Experiments for the adaptation of Text2Picto to French

Norré, Magali; Vandeghinste, Vincent; Bouillon, Pierrette; François, Thomas

Experiments for the adaptation of Text2Picto to French

Magali Norré [1 2] Vincent Vandeghinste [3] Pierrette Bouillon [2] Thomas François [1]

[1] Université catholique de Louvain [2] Université de Genève [3] Instituut voor de Nederlandse Taal

Context

Augmentative and Alternative Communication for disabled people
(Beukelman and Mirenda 1998)

Text-to-Pictograph systems
Imam et al. 2019, Vaschalde et al. 2018, Vandeghinste et al. 2015

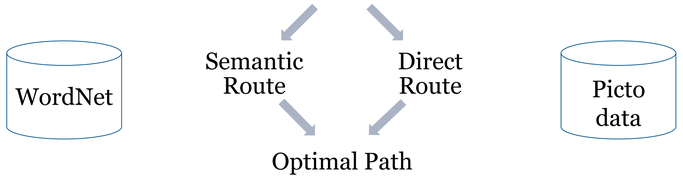
Text2Picto

Translation system for people with an intellectual **disability**

- from natural languages: Dutch | English | Spanish | **French**
  e.g. text: "Nicolas va à Anvers l'été au revoir"

Multi-Word Expression Detection, Tokenization, POS Tagging, Lemmatization, Named Entity Recognition, Sentence Detection

Shallow Linguistic Analysis



- into pictograph languages: Sclera | Beta | **Arasaac**

(perso | aller | à | ville | le | été | au_revoir)
(character | go | to | city | the | summer | bye bye)

- use cases:

1. Email Corpus (Sevens 2018)

"Où habites-tu ?" (Where do you live?)

2. Book Corpus (Vaschalde 2018)

"Abel et Bellina ne disent rien"
(Abel et Bellina say nothing)

3. Medical Corpus (Bouillon et al. 2017)

"Avez-vous des caries ?"
(Do you have cavities?)

Objectives

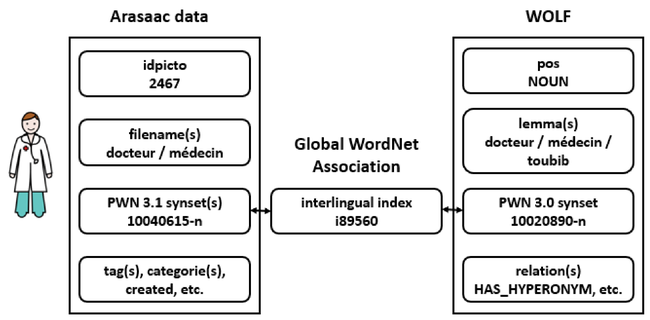
Adapt the Text2Picto system to French

Extend it to Arasaac pictographs

Compare Text2Picto to other studies

Methodology

Linking **pictographs** to **semantic resources**



- Pictographs: **Arasaac data** (JSON API)
- Collaborative Interlingual Index: Princeton WordNet 3.1 <-> 3.0
- Semantic resources: **French WordNets**

✓ **WOLF**
(Sagot and Fišer 2008)

✓ **WoNeF**
(Pradet et al. 2014)
coverage | fscore | precision versions

System Tuning

- Which **WordNet**?

	Sclera	Beta	Arasaac
WOLF	**23.93**	**25.89**	**29.48**
WoNeF coverage	13.71	17.55	18.41
WoNeF fscore	13.77	17.69	18.52
WoNeF precision	12.48	12.29	6.52

Table 1: Results of the French Text-to-Picto for WOLF and WoNeF by pictograph set with BLEU metric.

- Which **parameters**?

Optimization with a local hill climbing algorithm for each pictograph set (Vandeghinste et al. 2015)

Automated Evaluation

Email Corpus

	BLEU	NIST	WER	PER
Sclera				
Dictionary	12.05	2.02	58.79	55.86
(Sevens, 2018)	14.17	3.68	71.96	65.88
+ Synonyms	17.87	**4.00**	**56.21**	**50.34**
(Sevens, 2018)	16.55	3.97	67.54	60.50
+ Relations	**17.90**	**4.00**	56.24	50.85
(Sevens, 2018)	16.12	3.96	68.78	61.33
Beta				
Dictionary	10.97	1.00	63.02	62.11
(Sevens, 2018)	16.98	4.43	63.44	53.77
+ Synonyms	21.69	4.22	**57.56**	**52.15**
(Sevens, 2018)	23.01	5.00	52.42	43.31
+ Relations	**22.40**	**4.29**	57.93	52.61
(Sevens, 2018)	25.91	5.17	51.29	42.07
Arasaac				
Dictionary	7.38	0.41	**59.34**	58.85
+ Synonyms	24.88	4.41	68.14	57.65
+ Relations	**24.97**	**4.44**	68.00	**57.29**

Table 2: Results of the French Text-to-Picto and Dutch Text-to-Picto on Email Corpus by pictograph set with BLEU, NIST, WER and PER metrics.

Book Corpus

	BLEU	NIST	WER	PER
Dictionary	18.00	1.93	**49.82**	**48.57**
+ Synonyms	28.07	5.26	58.04	49.82
+ Relations	**28.34**	**5.28**	57.71	49.37
- Named entities	17.45	4.79	68.18	57.19
(Vaschalde, 2018)	25.45	–	–	–

Table 3: Results of the French Text-to-Picto compared to another system on Book Corpus for Arasaac pictograph set with BLEU, NIST, WER and PER metrics.

Medical Corpus

	BLEU	NIST	WER	PER
Dictionary	8.66	1.02	52.80	52.02
+ Synonyms	31.19	5.24	51.41	46.39
+ Relations	**31.36**	**5.26**	**51.13**	**46.12**

Table 4: Results of the French Text-to-Picto on Medical Corpus for Arasaac pictograph set with BLEU, NIST, WER and PER metrics.

Manual Evaluation

	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 5: 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.

Conclusion

- Results in line with those of other studies
- But only one reference translation possible into pictographs?
- System easily extensible to other (natural/pictograph) languages
- Room for improvement: user tests, automatic simplification, WSD