 eux 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 »). Nous constatons donc un réel besoin que les médecins expliquent mieux, quelle que soit la façon (oral, écrit, plus simple, avec ou sans images).

IV. DISCUSSION

Nous avons vu qu'il y avait bien une différence significative entre la compréhension de tous les pictogrammes et la compréhension des phrases (H1). Plus le niveau de langue était élevé, plus les scores de compréhension des pictogrammes et des phrases étaient élevés (H2). Les pictogrammes et les phrases de PictoDr étaient mieux compris que ceux de Text-to-Picto (H3). Les participants avaient plus facile à comprendre les instructions en pictogrammes que les questions (H4). À notre connaissance, des phrases complètes en pictogrammes ne sont jamais utilisées pour poser des questions.

Enfin, nous avons constaté des différences significatives de compréhension des pictogrammes par catégorie morphosyntaxique représentée (H5). Cette analyse a par exemple permis de montrer que représenter le temps (« passé » et « futur ») par des horloges – comme plusieurs auteurs l'avaient proposé [7], [9] – perturbe plus qu'autre chose les personnes qui n'ont jamais appris à utiliser des pictogrammes. L'apprentissage des images et la possibilité de les enlever dans ces technologies, autrement dit, leur adaptation en fonction des besoins individuels reste essentielle.

Bien que nos participants avec une DI n'utilisaient pas d'images pour communiquer, nous avons vu qu'ils avaient parfois des difficultés à comprendre ce que le médecin leur disait. Certains nous ont dit que les pictogrammes pouvaient être utiles pour d'autres personnes et ont évoqué la piste de la multimodalité (texte + image) pour améliorer la compréhension. Des participants ont également dit avoir préféré les phrases avec moins de pictogrammes. Plusieurs ont montré leur découragement à la vue de longues suites de pictogrammes à interpréter. De plus, nous avons vu que l'explication et la clarté des images doivent être améliorées. Nous travaillons d'ailleurs sur de nouvelles versions de ces systèmes qui prendront en compte nos observations.

Il faut aussi signaler certaines limites de la présente étude. D'une part, les participants retenus n'avaient pas de DI très sévère. Il serait intéressant de reproduire cette étude avec des personnes ayant une DI plus sévère ou celles avec des comorbidités (aphasie, accident vasculaire cérébral, etc.). 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. 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 orthophonistes et ergothérapeutes) réfléchissent ensemble sur le développement ou la validation d'un grand ensemble international de pictogrammes médicaux open source les plus facilement interprétables possible qui puisse être utilisé par tous dans un grand nombre de situations différentes.

## V. CONCLUSION

En conclusion, nous avons proposé une méthodologie originale d'évaluation d'interactions médicales en pictogrammes, qui comprend des entretiens entre une facilitatrice et des personnes avec une DI. Nous avons montré que plusieurs facteurs peuvent influencer la compréhension des – phrases traduites en – pictogrammes : le niveau de représentation (phrase ou pictogramme), le niveau de langue des personnes, la nature de la traduction, le type de phrases à traduire et la catégorie morphosyntaxique du mot à traduire. À notre connaissance, tous ces facteurs n'avaient pas encore été étudiés auprès de personnes avec une DI. Bien qu'elles aient souvent des difficultés à communiquer avec un médecin, recourir uniquement à des phrases traduites en pictogrammes n'apparaît pas comme une solution pour tous les patients avec une DI.

Nous avons également vu qu'il n'était pas toujours facile de déterminer avec certitude si les personnes avec une DI comprenaient ou non les pictogrammes. Néanmoins, les personnes interrogées ont quand même eu plus de difficultés à comprendre le sens global des phrases que le sens de certains pictogrammes individuels. Cette étude met donc 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.

## REMERCIEMENTS

Nous exprimons nos remerciements à tous nos participants. Magali Norré est soutenue par le mandat FSR N° 17005.2022 de l'Université catholique de Louvain (UCLouvain). Ce travail est aussi une partie du projet PROPICTO (www.propicto.unige.ch), financé par le Fonds National Suisse (N° 197864) et l'Agence Nationale de la Recherche (ANR-20-CE93-0005). Il a bénéficié de conseils statistiques de la part de la plateforme technologique de Support en Méthodologie et Calcul statistique de l'UCLouvain, SMCS/LIDAM. Les pictogrammes utilisés sont la propriété du Gouvernement d'Aragon et ont été créés par Sergio Palao pour Arasaac. Le Gouvernement d'Aragon les distribue sous la licence Creative Commons.

## RÉFÉRENCES

- [1] D. R. Beukelman, P. Mirenda. *Communication alternative et améliorée : Aider les enfants et les adultes avec des difficultés de communication*. De Boeck Supérieur, 2017.
- [2] E. Cataix-Nègre. *Communiquer autrement. Accompagner les personnes avec des troubles de la parole ou du langage : Les communications alternatives*. De Boeck Supérieur, 2017.
- [3] C. 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), pp. 84–119, 2016.
- [4] I. Vaz. « Visual symbols in healthcare settings for children with learning disabilities and autism spectrum disorder ». *British Journal of Nursing*, 22(3), pp. 156–159, 2013.
- [5] R. C. Gibson, M. Bouamrane, M. D. Dunlop. « Alternative and Augmentative Communication Technologies for Supporting Adults With Mild Intellectual Disabilities During Clinical Consultations : Scoping Review ». *JMIR Rehabilitation and Assistive Technologies*, 8(2), 2021.
- [6] M. Norré, V. Vandeghinste, P. Bouillon, T. François. « Extending a text-to-pictograph system to French and to Arasaac ». *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pp. 1050–1059, 2021.
- [7] M. Norré, V. Vandeghinste, T. François, P. Bouillon. « Investigating the Medical Coverage of a Translation System into Pictographs for Patients with an Intellectual Disability ». *Proceedings of the Ninth Workshop on Speech and Language Processing for Assistive Technologies*, pp. 44–49, 2022.
- [8] J. Mutal, P. Bouillon, M. Norré, J. Gerlach, L. Ormaechea Grijalba. « A Neural Machine Translation Approach to Translate Text to Pictographs in a Medical Speech Translation System-The BabelDr Use Case ». *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas*, pp. 252–263, 2022.
- [9] L. Sevens. *Words Divide, Pictographs Unite : Pictograph Communication Technologies for People with an Intellectual Disability*. LOT, Utrecht, The Netherlands, 2018.
- [10] B. Bulté, V. Vandeghinste, L. Sevens, I. Schuurman, F. Van Eynde. « Can Pictograph Translation Technologies Facilitate Communication and Integration in Migration Settings ? ». *Computational Linguistics in the Netherlands Journal*, 11, pp. 189–212, 2021.
- [11] M. Norré, P. Bouillon, J. Gerlach, H. Spechbach. « Évaluation de la compréhension de pictogrammes Arasaac et Sclera pour améliorer l'accessibilité du système de traduction médicale BabelDr ». *Actes de la 11e conférence de l'Institut Fédératif de Recherche sur les Aides Techniques pour personnes Handicapées*, pp. 177–182, 2020.
- [12] P. Bouillon, J. Gerlach, M. Norré, H. Spechbach. « Enhancing Speech Translation in Medical Emergencies with Pictographs : BabelDr ». *Translation Technology in Accessible Health Communication, Studies in Natural Language Processing*, Cambridge University Press, pp. 129–151, 2023.
- [13] N. Beail, W. Katie. « Using Qualitative Methods in Research with People Who Have Intellectual Disabilities ». *Journal of Applied Research in Intellectual Disabilities*, 27(2), pp. 85–96, 2014.
- [14] S. Rodríguez Vázquez, P. Bouillon. « A Reception Study of Machine-Translated Easy Language Text by Individuals with Reading Difficulties ». *3rd International Conference on Translation, Interpreting and Cognition*, 2021.
- [15] P. Bouillon, J. Gerlach, J. Mutal, N. Tsourakis, H. Spechbach. « A Speech-enabled Fixed-phrase Translator for Healthcare Accessibility ». *Proceedings of the 1st Workshop on NLP for Positive Impact*, 2021.
- [16] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever. « Robust Speech Recognition via Large-Scale Weak Supervision ». *International Conference on Machine Learning*, PMLR, pp. 28492–28518, 2023.
- [17] J. Cohen. « A coefficient of agreement for nominal scales ». *Educational and psychological measurement*, 20(1), pp. 37–46, 1960.
- [18] J. R. Landis, G. G. Koch. « The Measurement of Observer Agreement for Categorical Data ». *Biometrics*, 33(1), pp. 159–174, 1977.
- [19] M. E. Brooks, K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Mächler, B. M. Bolker. « glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling ». *The R Journal*, 9(2), pp. 378–400, 2017.