

Extending a Text-to-Pictograph System to French and to Arasaac

Magali Norré
CENTAL, UCLouvain,
Belgium
FTI, UNIGE,
Switzerland

Vincent Vandeghinste
Instituut voor de
Nederlandse Taal,
The Netherlands

Pierrette Bouillon
FTI, UNIGE,
Switzerland

Thomas François
CENTAL, UCLouvain,
Belgium

Abstract

We present an adaptation of the Text-to-Picto system, initially designed for Dutch, and extended to English and Spanish. The original system, aimed at people with an intellectual disability, automatically translates text into pictographs (Sclera and Beta). We extend it to French and add a large set of Arasaac pictographs linked to WordNet 3.1. To carry out this adaptation, we automatically link the pictographs and their metadata to synsets of two French WordNets and leverage this information to translate words into pictographs. We automatically and manually evaluate our system with different corpora corresponding to different use cases, including one for medical communication between doctors and patients. The system is also compared to similar systems in other languages.

1 Introduction

Augmentative and Alternative Communication (AAC) is used by disabled people to help them to communicate in daily life and to be more independent in their interactions with others (Beukelman and Mirenda, 1998). As such, AAC technologies also improve the social inclusion of disabled people, including those with an Intellectual Disability (ID), which are the focus of this paper.

One of the characteristics of AAC is to represent the natural language in the form of pictures or pictographs, to support the communication of people with language impairment. Several sets of pictographs have been designed specifically for people with an ID to express their basic needs to their family, friends, teachers, or healthcare professionals. Pictures can be used by these persons for communication in various situations such as for social media (Vandeghinste et al., 2015), access to school (Vaschalde et al., 2018), or in the (pre-)hospital setting (Vaz, 2013; Eadie et al., 2013).

Recent research focuses on applications with pictographs for medical settings and people with or without disabilities such as foreigners and allophone patients. This is the case for the Smartwatch prototype of Wolk et al. (2017) and the BabelDr project (Bouillon et al., 2017; Norré et al., 2021a,b). Current systems are often limited and do not always use NLP techniques, especially the applications available for the general public. My Symptoms Translator (Alvarez, 2014; Alvarez and Fortier, 2014) and MediPicto AP-HP¹ are examples of mobile app for medical communication with images between doctors and patients.

This article focuses on the Text-to-Picto system, which automatically translates text into pictographs for people with an ID (Sevens, 2018; Vandeghinste et al., 2015). The system was originally designed for Dutch, and later extended to English and Spanish (Sevens et al., 2015). In this work, we adapt it to French. In addition, we extend the system by linking it to a third pictograph set, namely Arasaac². Until now, two pictograph sets had been used in Text-to-Picto: Sclera³ and Beta⁴, but adding Arasaac was relevant in view of its growing popularity and coverage (more than 15,000 coloured pictographs, which are specifically designed for people with an ID).

This paper first refers to some related work (Section 2), before introducing the methodology used to adapt Text-to-Picto to French (Section 3). Then, we automatically and manually evaluate the French translation system with the three pictograph sets, using three corpora corresponding to different use cases of AAC. Results are also compared to those of similar systems (Section 4). Finally, we discuss the different evaluations of the system (Section 5).

¹<https://www.aphp.fr/medipicto>

²<https://arasaac.org>

³<http://www.sclera.be>

⁴<http://www.betasymbols.com>

2 Related Work

In this section, we present some work about text-to-pictograph translation systems integrating various NLP techniques for different languages.

In their translation system, [Mihalcea and Leong \(2008\)](#) used the WordNet resource ([Miller, 1995](#)), but without exploiting the relations between concepts. In addition, their system aimed to translate only the content words (nouns and verbs). The Glyph automatically translated patient instructions using NLP (e.g. preprocessing stage, including sentence splitter, word and synonym normalization, etc.), terminology or medication databases, but also computer graphics techniques ([Zeng-Treitler et al., 2014](#); [Bui et al., 2012](#)). This application was not designed or tested with disabled people unlike the work of [Sevens \(2018\)](#) and [Vandeghinste et al. \(2015\)](#) on the Dutch Text-to-Picto system, later extended to English and Spanish in the framework of the Able to Include project. The Text-to-Picto system had a certain success: [Kultsova et al. \(2017\)](#) integrated it in an assistive mobile application for travel and communication in Russian, intended for people with an ID, whereas [Nandy \(2019\)](#) adapted it for Indian languages. Other systems were also developed recently such as AraTraductor, an application for Spanish using NLP to improve its pictograph translations ([Bautista et al., 2017](#)) or the system of [Imam et al. \(2019\)](#) for English, which uses WordNet and ImageNet ([Deng et al., 2009](#)).

For French, [Vaschalde et al. \(2018\)](#) implemented a speech-to-picto tool with an automatic speech recognition module. It includes a system to automatically translate text into Arasaac pictographs, the set of images for AAC used for this study. Based on the work of [Vandeghinste et al. \(2015\)](#), [Vaschalde et al. \(2018\)](#) also evaluated their prototype by testing word sense disambiguation and automatic simplification for the passive structures or the deletion of some grammatical words. All these techniques were already tested by [Sevens \(2018\)](#), but only for translation from Dutch. [Sevens et al. \(2017\)](#) developed a rule-based module for simplification of twelve syntactic phenomena (relative clause, non subject-verb-object order, etc.), including compression. For pictograph translation, [Sevens et al. \(2016\)](#) also used a Dutch word sense disambiguation tool. It is worth mentioning that for the French language, there are not many large resources (e.g. sense-annotated corpus) – compared for example to English – for these NLP tasks.

3 Methodology

In this section, we describe how we automatically linked the pictographs to lexical-semantic resources such as WordNet (Section 3.1). Then, we present the NLP architecture of the Text-to-Picto system (Section 3.2), followed by a description of our use cases (Section 3.3).

3.1 Linking Pictographs to WordNets

In order to convert natural texts into pictographs, the aforementioned systems rely on lexical-semantic resources. The Princeton WordNet or PWN ([Miller, 1995](#)) is one of the largest lexical databases for English. It classifies verbs, nouns, adjectives and adverbs into sets of cognitive synonyms, called synsets, which are linked by semantic relations. Its latest versions are the PWN 3.0 and the PWN 3.1, whose synsets do not have the same numeric identifiers.

For French, we found the WOrdnet Libre du Français or WOLF ([Sagot and Fišer, 2008](#)) and the WoNeF ([Pradet et al., 2014](#)), two automatic translations of PWN 3.0 that differ in the way they were built. In this work, we use the WOLF 1.0b4 (2014) and the three versions of WoNeF 0.1 (2012): coverage (c), fscore (f), and precision (p). The WOLF is considered as the standard French WordNet and is cited more often than WoNeF. Compared to the 117,659 synsets of PWN 3.0, it contains 56,475 synsets with at least one lemma translated into French (see Table 1). As regards the WoNeF, the high coverage version contains 109,447 pairs (literal, synset), the main WoNeF has a F-score of 70.9%, and the high precision version has a precision of 93.3%. In addition, as a result of optimizing the three metrics, the coverage version includes 55,697 synsets, the fscore version has 53,440 and the precision has only 15,482 ([Pradet et al., 2014](#)).

	WOLF 1.0b4	WoNeF 0.1		
		c	f	p
N	42,427	37,685	37,335	10,920
	51.66%	45.89%	45.49%	13.29%
V	5,870	5,772	3,845	1,250
	42.63%	41.92%	27.92%	9.07%
ADJ	6,691	10,238	10,238	2,755
	36.85%	56.38%	56.38%	15.38%
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 1: Number of non-empty synsets of French WordNets and percentage compared to PWN 3.0 per POS.

The Text-to-Picto system was designed to be language independent and easily extensible to other languages. For the Dutch version, Vandeghinste and Schuurman (2014) had manually linked 5,710 Sclera pictographs and 2,760 Beta pictographs to Cornetto WordNet before linking them automatically to PWN 3.0 for English and to MCR WordNet 3.0 for Spanish (Sevens et al., 2015).

To extend the Text-to-Picto system to French and to the Arasaac pictograph set, we could not get access to the links between the PWN 3.0 and 800 Arasaac pictographs of Schwab et al. (2020). We therefore used the Arasaac API⁵ to get JSON data, including the manual links to the PWN 3.1, different pictograph filenames (i.e. the lemmas), and their numeric identifiers (allowing to access the pictograph url). The pictograph filenames are available for at least 30 languages, including English, Spanish, Dutch, Russian, Arabic, etc. The other data of Arasaac API are the same for all languages.

Using OpenRefine (Verborgh and De Wilde, 2013), we cleaned the data: sort, deletion, pre-process duplicate filenames (renamed by adding numbers to distinguish the pictographs in the reference corpora), etc. We automatically linked the Arasaac pictographs associated with one or more synsets of PWN 3.1, WOLF and WoNeF through the PWN 3.0 identifiers and the Collaborative Interlingual Indexes (CILI) available on the GitHub repository of the Global WordNet Association.⁶ It would also be possible to have translations from the Open English WordNet⁷ (McCrae et al., 2020) into Arasaac pictographs. This WordNet and the PWN 3.1 have the same synset identifiers.

For example (see Figure 1), the Arasaac pictograph *docteur* or *médecin* (doctor) has the identifier 2467. It is associated with a PWN 3.1 synset and other information (e.g. one or several tags) that we separated in other tables for future work. By transferring the English synset automatically with CILI, we obtain the PWN 3.0 synset, POS, relation(s) and lemma(s) of French WordNets.⁸ In this case, {*docteur/médecin/toubib*} for WOLF and only {*médecin*} for WoNeF (coverage and fscore versions) in which the lemmas {*docteur*}

and {*toubib*} are linked to another synset. Surprisingly, in the WoNeF (precision), there is no lemma {*docteur*} and {*médecin*} although they are frequent terms for *doctor*, but only the rare term {*toubib*}, linked to two other synsets.

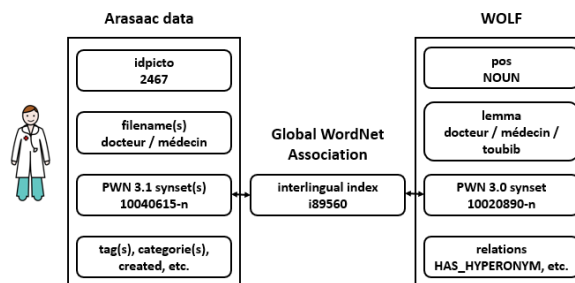


Figure 1: Mapping Arasaac pictographs with French WordNets.

As a result of this process, our Text-to-Picto system is the first that uses synsets of the French WordNets. In contrast, Schwab et al. (2020) directly used the PWN 3.0 and automatically translated the text because they consider the original PWN as the most complete and reliable database. As mentioned before, the automatic translations of WOLF and WoNeF versions differ and are, indeed, less complete compared to PWN 3.0.

3.2 Description of the System

We describe the architecture of our system used to translate a textual input into a sequence of Sclera, Beta, or Arasaac pictographs (see Figure 2).

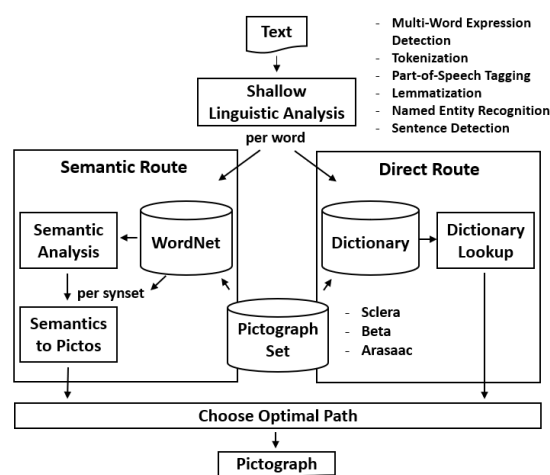


Figure 2: Architecture of the Text-to-Picto system adapted from Vandeghinste et al. (2015).

⁵<https://arasaac.org/developers/api>

⁶<https://github.com/globalwordnet/cili>

⁷<https://github.com/globalwordnet/english-wordnet>; <https://en-word.net/>

⁸We automatically checked the quality of the links between 300 PWN 3.1 synsets of Arasaac pictographs and PWN 3.0 synsets, by matching them to those obtained through the website of PWN 3.1, about 99% of links were correct.

Text Input	<i>Max</i> (Max)	<i>ira</i> (will go)	<i>à</i> (to)	<i>Leuven</i> (Leuven)	<i>l'</i> (the)	<i>été</i> (summer)	,	<i>au revoir</i> (goodbye)
MWE Detection								<i>au_revoir</i>
Tokenization	<i>Max</i>	<i>ira</i>	<i>à</i>	<i>Leuven</i>	<i>l'</i>	<i>été</i>		<i>au_revoir</i>
Part-of-Speech Tagging	NAM	VER:futu	PRP	NAM	DET:ART	NOM	PUN	NOM
Lemmatization	<i>Max</i>	<i>aller</i> (go)	<i>à</i>	<i>Leuven</i>	<i>le</i> (the)	<i>été</i>	,	<unknown>
Named Entity Recognition	(character)			(city)				
Semantic Route – WordNet		00341560-v 00359916-v ...				15237250-n		
Direct Route – Dictionary	(character)		<i>à</i>	(city)	<i>le</i>			<i>au_revoir</i>
Optimal Path and Picto ID	34560	2432	7041	5963	8476	5604		5896
Pictograph Output							,	

Figure 3: Example of French sentence translated into Arasaac pictographs: *Max ira à Leuven l'été, au revoir* (Max will go to Leuven the summer, goodbye).

tion are carried out by TreeTagger (Schmid, 1994); on the other hand, we added detection of Multi-Word Expressions (MWE), processing of specific French phenomena (e.g. elision, negation variants), and simple Named Entity Recognition (NER) based on rules and dictionaries. As in Vaschalde (2018), the named entities detected are substituted by generic placeholders such as *character* or *city*.

For example (see Figure 3), the sentence *Max ira à Leuven l'été, au revoir* (Max will go to Leuven the summer, goodbye) is translated by a sequence of seven pictographs *perso, aller, à, ville, le, été, au_revoir* (character, go, to, city, the, summer, goodbye) after the shallow linguistic analysis. Without the rules of elision and MWE detection, the article *le* (the) would not be translated and the MWE *au revoir* (goodbye) would be incorrectly translated, i.e. by two pictographs: the preposition *au* (at) and the verb *revoir* or *corriger* (revise).

In the next step, two routes are possible depending on the word to translate: the semantic route and the direct route. In the semantic route, each word is looked up in the WordNet database. In case a word is not found, we leverage two WordNet relations – *has_hyperonym* and *near_antonym* – to get substitute translation. For example, there is no pictograph for *saumon* (salmon), the word is therefore translated by its hyperonym *poisson* (fish). The word *infecter* (infect) does not have a pictograph either and is translated by its antonym followed by the negative pictograph, i.e. *désinfecter pas* (desinfect no).

In their annotations for Sclera and Beta sets, Vandeghinste and Schuurman (2014) indicated whether the pictograph is complex or not, i.e. whether it represents several concepts (verb + noun, noun + noun, noun + adjective). For example,

manger un sandwich (eat a sandwich) is translated by the single pictograph *boterham-eten* in Sclera. This filename is linked to the head synset {*manger/alimenter/déjeuner*} (eat/feed/lunch) and to the dependent synset {*sandwich*}. Information about pictographs corresponding to a MWE is missing for Arasaac. It is worth mentioning that inflection is taken into account for MWE annotated with two synsets, unlike the MWE detection used in the shallow linguistic analysis.

For the direct route, we build a dictionary for each of our three pictograph sets for the words not covered by WordNet, i.e. pronouns, prepositions, etc. In Sclera and Beta, the pictographs were linked to their Dutch filename by Vandeghinste et al. (2015). We have then manually translated these two Dutch dictionaries into French. For Arasaac, pictographs were manually linked to French lemmas through their identifiers. A part-of-speech tag were also used to distinguish certain homonyms, e.g. the negative adverb *pas* (not) and the noun *pas* (step). As a result, our dictionaries provide respectively 412, 298 and 420 direct links between pictographs and French tokens or lemmas.

To choose the optimal path while converting a sequence of lemmas to a sequence of pictographs, we use the search algorithm A* described in detail by Vandeghinste et al. (2015). It works with different parameters (i.e. penalties) related to WordNet relations, pictograph features and route preference.

3.3 Use Cases

We briefly describe our three corpora representing several use cases of AAC. They are used for the automated and manual evaluation of our system in which we use different metrics to compare the system's output to a reference translation.

1. The **Email Corpus** (130 sentences), manually translated into Sclera and Beta pictographs by [Sevens \(2018\)](#). The emails, written by people with an ID, their teachers, or their parents, were extracted from the WAI-NOT Belgian website. We manually translated this Dutch corpus into French and Arasaac pictographs. We have slightly pre-edited the reference translations into Sclera and Beta to maintain French word order of our corpus. We did not reproduce the spelling mistakes of people with an ID because we do not evaluate the automated spelling correction.
2. The **Book Corpus** (254 sentences), consisting of six copyright-free children stories manually translated into Arasaac pictographs by [Vaschalde \(2018\)](#). We have slightly pre-edited this French corpus to make it compliant with our evaluation format: e.g., we did not translate plural words twice as is the case in [Vaschalde \(2018\)](#). For example, *les couvertures* (the covers) in its source corpus is manually translated by *le couverture couverture* (the cover cover) in its reference corpus. In our reference corpus, we replaced it by *les couverture* without repetition and keeping the plural for the article because we added the pictograph *les* in our dictionary. We also modified some filenames in their reference translations because we renamed the duplicates when we preprocessed the Arasaac data.
3. The **Medical Corpus** (260 sentences), is a subset from BabelDr, a medical translation system ([Bouillon et al., 2017](#)). These sentences are relatively simple compared to some variations offered by the system. There are mainly questions from doctors to patients, i.e. *pouvez-vous décrire la douleur ?* (can you describe the pain?) or *à combien était votre température la dernière fois que vous l'avez mesurée ?* (what was your temperature the last time you measured it?). There are also patient instructions such as *je vais m'occuper de vous aujourd'hui* (I will take care of you today). As for the Email Corpus, we manually translated the Medical Corpus into Arasaac pictographs. The Figure 4 shows an example of reference translation, the sequence of five filenames: *avoir, vous, des, carie, ?* (have, you, the, cavities, ?).



Figure 4: Example of sentence in Arasaac pictographs from the Medical Corpus: *avez-vous des caries ?* (do you have cavities?).

As regards manual translations, all the words in the source text are translated into pictograph filenames in our reference corpora. However, the process is not a literal translation. We have sometimes translated several words into a single pictograph: e.g. for MWE such as *bouteille de coca* (bottle of coca-cola) or *envoyer une lettre* (send a letter), which can be translated by the complex pictograph *Coca-Cola* or *envoyer_2* in Arasaac (see Figure 5). Some Arasaac pictographs can also have different filenames or meanings, e.g. the pictograph for *mal de tête* (headache) or *faire mal* (hurt).



Figure 5: Example of complex pictographs for *bouteille de coca* (Sclera/Arasaac), *envoyer une lettre* (Sclera/Arasaac) and *mal de tête* or *faire mal* (Arasaac).

4 Results

This section presents how we tuned the system (Section 4.1) and describes the results of the automated evaluation (Section 4.2), followed by the manual evaluation (Section 4.3).

4.1 Tuning the System

For tuning our system, we used 56 sentences sampled from the Email Corpus, that is our development set. We should also stressed that running several times the Text-to-Picto system on the same sentence may yield slightly different translations, even with the same parameters (cf. Section 5). Therefore, each BLEU score has been computed as the average over 10 runs (translations). We also report the standard deviation (SD) over the 10 runs.

We first experimented with the WordNets: the WOLF and the three versions of WoNeF – coverage (c), f-score (f), and precision (p). At this step, as we did not know the optimal parameters for Text-to-Picto yet, we used the best ones reported by [Sevens \(2018\)](#) for Sclera and Beta sets (for Arasaac, we

took Beta’s parameters). The standard BLEU score (Papineni et al., 2002), reported in Table 2, allowed us to choose WOLF as the best French WordNet, as it obtains the highest BLEU scores regardless of the pictograph set. Therefore, WOLF will be used for all our evaluations. These results can be explained because WOLF is often connected to more synsets than the WoNeF; therefore it is more likely that there is a link to the pictograph. For WOLF, the SD is always higher than for WoNeF (SD equals 0 for the high precision version, as its small size makes it less likely to refer to several lemmas).

	Sclera	Beta	Arasaac
WOLF	23.9 (0.36)	25.8 (1.42)	29.4 (1.49)
WoNeF c	13.7 (0.10)	17.5 (0.26)	18.4 (1.30)
WoNeF f	13.7 (0.09)	17.6 (0.19)	18.5 (1.22)
WoNeF p	12.4 (0.00)	12.2 (0.00)	6.5 (0.00)

Table 2: Results of the French Text-to-Picto for WOLF and WoNeF by pictograph set with BLEU metric.

In the next step, we tuned the parameters (cf. Section 3.2) through an automated procedure, using a local hill climbing algorithm (Vandeghinste et al., 2015) with BLEU as the evaluation metric. For each pictograph set – Sclera (S), Beta (B) and Arasaac (A) –, we run five trials of 50 iterations with different random initialisation of the parameters and using a granularity of one, in order to cover different areas of the search space. Finally, we took the best scoring parameter values (see Table 3).

	Min	Max	S	B	A
WOLF relations					
Threshold	5	20	11	12	9
Hyperonym penalty	0	15	13	9	15
Antonym penalty	0	15	6	6	9
Pictograph features					
Wrong number	0	10	10	8	4
No number	0	10	2	3	6
Route preference					
Out-Of-Vocabulary	0	10	5	2	2
Direct route advantage	0	15	12	5	12

Table 3: Results of the parameter tuning of French Text-to-Picto for Sclera, Beta and Arasaac pictograph sets.

4.2 Automated Evaluation

We automatically evaluated the performance of our French Text-to-Picto system on the Email Corpus, the Book Corpus and the Medical Corpus. Different experimental conditions were tested, progressively activating more features of the system: a) only with dictionary, b) with dictionary and synonyms of WOLF, c) with dictionary, synonyms

and other relations of WOLF (i.e. hyperonyms, antonyms). In addition, we compared our results with those of the Dutch Text-to-Picto system (Sevens, 2018) and those of the French system of Vaschalde et al. (2018), when available. Such comparisons should be taken with caution, as the experiments are not strictly comparable.

For a better comparison with these studies, we reused the metrics of Sevens (2018) and Vandeghinste et al. (2015): the BLEU (Papineni et al., 2002),⁹ the Word Error Rate (WER) and the Position-independent word Error Rate (PER). BLEU is designed for machine translation, while WER and PER are used in speech recognition. All of these metrics compare the system output to one or more reference translations. In our case, we used sentences translated manually into pictograph filenames (cf. Section 3.3). As described in Section 4.1, each evaluation metric was estimated based on an average over 10 runs of the system. Standard deviations are reported in brackets in the tables.¹⁰

	BLEU	WER	PER
Sclera			
Dictionary	12.0 (0.0)	58.7 (0.0)	55.8 (0.0)
(Sevens, 2018)	14.1	71.9	65.8
+ Synonyms	17.8 (0.4)	56.2 (0.4)	50.3 (0.4)
(Sevens, 2018)	16.5	67.5	60.5
+ Relations	17.9 (0.4)	56.2 (0.3)	50.8 (0.3)
(Sevens, 2018)	16.1	68.7	61.3
Beta			
Dictionary	10.9 (0.0)	63.0 (0.0)	62.1 (0.0)
(Sevens, 2018)	16.9	63.4	53.7
+ Synonyms	21.6 (1.4)	57.5 (1.1)	52.1 (1.1)
(Sevens, 2018)	23.0	52.4	43.3
+ Relations	22.4 (1.2)	57.9 (0.5)	52.6 (0.6)
(Sevens, 2018)	25.9	51.2	42.0
Arasaac			
Dictionary	7.3 (0.0)	59.3 (0.0)	58.8 (0.0)
(Sevens, 2018)	–	–	–
+ Synonyms	24.8 (0.7)	68.1 (1.5)	57.6 (1.4)
(Sevens, 2018)	–	–	–
+ Relations	24.9 (0.8)	68.0 (1.9)	57.2 (1.7)
(Sevens, 2018)	–	–	–

Table 4: Results of the French Text-to-Picto and Dutch Text-to-Picto on Email Corpus by pictograph set with BLEU, WER and PER metrics.

We compared our results with those of Sevens (2018) for Dutch on the same 84 sentences from the Email Corpus (see Table 4). As regards BLEU scores, we got very comparable results, especially

⁹The standard BLEU based on average of 1-grams, 2-grams, 3-grams, and 4-grams (mteval-v011b.pl on <https://github.com/moses-smt/mosesdecoder>).

¹⁰For the first condition (a), the SD is always 0 because, as without WordNet, there is no variation in the results.

for Sclera. For Beta, our scores are lower than those of [Sevens \(2018\)](#). Arasaac obtains the best BLEU scores, except for the first condition (a). This can be explained by the fact that dictionaries for Sclera and Beta are more suitable for this task. They were built from frequent untranslated words from the Email Corpus. For the conditions (b) and (c), differences does not seem meaningful.

For the Book Corpus, as proper names occur frequently, we added one condition: d) without the step of Named Entity Recognition (NER), based on the work of [Vaschalde \(2018\)](#). Results are reported at Table 5. Our BLEU scores for conditions (b) (28.0) and (c) (28.3) are in line with those of [Vaschalde et al. \(2018\)](#). Their translation system from French into Arasaac pictographs obtains a BLEU score of 25.45 when all the words are translated, without word sense disambiguation and without a specific treatment for plurals (26.65 with it). The rather decent score of the dictionary condition (a) on this corpus – in contrast with its performance on the two other use cases – can probably be explained by the effect of the NER module. Indeed, the test set includes 152 occurrences of proper names and we observe that, without NER module (d), scores drop a lot.

	BLEU	WER	PER
Dictionary	18.0 (0.0)	49.8 (0.0)	48.5 (0.0)
+ Synonyms	28.0 (0.4)	58.0 (0.9)	49.8 (0.8)
+ Relations	28.3 (0.6)	57.7 (0.7)	49.3 (0.6)
- NER	17.4 (0.7)	68.1 (1.1)	57.1 (0.8)

Table 5: Results of the French Text-to-Picto on Book Corpus for Arasaac pictograph set with BLEU, WER and PER metrics.

	BLEU	WER	PER
Dictionary	8.6 (0.0)	52.8 (0.0)	52.0 (0.0)
+ Synonyms	31.1 (0.7)	51.4 (0.8)	46.3 (0.5)
+ Relations	31.3 (0.7)	51.1 (1.2)	46.1 (1.2)

Table 6: Results of the French Text-to-Picto on Medical Corpus for Arasaac pictograph set with BLEU, WER and PER metrics.

Unlike the Book Corpus, there is no named entity in the Medical Corpus from the Arasaac dictionary (a). The BLEU scores (see Table 6) are higher than for the other corpora when we use the synonyms and relations of WOLF (b) and (c). The WER and PER metrics are also better than those obtained with the two other corpora. For PER, this may be due to a higher similarity of easily translatable syntactic structures (e.g. do you have...?, etc.).

4.3 Manual Evaluation

We also carried out a manual evaluation of one automatic translation of the 260 sentences of the Medical Corpus into Arasaac pictographs generated by our tuned system (i.e. the WOLF and the other parameters). For each of the translated words, a judge checked whether the pictograph generated was a coherent semantic representation of the word, in order to calculate the precision. She removed untranslated words, in order to calculate the recall. The system reached a precision of 83.7%, a recall of 90.14%, and a F-score of 86.92% (see Table 7).

	P	R	F1
Sclera (Sevens, 2018)			
Email – Dutch	89.24%	86.23%	87.71%
Email – English	93.30%	73.04%	81.94%
Email – Spanish	93.31%	83.14%	87.93%
Beta (Sevens, 2018)			
Email – Dutch	85.91%	89.45%	87.64%
Email – English	82.56%	86.14%	84.31%
Email – Spanish	94.64%	86.83%	90.57%
Arasaac			
Medical – French	83.70%	90.14%	86.92%

Table 7: Results of the French Text-to-Picto on Medical Corpus and Dutch/English/Spanish Text-to-Picto on Email Corpus by pictograph set with Precision, Recall and F-score metrics.

By comparison, [Sevens \(2018\)](#) obtained on its Email Corpus a F-score between 81-85% for English-to-Sclera/Beta system and 87-91% for Dutch/Spanish-to-Sclera/Beta (without counting proper names). Our French system obtains a higher recall than precision and recall of other linguistic versions. In our manual evaluation, we assume that all words must be translated.

5 Discussion

As regards the efficiency of the different pictograph sets, tested on the Email Corpus, we see that our results are better for the Beta set than for the Sclera set. [Sevens \(2018\)](#) explained that Beta contains less pictographs than Sclera. As a result, more paraphrasing translations are possible in Sclera, resulting in a less accurate measurement of translation quality by BLEU. However, our scores for Arasaac, the largest set of pictograph, are the best. Compared to others, this set includes more function words (articles, prepositions, etc.) that have been encoded in our dictionary. Therefore, they will always be well translated, which improves the results. It should also be mentioned that, despite

relatively good results, some pictographs generated by our system are not also easily comprehensible, depending on the context of the sentence. This is especially the case for function words and pain description on a specific body part. In addition, medical words are not always translated because there is no corresponding pictographs, e.g. *symptôme* (symptom), *cancer* (tumor), etc.

Our experiments also aimed at assessing the contribution of different components to our Text-to-Picto system for French. Using only the dictionary clearly yields unsatisfactory results. The system improves when we add the synonym or the relation component, regardless of the pictograph sets and corpora. However, the difference between both components appears marginal. In our reference corpora, a manual inspection of the data reveals that the relation of synonymy is much more frequent than those of hyperonymy and antonymy – which is very rare –, especially for the largest Arasaac pictograph set. Translating text into pictographs is a meticulous and time-intensive process (Sevens et al., 2016). This explains why the corpora are small. It is worth mentioning that the BLEU score is very dependent on the reference translation, which may be partially subjective.

Finally, as explained above, every time we run the French system with the same parameters, we get slightly different translations. This happens when the optimal path calculation step has to choose randomly between several pictographs that have an equal weight, using WordNet. For future work, it would be possible to associate the pictographs with frequency information to regulate this issue. Some studies (Imam et al., 2019; Sevens, 2018; Sevens et al., 2016) also showed that word sense disambiguation improves the results of text-to-picto systems. This would avoid translation errors related to homonyms identified in our manual evaluation, e.g. *enceinte* (to be pregnant or a speaker) and *bleu* (the colour or a bruise).

6 Conclusion

We presented the French version of the Text-to-Picto system, which automatically translates a textual input into pictographs for people with an ID. Our experiments show that this system is easily extensible to other natural or pictograph languages.¹¹

¹¹The source code of the system and the French corpora used for evaluation will be made available for the research community at the following address: <https://github.com/VincentCCL/Picto>.

Compared to the Dutch version, we adapted the shallow linguistic analysis by adding new steps (detection of MWE, preprocessing of specific phenomena, and simple NER). Data cleaning was performed to link the Arasaac pictographs to French semantic resources. The evaluations on the Email and the Book Corpus with WOLF show that our results are indeed in line with those of previous studies. However, there is room for further improvement, for instance adding a word sense disambiguation step to select the right pictograph for a given meaning. We also carry out automated and manual evaluations on a new use case: medical data, which raised new challenges related to the translation of technical terms. We have seen above that our system currently tends to poorly handle technical terms, often missing from WOLF. We plan to investigate solutions to this limitation, for example by applying automatic text simplification for the medical domain (Cardon and Grabar, 2020) on the original sentences.

We also plan to run tests with target users to tune this Text-to-Picto system for medical communication between doctors and patients in the hospital setting. The Dutch version of the presented system has already been tested in real situations with a focus group of five adults with an ID and two coaches in a day centre in Belgium (Sevens, 2018).

Acknowledgments

This research was funded by the UCL-FSR mandate N°13936.2020. This work is also part of the PROPICTO project, funded by the Fonds National Suisse (N°197864) and the Agence Nationale de la Recherche (ANR-20-CE93-0005). The pictographs used are property of the Aragon Government and have been created by Sergio Palao to Arasaac (<http://arasaac.org>). Aragon Government distributes them under Creative Commons License (BY-NC-SA). The other pictographs used are property of Sclera vzw (<https://www.sclera.be/>) which distributes them under Creative Commons License.

References

- Juliana Alvarez. 2014. Visual design. A step towards multicultural health care. *Arch Argent Pediatr*, 112(1):33–40.
- Juliana Alvarez and Simon Marino Fortier. 2014. Two way communication: How to test both sides of an emergency tool. In *2014 IEEE Healthcare Innovation Conference (HIC)*, pages 145–148. IEEE.

- Susana Bautista, Raquel Hervás, Agustín Hernández-Gil, Carlos Martínez-Díaz, Sergio Pascua, and Pablo Gervás. 2017. Aratrador: text to pictogram translation using natural language processing techniques. In *Proceedings of the XVIII International Conference on Human Computer Interaction*, pages 1–8.
- David R. Beukelman and Pat Mirenda. 1998. *Augmentative and alternative communication*. Paul H. Brookes Baltimore.
- Pierrette Bouillon, Johanna Gerlach, Hervé Spechbach, Nikolaos Tsourakis, and Sonia Halimi. 2017. BabelDr vs Google Translate: A user study at Geneva University Hospitals (HUG). In *20th Annual Conference of the European Association for Machine Translation (EAMT)*.
- Duy Duc An Bui, Carlos Nakamura, Bruce E. Bray, and Qing Zeng-Treitler. 2012. Automated illustration of patients instructions. In *AMIA Annual Symposium Proceedings*, volume 2012, page 1158. American Medical Informatics Association.
- Rémi Cardon and Natalia Grabar. 2020. French biomedical text simplification: When small and precise helps. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 710–716.
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255.
- Kathy Eadie, Marissa J. Carlyon, Joanne Stephens, and Matthew D. Wilson. 2013. Communicating in the pre-hospital emergency environment. *Australian Health Review*, 37(2):140–146.
- Mai Farag Imam, Amal Elsayed Aboutabl, and Ensaf H. Mohamed. 2019. Automating text simplification using pictographs for people with language deficits. *I.J. Information Technology and Computer Science*, 9(1):26–34.
- Marina Kultsova, Dmitry Matyushechkin, Andrey Usov, Svetlana Karpova, and Roman Romanenko. 2017. Generation of pictograph sequences from the Russian text in the assistive mobile application for people with intellectual and developmental disabilities. In *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–4. IEEE.
- John Philip McCrae, Alexandre Rademaker, Ewa Rudnicka, and Francis Bond. 2020. 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.
- Rada Mihalcea and Chee Wee Leong. 2008. Toward communicating simple sentences using pictorial representations. *Machine translation*, 22(3):153–173.
- George A. Miller. 1995. WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41.
- Ankita Nandy. 2019. Beyond words: Pictograms for Indian languages. *International Journal of Research in Science and Technology*, 9(1):19–25.
- Magali Norré, Pierrette Bouillon, Johanna Gerlach, and Hervé Spechbach. 2021a. 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 AFIA/TLH-ATALA : La santé et le langage*.
- Magali Norré, Pierrette Bouillon, Johanna Gerlach, and Hervé Spechbach. 2021b. 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 (BfC 2020)*, pages 55–63. ZHAW Zurich University of Applied Sciences.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. 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.
- Quentin Pradet, Gaël De Chalendar, and Jeanne Baguenier Desormeaux. 2014. WoNeF, an improved, expanded and evaluated automatic French translation of WordNet. In *Proceedings of the Seventh Global Wordnet Conference*, pages 32–39.
- Benoît Sagot and Darja Fišer. 2008. Building a free French WordNet from multilingual resources. In *OntoLex*, Marrakech, Morocco.
- Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *New methods in language processing*, page 154.
- Didier Schwab, Pauline Trial, Céline Vaschalde, Loïc Vial, Emmanuelle Esperança-Rodier, and Benjamin Lecouteux. 2020. Providing semantic knowledge to a set of pictograms for people with disabilities: a set of links between WordNet and Arasaac: Arasaac-WN. In *LREC*.
- Leen Sevens. 2018. *Words Divide, Pictographs Unite: Pictograph Communication Technologies for People with an Intellectual Disability*. LOT, JK Utrecht, The Netherlands.
- Leen Sevens, Gilles Jacobs, Vincent Vandeghinste, Ineke Schuurman, and Frank Van Eynde. 2016. Improving text-to-pictograph translation through word sense disambiguation. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 131–135.
- Leen Sevens, Vincent Vandeghinste, Ineke Schuurman, and Frank Van Eynde. 2015. 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.

Leen Sevens, Vincent Vandeghinste, Ineke Schuurman, and Frank Van Eynde. 2017. 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.

Vincent Vandeghinste and Ineke Schuurman. 2014. Linking pictographs to synsets: Sclera2Cornetto. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, volume 9, pages 3404–3410. ELRA, Paris.

Vincent Vandeghinste, Ineke Schuurman, Leen Sevens, and Frank Van Eynde. 2015. Translating text into pictographs. *Natural Language Engineering*, 23(2):217–244.

Céline Vaschalde. 2018. Génération automatique de pictogrammes à partir de la parole pour faciliter la mise en place d'une communication médiée. Master's thesis, Université d'Orléans.

Céline Vaschalde, Pauline Trial, Emmanuelle Esperança-Rodier, Didier Schwab, and Benjamin Lecouteux. 2018. Automatic pictogram generation from speech to help the implementation of a mediated communication. In *Conference on Barrier-free Communication*.

Irene Vaz. 2013. Visual symbols in healthcare settings for children with learning disabilities and autism spectrum disorder. *British Journal of Nursing*, 22(3):156–159.

Ruben Verborgh and Max De Wilde. 2013. *Using OpenRefine*. Packt Publishing Ltd.

Krzysztof Wołk, Agnieszka Wołk, and Wojciech Glinkowski. 2017. 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.

Qing Zeng-Treitler, Seneca Perri, Carlos Nakamura, Jinqiu Kuang, Brent Hill, Duy Duc An Bui, Gregory J. Stoddard, and Bruce E. Bray. 2014. Evaluation of a pictograph enhancement system for patient instruction: a recall study. *Journal of the American Medical Informatics Association*, 21(6):1026–1031.