Evaluating the comprehension of Arasaac and Sclera pictographs for the BabelDr patient response interface

Norré, Magali; Bouillon, Pierrette; Gerlach, Johanna; Spechbach, Hervé

Jekat, Susanne J., Steffen Puhl, Luisa Carrer, and Alexa Lintner, eds. 2021. *Proceedings of the 3rd Swiss Conference on Barrier-free Communication (BfC 2020), Winterthur (online), June 29–July 4, 2020.* Winterthur: ZHAW Zurich University of Applied Sciences. DOI: 10.21256/zhaw-3001.

Magali Norré, Pierrette Bouillon, Johanna Gerlach & Hervé Spechbach

Evaluating the comprehension of Arasaac and Sclera pictographs for the BabelDr patient response interface

Research track

Abstract

This paper summarises some of the findings from a preliminary survey conducted to evaluate the comprehension of pictographs for patient responses in the speech-to-speech translation tool BabelDr, a system designed to improve communication between doctors and allophone patients or minority groups. Despite the relatively low number of respondents, the gathered data could serve as a starting point for discussion in future decision-making processes about how to design a bidirectional interface for patients with a range of pictographs and how to evaluate their comprehension.

1 Introduction

The BabelDr project is a collaboration between the Faculty of Translation and Interpreting (FTI) in Geneva and the Geneva University Hospitals (Bouillon et al. 2017). The aim of the project is to design a reliable translation system for emergency settings to improve communication between doctors and allophone patients (e.g. refugees) or minority groups (sign language users). Until recently, BabelDr had a unidirectional interface. The doctor had to ask closed questions and the patient must then respond nonverbally with a gesture or by pointing. The aim of this paper is to evaluate different open source pictograph sets that could be used for allowing patients to answer more precisely to the doctors' questions. This bidirectional version could improve the communication and possibly reduce doctors' feelings of being constrained by the unidirectional version (Spechbach et al. 2019). In particular, it would allow doctors to ask open questions which is more natural in the diagnostic task and is known to encourage patients to report any and all problems. It could also save time, reducing the number of necessary questions for the diagnosis (one open question replacing multiple yes/no questions).

Even if pictures are used in medical settings to communicate with patients with special needs (Eadie et al. 2013), online medical applications which use pictographs are very limited and remain technically unsophisticated (Wołk et al. 2017). In the Medipicto AP-HP mobile application, the patient chooses pictographs labelled in his language to communicate with the caregiver who can ask questions by choosing a pictograph translated into the patient's and caregiver's languages from a predefined list. The caregiver is limited in terms of questions and this application does not offer speech functionalities (recognition or synthesis), as opposed to BabelDr. Subject to certain conditions, medical images and pictographs are available, e.g. SantéBD, Widgit Health or a graphic chart for symptoms (Alvarez 2011). There are also different pictograph sets that are designed for augmentative and alternative communication used by people with disabilities (Cataix-Nègre 2017; Beukelman and Mirenda 1998).

In this study, we describe the result of the survey set up to evaluate two large pictograph sets that are freely available to represent patient responses: Arasaac and Sclera. These two sets were already used in different pictograph-based systems in other domains, in particular Schwab et al. (2020) build a semantic resource to e.g. design a

translation system from speech into Arasaac pictographs. Sevens (2018), Vandeghinste and Schuurman (2014) used the Sclera set in their Text-to-Picto and Picto-to-Text systems for people with an intellectual disability. Although pictographs were already evaluated in some studies (Wołk et al 2017), this is the first study which evaluates pictographs in the context of a speech application for diagnostic interviews.

In Section 2, we present in more detail BabelDr and the new bidirectional interface which includes pictographs for patient responses. Section 3 focuses on the design and implementation of the questionnaire (available here). Section 4 summarises the findings and results by question types (4.1), response groups (4.2), and pictograph sets: Arasaac or Sclera (4.3). Finally, Section 5 draws conclusions and describes our plans for future developments.

2 The bidirectional version of BabelDr

BabelDr is an online speech-enabled fixed-phrase translator, specifically designed for medical dialogue. Similar to other fixed-phrase translators (such as Medibabble or UniversalDoctor), the system relies on a predefined list of human-translated sentences in order to insure translation reliability, but instead of searching for sentences in this list with keywords, doctors can ask their questions orally, which improves the ergonomics (Spechbach et al. 2019). The doctor speaks freely and the system links the recognition result to the closest human-translated sentence using neural methods (Mutal et al. 2019). At the time of writing, the BabelDr platform is accessible online at the address http://babeldr.unige.ch and is in use at Geneva University Hospitals (HUG) for translation between French and six migrant languages, with a high satisfaction from doctors and patients (Janakiram et al. 2020).

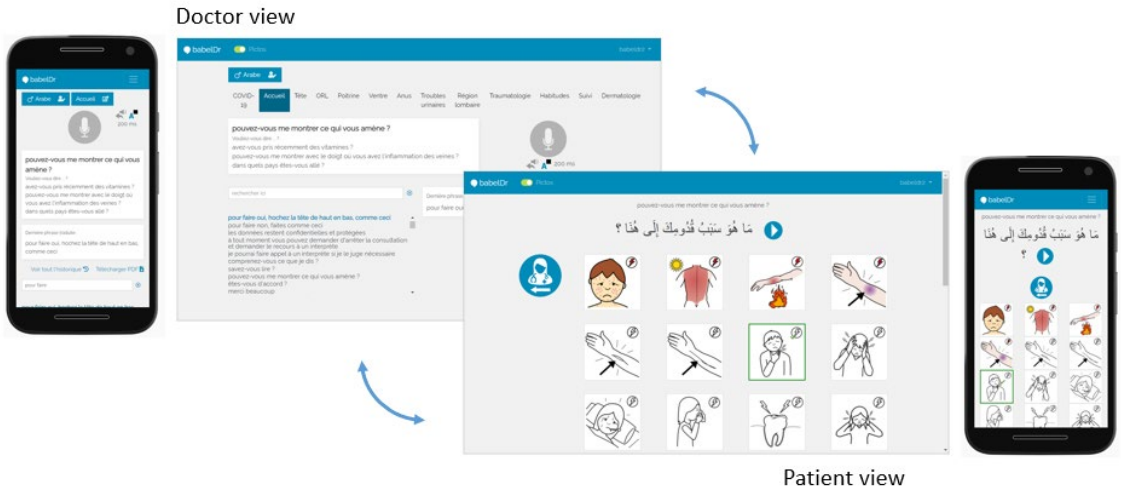

Figure 1. BabelDr bidirectional interface

The bidirectional interface includes two different views, one for the doctor and one for the patient (see Figure 1). The doctor view allows doctors to ask questions orally or to search for questions in a list with keywords. When the doctor confirms the recognition result after speaking, or picks a question in the list, the system switches to the patient view and speaks the question for the patient in the target language. The patient view presents a selection of clickable response pictographs corresponding to the question, among which the patient can select his answer. All questions and answers are saved for

the doctor. If necessary, the doctor can ask a new question in order to confirm the patient's answer.

One of the aims of BabelDr is to make its content easily expandable in order to follow demographics. An online interface allows doctors or developers to link BabelDr questions with different sets of pictographs, as shown in Figure 2. This allows to easily integrate different sets of pictographs in the system depending on the needs. Evaluation can also be done directly on the task.

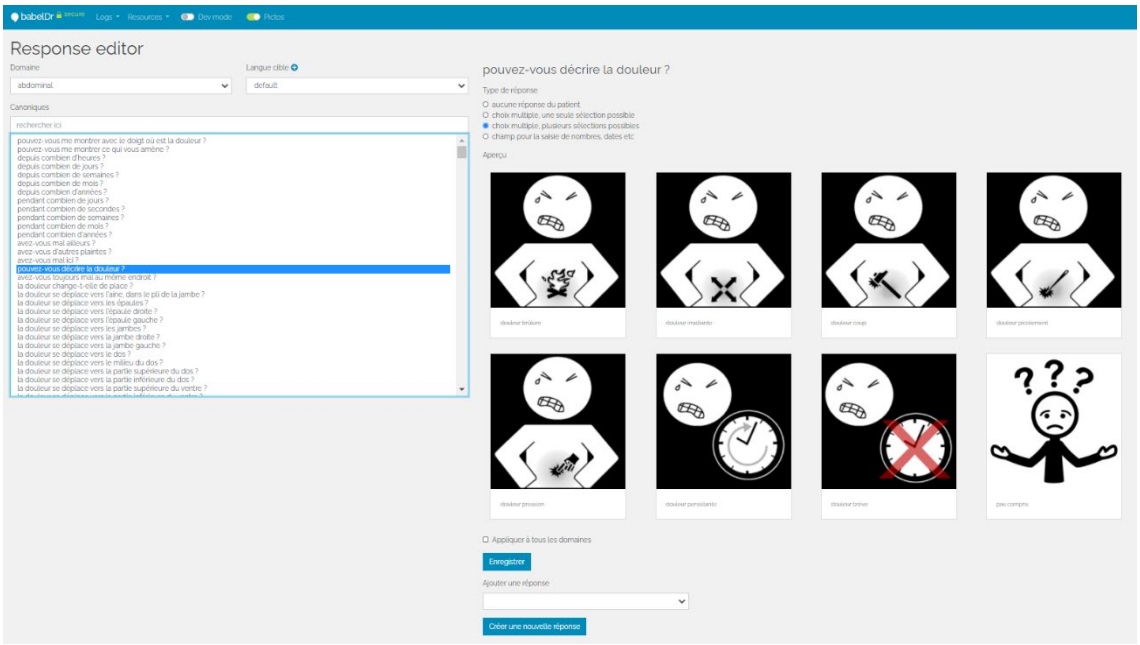

Figure 2. BabelDr response editor

3 Questionnaire design and implementation

This preliminary study aimed (1) to evaluate the comprehension of different sets of pictographs for patient responses in the BabelDr bidirectional interface and (2) to investigate how comprehension can be assessed. The survey was launched in three languages (French, Spanish and Arabic) in order to get the most diversified audience as possible and was implemented using Google Forms, an accessible online survey administration platform.

A snowball sampling method was used to recruit respondents, who were given two weeks to complete the online survey. It featured 33 questions and covered seven response groups, corresponding to frequent types of medical questions in BabelDr: (i) yes/no (for example, "Do you have pain?"); (ii) location of the pain ("Where is your pain?"); (iii) pain description ("Can you describe your pain?"); (iv) time of day; (v) cause of the pain (activity, etc.); (vi) pain evaluation ("Can you evaluate your pain on a scale of 0 to 10?"); (vii) and visual field ("What do you see?"). Figure 3 gives an example of Sclera and Arasaac pictographs for pain description.

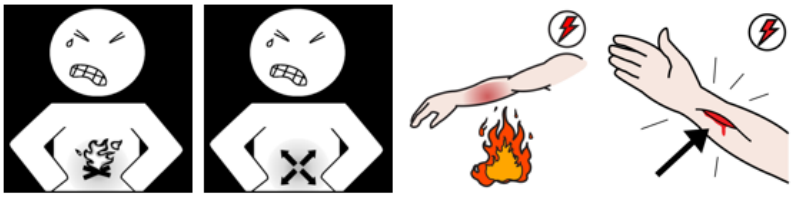

Figure 3. Sclera and Arasaac pictographs for pain description: "burning sensation", "radiating pain", "burn" and "cut"

The respondents had to guess the meaning of a pictograph response associated with a BabelDr question, for example the Sclera *"say no"* pictograph as response to "Avez-vous mal au ventre?" ("Do you have pain in the abdomen?"). The survey contained three different types of questions, as shown in Figure 4:

(a) 10 multiple-choice questions with several distractors where respondents had to select the correct meaning of a pictograph in the context of a specific question (for i/ii response groups and Arasaac and Sclera sets);

(b) 20 open questions where respondents had to describe the pictograph with a short text (for iii/iv/v groups and Arasaac and Sclera);

(c) 3 multiple-choice questions without distractors (for vi/vii groups, Arasaac only). In this case, respondents had to link a BabelDr sentence to a picture that we created with Arasaac pictographs (for example, "Have you lost your sight in your right eye?" with the picture 1). They could choose several possible responses, contrary to question type (a).

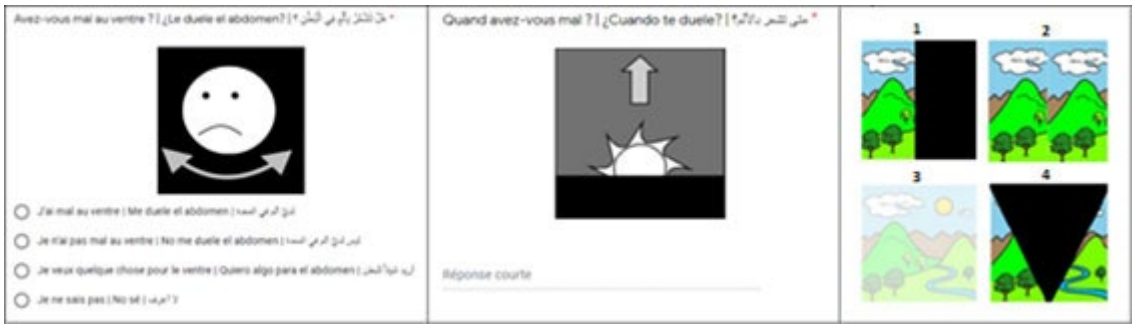

Figure 4. Example of three question types with Sclera and Arasaac pictographs

4 Findings

A total of 67 usable responses were collected through all language versions of the survey from three countries: 44 in Belgium, 18 in France and 5 in Switzerland. All were French-speaking (for 88.1%, it was their mother tongue), with the majority having always lived in Europe (85%) and speaking English as second language. The other mother tongues were: Arabic (10.4%), Spanish (4.5%), Armenian (1.5%), Czech (1.5%) and English (1.5%). In addition, some also spoke other languages such as Dutch or German. The age of respondents ranged between 18 and 66 years. Almost half were students (47.8%). Their fields of study or work were very diverse: literature, science, speech therapy (Norré et al. 2020), etc.

4.1 Question types

To investigate how to assess pictograph comprehension with different BabelDr sentences, we tested three types of questions.

In multiple-choice questions with distractors (a), we defined a single correct response per pictograph, which made it easy to obtain quantitative results. We obtained an average of 75.2% (percentage of correct responses). For open questions (b), we often noticed several possible interpretations for the same pictograph in a given context, especially for pain description. Open questions are more complex to evaluate, but we obtained a core meaning and some interesting variations for each case which can also be useful to identify possible interpretation problems. For example, for the "*radiating pain*" Sclera pictograph (Figure 3), responses included "diffuse pain", "cramps", "swelling" or "vomiting". For multiple-choice questions without distractors (c), we obtained an average of 86% (percentage of correct responses) for the three questions.

4.2 Response groups

We observed differences in comprehensibility between the response groups (i-vii).

Pictographs representing a part of the human body (82.5%) were recognized more easily than yes/no pictographs (64.1%). We obtained more different interpretations for pictographs for pain description, than for activity pictographs such as "*eat*" or "*go to bed*" ("Is belly pain triggered/relieved by something/anything?"). For time of day, we evaluated only one pictograph for the question: "Quand avez-vous-mal?" ("When does it hurt?"). More than 85% of participants responded correctly "matin" ("morning") or "quand je me réveille" ("when I wake up") for a sunrise representation with an upwards arrow (Figure 4). Some people confused this with "midi" ("noon") or "soir" ("evening"). For pain evaluation, correct responses by sentence ranged between 62.6% and 98.5%. The sentences at the first and last position on the scale (0: "I am not hurt", 2.5: "The pain is bearable", 5: "I am in pain", 7.5: "I am in a lot of pain", 10: "The pain is unbearable") obtained the best results. The visual field pictographs were very comprehensible (93.6%). However, no response group was recognized by all the participants; this shows that it is difficult to design "universal" pictographs.

4.3 Arasaac versus Sclera

For the multiple-choice questions with distractors, we obtained a score of 82% correct responses for Arasaac set and 68.3% for the Sclera set.

Several pictographs were comprehensible in both sets because the pictorial symbols are very similar (e.g. "*go to the toilet*" or "*morning*"). Other pictographs were less precise or more difficult to guess. For example: the "*belly*" Sclera pictograph had a good score (100% of correct responses) while the "*back*" pictograph in this set got a low score (55.2%) because of confusion with the "chest" distractor (41.8%) in the same question, due to the lack of the face and buttocks (Figure 5). In Sclera, pictographs are mainly black-and-white and often have few distracting details (Sevens 2018).

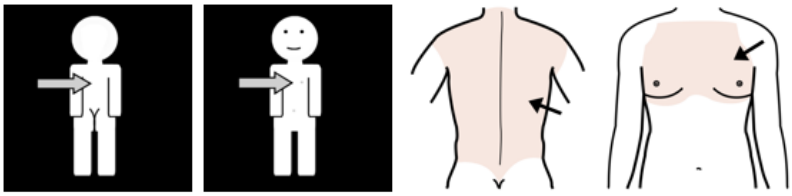

Figure 5. Sclera and Arasaac pictographs for human body: "back" and "chest"

For pain description, the Sclera set uses a human grimacing with a specific symbol in the belly, while the Arasaac set depicts pain on a specific part of the human body and includes a red lightening symbol, as shown in Figure 3. For the Arasaac "*chest pain*" pictograph, there was a majority of correct responses (74.6%), but 23.9% chose the bad interpretation "I have electricity in my chest". Alvarez (2011) had observed that lightening was a symbol often used for pain. However, it does not seem universal since in open questions we had an interpretation of "electric choc" for the "*cut*" pictograph (Figure 3).

There was also a large difference between yes/no pictographs (Figure 6) in Sclera (50%) and Arasaac (78.3%). In the medical context, the Sclera pictographs for "yes" and "no" are more difficult to understand because they combine the representation of yes/no movement and a happy/not happy face (mouth pulled down/up). If the doctor asks: "Avez-vous mal au ventre?" ("Do you have pain in the abdomen?"), the happy face of the "*yes*" pictograph was confusing: 17.9% did not know, 38.8% chose the wrong response, and 43.3% chose the correct.

Figure 6. Sclera and Arasaac pictographs for yes/no: "say yes", "say no", "yes" and "no"

5 Conclusion

This preliminary study did not show that one set is globally better than the other for all types of questions. The survey has also provided insight into the difficulties to represent responses to specific questions in medical settings with pictographs (or pictures) and showed the need to give the doctor the possibility of checking whether the patient has understood correctly. Since we had a heterogeneous group of participants, we could not observe if there were any cultural differences concerning the acceptance of pictographs.

Regarding methods for assessment of the pictographs, results for all three question types reveal the difficulty of evaluating pictures with textual glosses, which introduce ambiguity, the language being ambiguous. It is also possible that the respondent understands the pictograph but uses an incorrect word in the case where open questions are used. The choice of a single interpretation among several responses in a questionnaire remains limited. In BabelDr, the patient must choose one or more pictographs among several, which could help him to better understand the meaning of the pictographs in context. For example, if all the pictographs represent lightning, the patient can infer that the lightning probably does not mean "electricity", but "pain".

It would be interesting to conduct another study with more participants and the same pictographs for the three types of evaluation methods even if some pictographs do not exist in all the sets or with different names. Evaluating pictographs is often costly and time-consuming (Kim et al. 2009). To solve this problem, one possibility would be to use a crowdsourcing approach to validate images by a larger community (Christensen et al. 2017; Yu et al. 2013).

Future work will consist of evaluating the BabeldDr interface with Arasaac and Sclera pictographs directly with allophone users. This work started in August 2020. Another aspect that we are working on is the patients' satisfaction regarding pictographic

responses. More future studies need to be carried out with respect to the translation of questions into pictographs to allow bidirectional communication between doctors and more allophone patients in hospitals.

Acknowledgements

This work was partly funded by the Fond National Suisse (No. 197864) and the Agence Nationale de la Recherche, via the PROPICTO project (ANR-20-CE93-0005).
The pictographs used are property of Aragon Government and have been created by Sergio Palao for Arasaac. Aragon Government distributes them under Creative Commons License (BY-NC-SA). The other pictographs used are property of Sclera vzw which distributes them under Creative Commons License 2.0.

References

Alvarez, Juliana. 2011. “Évaluation des systèmes symboliques des symptômes biomédicaux pour la création d’une interface graphique pluriculturelle de l’échelle de triage canadien.” MA thesis, University of Montreal, Canada.

Beukelman, David R., and Pat Mirenda. 1998. *Augmentative & alternative communication: Supporting children and adults with complex communication needs*. Baltimore: Paul H. Brookes Publishing.

Bouillon, Pierrette, Johanna Gerlach, Hervé Spechbach, Nikos Tsourakis, and Sonia Halimi. 2017. “BabelDr vs Google Translate: a user study at Geneva University Hospitals (HUG).” In *Proceedings of the 20th Annual Conference of the EAMT*, *Prague, Czech Republic.* European Association for Machine Translation. https://archive-ouverte.unige.ch/unige:94511.

Cataix-Nègre, Elisabeth. 2017. *Communiquer autrement. Accompagner les personnes avec des troubles de la parole ou du langage : Les communications alternatives*. De Boeck Supérieur.

Christensen, Carrie, Doug Redd, Erica Lake, Jean P. Shipman, Heather Aiono, Roger Jr. Altizer, Bruce E. Bray, and Qing Zeng-Treitler. 2017. “Doodle Health: A Crowdsourcing Game for the Co-design and Testing of Pictographs to Reduce Disparities in Healthcare Communication.” In *AMIA Annual Symposium Proceedings*, 585–94. American Medical Informatics Association. https://hsrc.himmelfarb.gwu.edu/smhs_crl_facpubs/143.

Eadie, Kathy, Marissa J. Carlyon, Joanne Stephens, and Matthew D. Wilson. 2013. Communicating in the pre-hospital emergency environment. *Australian Health Review* 37 (2): 140–46.

Janakiram, Antony Akash, Johanna Gerlach, Alyssa Vuadens-Lehmann, Pierrette Bouillon, and Hervé Spechbach. 2020. “User Satisfaction with a Speech-Enabled Translator in Emergency Settings.” In *Digital Personalized Health and Medicine*, 1421–22.

Kim, Hyeoneui, Carlos Nakamura, and Qing Zeng-Treitler. 2009. “Assessment of Pictographs Developed Through a Participatory Design Process Using an Online Survey Tool.” In *Journal of Medical Internet Research* 11 (1), e5.

Mutal, Jonathan David, Pierrette Bouillon, Johanna Gerlach, Paula Estrella, and Hervé Spechbach. 2019. “Monolingual backtranslation in a medical speech translation system for diagnostic interviews-a NMT approach.” *In Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, *Dublin, Ireland,* 196–203. https://archive-ouverte.unige.ch/unige:123138.

Norré, Magali, Pierrette Bouillon, Johanna Gerlach, and Hervé Spechbach. 2020. “Évaluation de la compréhension de pictogrammes Arasaac et Sclera pour améliorer l'accessibilité du système de traduction médicale BabelDr.” In *Proceedings of the Eleventh Conference of IFRATH, Paris, France,* 177–82. Institut Fédératif de Recherche sur les Aides Techniques pour Personnes Handicapées. https://archive-ouverte.unige.ch/unige:144565.

Schwab, Didier, 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 *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, 166–171. Paris: ELRA. https://www.aclweb.org/anthology/2020.lrec-1.21.pdf.

Sevens, Leen. 2018. “Words Divide, Pictographs Unite: Pictograph Communication Technologies for People with an Intellectual Disability.” PhD diss., Katholieke Universiteit Leuven, Belgium.

Spechbach, Hervé, Johanna Gerlach, Sanae Mazouri Karker, Nikos Tsourakis, Christophe Combescure, and Pierrette Bouillon. 2019. “A Speech-Enabled Fixed-Phrase Translator for Emergency Settings: Crossover Study”. In *JMIR Medical Informatics* 7 (2), e13167.

Vandeghinste, Vincent, and Ineke Schuurman. 2014. “Linking pictographs to synsets: Sclera2Cornetto.” In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, vol. 9, 3404–10. Paris: ELRA. https://lirias.kuleuven.be/183055.

Wołk, Krzysztof, 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.” In *Computational and Mathematical Methods in Medicine, vol.* 2017. Hindawi. https://www.hindawi.com/journals/cmmm/2017/4306416/.

Yu, Bei, Matt Willis, Peiyuan Sun, and Jun Wang. 2013. “Crowdsourcing participatory evaluation of medical pictograms using Amazon Mechanical Turk.” In *Journal of Medical Internet Research* 15 (6), e108.

Authors

Magali Norré is a PhD Student at the University of Geneva and at the Center for Natural Language Processing (Cental) at the Catholic University of Louvain (UCLouvain) in Belgium. She is interested in translation systems with pictographs to improve communication between doctors and patients with an intellectual disability in hospitals.

Email: magali.norre@uclouvain.be

Pierrette Bouillon has been Professor at the FTI (Faculty of Translation and Interpreting), University of Geneva since 2007. She is currently Director of the Department of Translation Technology and Dean of the FTI. She has numerous publications in computational linguistics and natural language processing, particularly within lexical semantics (Generative lexicon theory), speech-to-speech machine translation for limited domains, post-editing and more recently accessibility.

Email: pierrette.bouillon@unige.ch

Johanna Gerlach is a Research and Teaching Fellow at the Department of Translation Technology of the Faculty of Translation and Interpreting at the University of Geneva. She holds a Master’s degree in translation and a PhD in Multilingual Information Processing. Her current work focuses on the BabelDr project, a spoken language translation system for the medical domain, where she contributes to the development and evaluation of interfaces and multilingual linguistic resources.

Email: johanna.gerlach@unige.ch

Hervé Spechbach is MD at the Geneva University Hospitals (HUG) since 2006 and head of outpatient emergency service since 2015. He has publications in informatics tools used in context of emergency department, which are mainly service-oriented. He represents the HUG in the Swiss Center for Barrier-free Communication project and co-leads the BabelDr project, an innovative and reliable medical phraselator, which received the 2018 innovation prize at the HUG.

Email: herve.spechbach@hcuge.ch