5

Enhancing Speech Translation in Medical Emergencies with Pictographs

BabelDr

WITH CONTRIBUTIONS FROM PIERRETTE BOUILLON, JOHANNA GERLACH, MAGALI NORRÉ AND HERVE SPECHBACH

5.1 Introduction

In emergency care settings, there is a crucial need for automated translation tools. In Europe, this need has been fueled by the migratory crisis (Spechbach et al., 2019), but the same need obtains in countries such as the USA (Turner et al., 2019) and Australia (Ji et al., 2020), where the foreign-born population is increasing. Emergency services often have to deal with patients who have no language in common with staff; and this issue has been shown to negatively impact both healthcare quality and associated costs (Meischke et al., 2013). In particular, a lack of clear communication can interfere with the prompt and accurate delivery of care (Turner et al., 2019). Language barriers also increase the risk of erroneous diagnoses and serious consequences (Flores et al., 2003).

According to Kerremans et al. (2018), various bridging solutions are currently used by services addressing asylum seekers or mental healthcare. They cite the use of plain language and professional or *ad hoc* interpreters, but also the use of gestures, communication technologies, and visual supports such as images or pictographs. In particular, in emergency settings where interpreters are not always available, there is a growing interest in the use of translation tools to improve communication (Turner et al., 2019). Fixed-phrase translators (Seligman and Dillinger, 2013), also known as "phraselators", are often used in the medical field for safety and accuracy reasons, for example, "Culturally and Linguistically Diverse (CALD) Assist," "Canopy Speak," "Dr. Passport (Personal)," "MediBabble Translator," "Talk To Me," and "Universal Doctor Speaker" (Panayotou et al., 2019; Khander et al., 2018). These are based on a limited list of pre-translated sentences, which then can be presented to the patient in written or spoken form, using either text-to-speech or human audio recordings. Some of these fixed-phrase systems are now relatively sophisticated and speech-enabled, for example, "BabelDr" (Spechbach et al., 2019).

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

These enable doctors to speak freely, with the system linking the recognition result to the closest source-language match that is a clear and explicit variant of the original sentence. This intermediate result can be presented to the doctor for confirmation, and can also be used as the input for translation into the system's target languages (Mutal et al., 2019; Bouillon et al., 2021).

Machine translation is another alternative, but the quality is too often low for this type of discourse, due in part to many context-dependent phenomena (ellipsis, etc.). Literal translation is often problematic as well, since cultural differences may influence the way questions are asked (Halimi et al., 2021). Some recent studies have showed that both patients and doctors tend to prefer a fixed-phrase translator to generic machine translation such as Google Translate (Turner et al., 2019; Panayotou et al., 2019; Bouillon et al., 2017).

We focus here on the BabelDr system, a speech-enabled phraselator used to improve communication in emergency settings between doctors and allophone patients (Bouillon et al., 2021). The aim of the chapter is two-fold. First, we wish to assess if a bidirectional version of the phraselator allowing patients to answer doctors' questions by selecting pictures from open-source databases will improve user satisfaction. Second, we wish to evaluate pictograph usability in this context. Our hypotheses are that images will in fact help to improve patient satisfaction and that multiple factors influence pictograph usability. Factors of interest include not only the comprehensibility of the pictographs *per se*, but also how the images are presented to the user with respect to their number and ordering.

Visual supports have been already suggested for medical dialogue in research studies among patients with limited English proficiency (Somers, 2007) or hospitalized individuals with language or motor disabilities (Eadie et al., 2013; Bandeira et al., 2011), and some systems are already available for medical use (see Section 5.2). However, to the best of our knowledge, BabelDr is the first system which integrates speech and automatically links doctors' spoken questions to specific pictographs for the patient. Some studies have evaluated the effect of pictographs on user satisfaction, but not in the context of a diagnostic interview or with a CALD population.

Section 5.2 of the chapter provides an overview for the reader of the broader context of pictographs in the medical domain. In Section 5.3, we describe the bidirectional BabelDr system and our method for selecting images and integrating them into the system. We then summarize two user studies intended to answer our research questions. The first focuses on user satisfaction (Section 5.4.1) and the second on pictograph usability (Section 5.4.2). Finally, in Section 5.5, we draw conclusions and briefly describe our future work on this topic.

5.2 Pictographs in Medical Communication

Patients, especially those with limited health literacy skills, often have trouble understanding health information. Pictographs are one proposal for clarifying and elucidating that information. As emphasized by Katz et al., (2006), "research in psychology and marketing indicates that humans have a cognitive preference for picture-based, rather than text-based, information".

In clinical settings, pictographs have been developed mainly for communication of health information and tested for delivery of specific instructions (concerning medication, etc.). In this domain, the use of images has been shown to positively affect patient comprehension by improving attention, recall, satisfaction, and adherence (Houts et al., 2006; Katz et al., 2006). For example, Hill et al., (2016) and Zeng-Treitler et al., (2014) evaluated automated pictograph illustrations generated by the Glyph system for communicating patient instructions (e.g., "Call your doctor if you experience fainting, dizziness, or racing heart rate"). They found that participants who received pictograph-enhanced discharge instructions recalled more of their instructions than those who received standard discharge instructions. In addition, patients were more satisfied with the understandability of their instructions. In the same context, several studies also highlighted the importance of using pictures together with written or oral instructions to avoid misinterpretation of picture-only instructions. That is, combinations of formats are generally preferred to picture- or text-alone (Houts et al., 2006).

Clearly, pictographs are of potential value, and in fact several sets are available. However, only a few are open-source, which limits actual usability. Some sets were developed for specific purposes. For example, USP pictograms were specifically developed to help convey medication instructions, precautions, and/or warnings to patients and consumers. Similarly, "Visualization of Concepts in Medicine" (VCM) (Lamy et al., 2008) is an iconic language based on a small number of graphical primitives and combinatory rules for facilitating access to drug monographs by practitioners. SantéBD is a French database, accessible under certain conditions, that provides educational content in the form of images, comics, or texts using the method "Easy-to-Read-and Understand" (FALC [Facile à Lire et à Comprendre]) is designed to aid individual comprehension in healthcare situations, but also to facilitate communication between doctors and patients during consultations (Figure 5.1). Similarly, "Widgit Health" (Vaz, 2013) offers a symbol board created to help medical staff to communicate quickly and easily in various domains, including coronavirus disease 2019 (COVID-19). Arassac and Sclera are two large open-source datasets (over 13,000 pictographs per set) designed for AAC

Figure 5.1 SantéBD

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

(Augmentative and Alternative Communication). They have been used in several contexts, including hospitals (Paolieri and Marful, 2018), and have been integrated into various online applications. In particular, the Sclera set was used by Vandeghinste and Schuurman (2014) in a text-to-pictograph translation system for people with disabilities, while the Arasaac set by Vaschalde et al., (2018) was used in a speech-to-pictograph system. Many other specific pictograph sets were designed for healthcare use, but are not accessible online (Cataix-Nègre, 2017; Beukelman and Mirenda, 1998).

Pictographs are unlikely to be universal (Sevens, 2018). Some medical research focused on the pictograph comprehensibility and crowdsourcing. Kim et al. (2009) concluded that "there is a large variance in the quality of the pictographs developed using the same design process". Yu et al. (2013) used Amazon Mechanical Turk (MTurk) workers to test a crowdsourcing approach in order to have 20 medical USP pictograms evaluated by 100 US "turkers." Their comprehensibility ranged between 45% and 98% (mean=72.5). Another study using a crowdsourced game called Doodle Health (Christensen et al., 2017) showed that it is possible to design a large set of medical images (596 drawings) and validate them by a larger community (114 volunteers made more than 1758 guesses). They obtained a score between 70% and 90%. According to the authors, this game had several limitations: not all participants had sufficient specialized knowledge to draw and/or recognize certain medical concepts, for example, the word "defibrillator". These studies show the importance of testing pictographs with a specific target group and task. In addition, most reports demonstrated an impact of the culture on comprehensibility. Yu et al. (2013) conclude that the "educational level is the only factor that affected participant performance". Kassam et al. (2004) similarly show that "basic education and time since immigration predicted interpretation accuracy better than first language or any other demographic characteristic".

Although the potential of pictographs for medical diagnosis is recognized (e.g., Somers, 2007), studies in this domain are very scarce (Alvarez, 2014). Existing medical phraselators generally do not contain pictographs (Wołk et al., 2017). Only a few medical pictographic fixed-phrase translators are available online, for example, "My Symptoms Translator" on Apple devices (Alvarez, 2014) or "Medipicto AP-HP" on Android and Iphone developed by the Hospitals of Paris; but these are quite limited and unsophisticated. "My Symptoms Translator" is aimed at reducing communication barriers and allowing patients to express their symptoms during medical emergencies. Pictographs represent types of pain, injuries, and medication. In the "Medipicto AP-HP" mobile application, the patient chooses pictographs labeled in his/her language to communicate with the caregiver, who can ask

questions by choosing pictographs translated into patient and caregiver languages from a predefined list. Wołk et al. (2017) also recently developed a cross-lingual medical aid application with pictographs on mobile devices (e.g., smartwatch) for communication between doctors, foreigners, and patients with speech, hearing, or mental disabilities. However, none of these applications can be adapted for specific needs or pictograph sets, and this limitation impedes use and evaluation. In the following sections, we describe BabelDr, conceived as a platform for experimentation in the domain of medical communication.

5.3 BabelDr and the Bidirectional Version

BabelDr is an online, speech-enabled phraselator for medical dialogue between doctors and patients (Bouillon et al., 2021, Spechbach et al., 2019). BabelDr is a project of the Faculty of Translation and Interpreting of the University of Geneva in collaboration with Geneva University Hospitals (Geneva, Switzerland). Several languages are available: Albanian, Arabic, Dari, (simple) English, Farsi, Spanish, Tigrinya, and Swiss-French sign language (LSF-CH) (Strasly et al., 2018).

The BabelDr interface was initially unidirectional and designed only for the translation of doctor's questions. Patients answered non-verbally using gestures (e.g., head movements for "yes" and "no"), facial expressions, etc. However, to allow doctors to ask open questions (likely to be faster, less restrictive, and more engaging), we have now designed a bidirectional interface (Figure 5.2) by manually associating BabelDr sentences with pictographs representing a range of possible responses for patients, for example, "burn," "sore throat," and "headache" pictographs in response to the question "Can you show me why you have come here?" ("*Pouvez-vous me montrer ce qui vous amène ?*").

The bidirectional interface includes two different views, one for the doctor and one for the patient. The doctors' view allows doctors to speak or to search for questions in a list, using keywords. When the doctor confirms the speech recognition result (based on the back-translation produced by the system (Spechbach et al., 2019)) or selects a sentence in the list, the system switches to the patient view and speaks the question for the patient in the target language. If desired, the patient can replay the spoken translation (or the video for the LSF-CH version). The patient view presents a selection of clickable response pictographs corresponding to the question, among which the patient can select his or her answer. To help patients use this interface, several animated visual

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

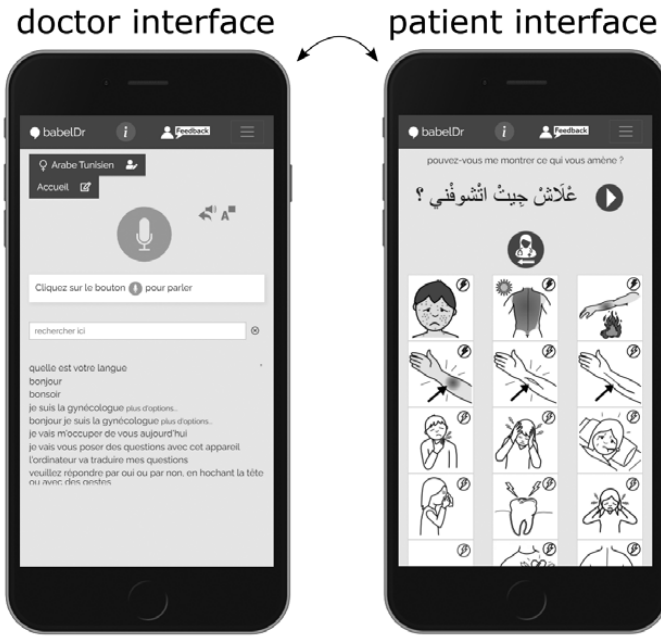

Figure 5.2 BabelDr bidirectional interface

hints are included. For example, the "Back" button is temporarily highlighted if the patient does not click on it within a given time after selecting a response. Once the patient has responded, the system switches back to the doctor view and displays the selected response(s) in written form in French. If necessary, the doctor can ask a new question to confirm the patient's answer. All questions and answers are automatically recorded in a history of the dialogue that the doctor can view at any time during the session or download as a pdf. The doctor can also deactivate the bidirectional version if required.

The pictographs were selected from the two open-source sets, Arasaac and Sclera, based on a previous study of comprehensibility in medical settings (Norré et al., 2020, 2021). In Sclera, the pictographs are mainly black-and-white and designed with few distracting details. As mentioned by Sevens (2018), the "characters that are depicted on the pictographs do not present a specific race, body type, age, or gender, thus referring to virtually any person in the world," as compared with Arasaac pictographs, for which we had to choose the gender of the character each time (Figure 5.3). The Arasaac pictographs provided by the Aragonese Portal of AAC are available in color and in

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

Figure 5.3 Examples of one Sclera and four Arasaac pictographs (no gender, female and male) for "headache"

Figure 5.4 Examples of Arasaac pictographs for "yes", "no" and "I don't understand" in BabelDr

black-and-white. They are often more detailed and there are sometimes several variations for the same concept.

In the previous study on comprehensibility, we concluded that neither set is superior for all question types (Norré et al., 2020, 2021). For closed questions, we used the Arasaac "yes" and "no" pictographs (Figure 5.4), which had obtained a higher comprehension score (78.3%) than those in Sclera (50%). In the medical context, the Sclera pictographs for "yes" and "no" are not appropriate, as they combine the representation of a yes/no movement and a happy/not happy face (mouth pulled down/up). If the doctor asks: "Do you have pain in the abdomen?" ("*Avez-vous mal au ventre ?*"), the happy face of the "yes" pictograph can be confusing. For all interactions (including introductory phrases such as "Hello, I am the doctor" or "I will take care of you today"), questions and patient instructions, we included the Arasaac "I don't understand" pictograph (Figure 5.4). We used Sclera pictographs for questions related to the pain description because they appear to be less problematic in our context.

We have noted various comprehension issues. In Arasaac, for instance, a given pictograph often represents several concepts. For example, a specific type of pain (burn, etc.) is always depicted on a certain part of the body (arm, etc.), so that the relevant pictograph conveys both "burn" and "arm". (Linguistically, the

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

combination might be expressed in a prepositional phrase, e.g., "burn on the/your arm"). The problem is that, in response to open questions such as "Can you describe your pain?", patients might not choose that pictograph if they have a burn elsewhere than pictured (or, conversely, if their arm hurts, but it is not a burn). There are no pictographs representing a burn in all possible places (and in fact the medical coverage of this set is limited overall). In the Sclera set, the type of pain is represented by a grimacing and identical character with a specific symbol ("fire", "hammer") for the symptom always located in the stomach area (Figure 5.5).

Additionally, the first Arasaac pictographs always used the same symbol to categorize pictographs related to health (a red cross) or pain description (a red lightning bolt) (Figures 5.5 and 5.6). In the preliminary study, these were often shown to be sources of ambiguity. When we asked the participants what the "chest pain" pictograph meant, they often gave the interpretation "I have electricity in my chest" (Figure 5.6). Even so, we can hope that, in the medical context, patients might after all infer that the lightning probably means "pain" rather than "electricity".

In any case, to improve the coverage of patient responses in BabelDr, we created and adapted some Arasaac pictographs, for example, those that were

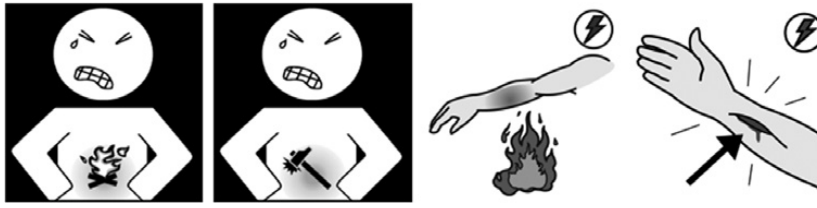

Figure 5.5 Examples of Sclera for pain description: "burning pain", "throbbing pain", and of Arasaac pictographs for "burn" and "cut"

Figure 5.6 Arasaac pictograph for "chest pain" and pictographs that we created or adapted for "Syria", "left ear" and "five glasses"

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

missing for some countries. The patient can choose from 61 countries; between a right ear and a left ear for the question "In which ear do you hear less well?"; and between one or more glasses/bottles of wine for the question "How many glasses of alcohol do you drink per day?", etc. (Figure 5.6).

One aim of BabelDr is to make its content easily expandable – first, to adapt to new health situations or demographics, but also to carry out experiments with various tool configurations for research purposes. An online interface allows developers to upload pictographs; define their corresponding (French) written forms, that is, the responses to be displayed for doctors; and finally link these pictographs to BabelDr questions, as shown in Figure 5.7. This interface enables easy integration of various sets of pictographs into the system, depending on needs, and enables direct evaluation of tasks, as proposed in these experiments. To aid linkage of the BabelDr sentences with pictographs, we manually categorized each BabelDr sentence according to the type of response expected by the doctor, for example, yes/no, pain description, cause and location of pain (e.g., activity, human body), time of day, ways to take medication, food, positions and movements, sports, countries and languages, colors, animals, professions, etc. Some pictographs were used for many questions. In total, BabelDr now includes approximately 395 unique pictographs that we sometimes had to rename to make them understandable in the context of the doctor's dialogue history. On average, each question is associated with twenty pictographs (not including yes/no questions with three possible responses or input fields with only one possible response). The maximum number of pictographs per question is sixty-one for questions related to countries (such as "Have you traveled recently?").

5.4 Usability of the Bidirectional Version of BabelDr

The usability of the bidirectional version of BabelDr was evaluated in two different studies. The first aimed at comparing patient satisfaction with the unidirectional and bidirectional versions, while the second focused on pictograph usability in the medical context.

5.4.1 Patient Satisfaction

5.4.1.1 Design

The first study aimed to compare patient satisfaction among foreign-speaking patients with the unidirectional as compared with the bidirectional version of BabelDr. The study was conducted online during the period of the COVID-19 epidemic in August and September 2020. In

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

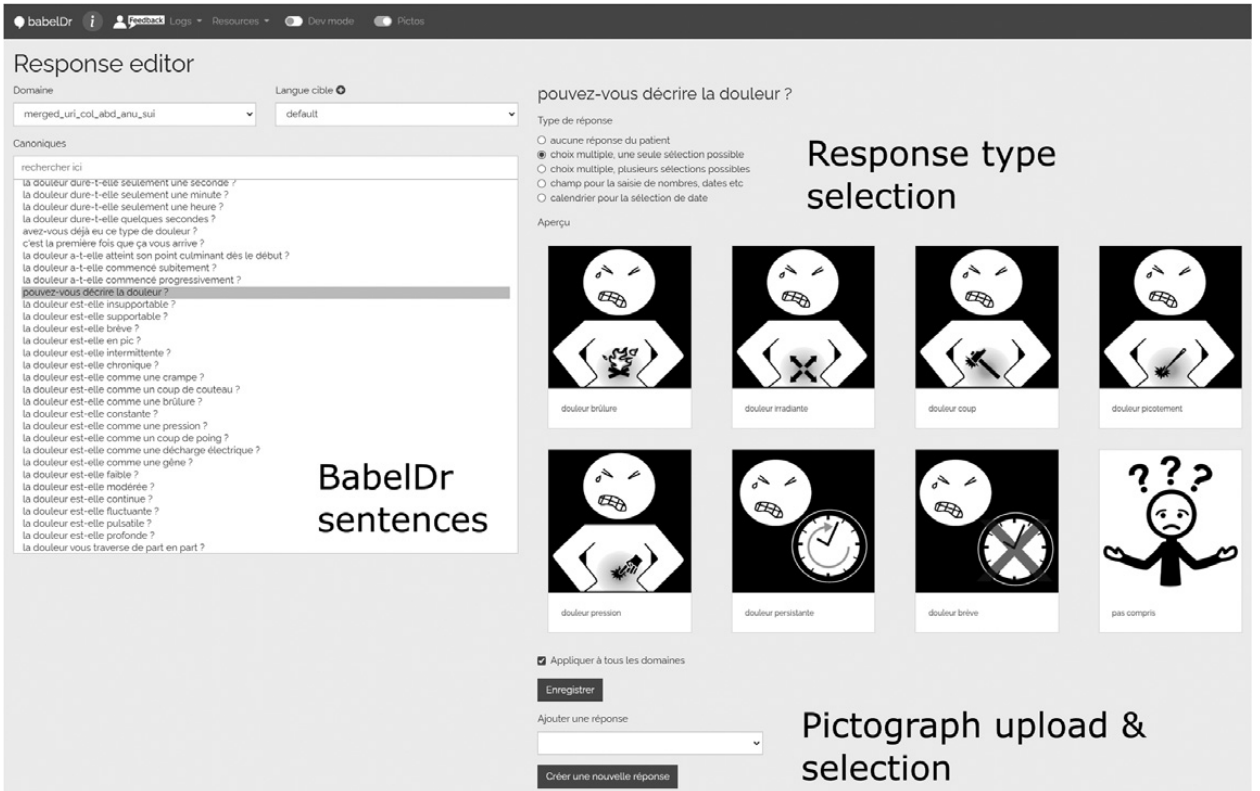

Figure 5.7 BabelDr response editor

these user tests, twelve Arabic-speaking participants were asked to answer 50 medical questions with the two BabelDr interfaces via the Zoom video conferencing tool. Questions included 36% of the yes-no questions and 64% of the open questions about COVID-19 and the patient's history. Patients received task instructions via email. For the bidirectional part, they had to respond by clicking on one or more pictographs relevant to the context of the question. For the unidirectional part, they did not have access to pictographs, and thus had to find the best way to respond without speaking, for example, using gestures or facial expressions.

At the end of each user test, patients had to complete a satisfaction questionnaire, which consisted of a total of twenty items (ten for each type of interface). Items were derived from the System Usability Scale (SUS) questionnaire by Brooke (1996) and adapted to the functionalities of BabelDr. A 5-point Likert scale ("strongly disagree", "disagree", "neutral", "agree", and "strongly agree") was used to rate agreement with items. Patients were also asked to indicate which version they preferred.

Participants were recruited on social networks in groups linked to refugees in Belgium, charitable associations, or academic groups. In total, twelve people tested the system, including eleven males living in Belgium and one female in France. The inclusion condition for all participants was Arabic as mother tongue.

5.4.1.2 Results

During the entire experiment, patients selected more than 200 pictographs, of which 81 were unique. Figure 5.8 summarizes the results of the SUS test. Overall, the results of the satisfaction questionnaire were very positive (no one strongly disagreed with the various statements, such as "The system was easy to use"), with most participants agreeing or strongly agreeing with most statements, for both interfaces, unidirectional (without pictographs) and bidirectional (with pictographs).

We calculated averages of scores by item (0: no response, 1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree). To produce an overall score on a range of 0 to 100 for each system following the SUS approach, we summed the score contributions from the 10 items (see Table 5.1 for scores by item) and multiplied the result by two. The two systems are very close, achieving overall scores of 85.1 and 86.2 for unidirectional and bidirectional, respectively.

All patients found both versions of the system easy to use, with a slightly higher score for the bidirectional version (Q1), and felt the system enabled

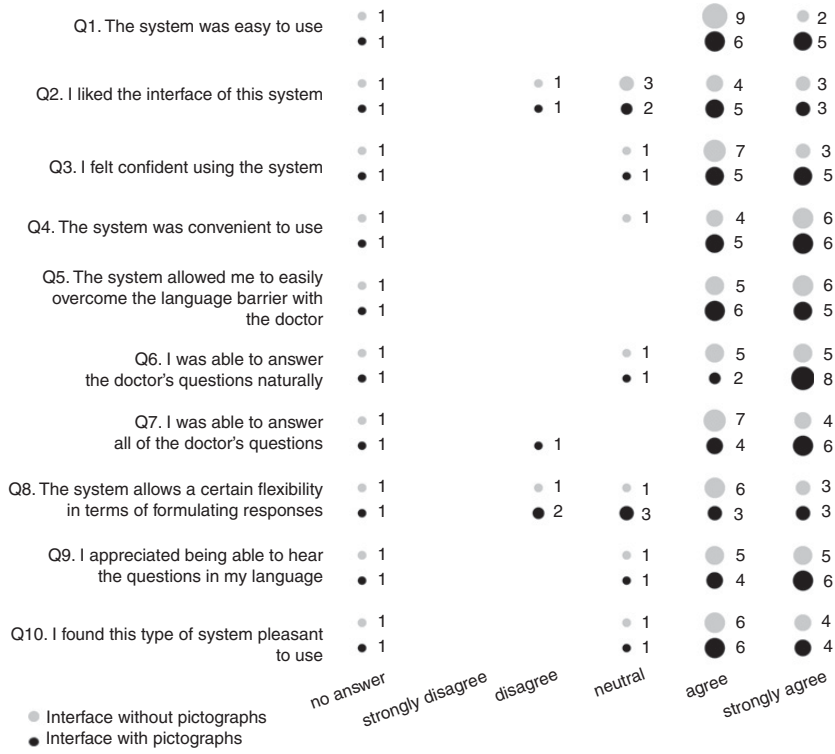

Figure 5.8 Results of the satisfaction questionnaire completed after the experiment for the interface without pictographs and the interface with pictographs. Numbers on the right side of the circles represent the number of patients (n=12)

them to easily overcome the language barrier with the doctor (Q5). They also felt more confident using the bidirectional version (Q3). The system was judged convenient to use (Q4), even though it was tested remotely via videoconference. The statements concerning appreciation of the interface (Q2) and flexibility for formulating responses (Q8) received slightly more mixed opinions than the others (Figure 5.8). Thus there seems to be room for improvement, even though the bidirectional interface allowed the clear majority (8 "strongly agree") to answer doctors' questions more naturally (Q6). Surprisingly, all patients ("strongly") agreed that they were able to answer all of the doctor's questions even with the unidirectional version (Q7), although in fact they actually did not respond to all the questions. We conclude that the results for the assessment of the text-to-speech (Q9) and the complete system (Q10) are similar for both interfaces.

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

Table 5.1 *Results (/5) of the satisfaction questionnaire for the unidirectional and bidirectional interfaces (n=11). Standard deviation is given in brackets.*

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Uni	4.2 (0.4)	3.8 (1.0)	4.2 (0.6)	4.5 (0.7)	4.5 (0.5)	4.4 (0.7)	4.4 (0.5)	4.0 (0.9)	4.4 (0.7)	4.3 (0.6)
Bidi	4.5 (0.5)	3.9 (0.9)	4.4 (0.7)	4.5 (0.5)	4.5 (0.5)	4.6 (0.7)	4.4 (0.9)	3.6 (1.1)	4.5 (0.7)	4.3 (0.6)

Of the twelve participants, almost all (n=10) preferred the bidirectional version, except one who preferred the interface without pictographs and one who did not answer. We received several comments highlighting the advantages of the bidirectional version: "It makes it easier for the person to answer and communicate" (translated from: "تسهل على الشخص الاجابة و التواصل"); "It makes it easier to clarify the problem, because we can show exactly where the pain is for example" (translated from: "*Parce que on peut montrer exactement où se trouve la douleur par exemple*"); "Photos make the expression easier in order to answer the questions better!" or "I found it better and useful for people". We received no comments about the interface without pictographs.

5.4.2 Pictograph Usability

5.4.2.1 Design

In the second study, we looked at the usability of the pictographs in the medical context, with a focus on: 1) their comprehensibility; and 2) for each question, how the number and order of pictographic response choices affect users' (a) ability to correctly find predefined responses and (b) response time. Our hypotheses are the following:

- responses to questions (including, for example, symptoms, actions, or pain descriptions) can be illustrated understandably using pictographs;
- including more response choices per question will lead to longer response times and/or more errors;
- the order in which the pictographic responses are presented will affect the selection.

For this experiment, we created a customized version of the bidirectional BabelDr system showing only the patient view. Participants were presented with a doctor's questions in French, accompanied by French audio produced by speech synthesis (and replayable at will), and the French written form of the "correct" response that should be chosen among the proposed response pictographs (e.g., headache). The French form was the official name of the

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

pictographs (i.e., filenames in the Sclera and Arasaac sets). Participants were allowed to select only one response per question. The system logged the selected responses, as well as the response time for each question – the time between presentation of the question with its response choices and the validation of the response by the user. Figure 5.9 shows an example of the interface.

The study included six questions: three open questions repeated twice, with different correct responses. A closed question ("Do you understand what I am saying to you?") was used to introduce the test interface and response mechanism. This was followed by a question asking users to select from a list the languages with which they were familiar. These two questions were not counted in the results.

We used a between-subjects study design, in which each participant answered the same six question/response combinations in one of three different versions of the test. The versions were created by varying the number of response choices shown to the participant (five, ten, or fifteen) for each of the doctor's questions (Table 5.2), with each version including two questions with five choices, two with 10, and two with 15. In addition, the position (at the beginning, in the middle or at the end) of the correct response (in bold in Table 5.2) was automatically randomized for each participant.

The correct responses are presented in Figure 5.10. We used Arasaac (for Q1, Q3, Q4, Q6) and Sclera (Q2, Q5) pictographs in black-and-white. In addition,

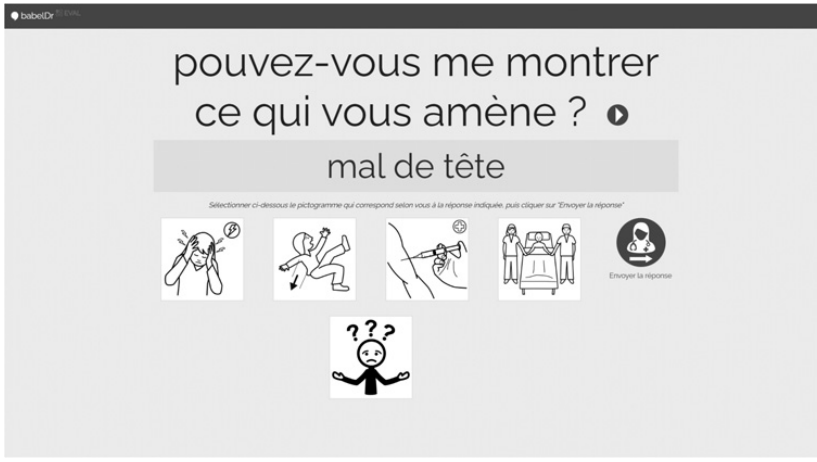

Figure 5.9 Example of the evaluation interface showing the patient's view of the bidirectional version of BabelDr, with an additional text field (green background) to display the "correct" response to select, here "headache" ("mal de tête")

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

Table 5.2 *Question and response choices. The responses are given using the names of the Arasaac and Sclera pictographs*

	Can you show me what's going on? (Q1|Q4)	Can you describe your pain? (Q2|Q5)	Show me the movements that make the pain worse (Q3|Q6)
5 responses	**Fall**, **headache**, I don't know, injection, visit	Burning pain, I don't know, **nagging pain**, **pain radiating**, prickling pain	Eat, **go to sleep**, I don't know, **lean**, sit on the toilet
10 responses	5 Previous responses + blow the nose, cough, fever, shivery, sore throat	5 Previous responses + cramping pain, pain insensitively, pain numbness, pain pressure, throbbing pain	5 Previous responses + drink, get out of bed, sit on the chair, sleep, stand up from chair
15 responses	10 Previous responses + heart attack, hot, rehabilitation specialist, stomach ache, vomit	10 Previous responses + brief pain, little pain, pain always, pain sometimes, pressing pain	10 Previous responses + pick up, run, sport, urinate, work out

Figure 5.10 Correct pictographs for "headache" (Q1), "fall" (Q4), "nagging pain" (Q2), "pain radiating" (Q5), "lean" (Q3) and "go to sleep" (Q6)

the "I don't understand" pictograph was always presented as a response option, positioned after all the other pictographs.

The participants received a link, which brought them to one of the three versions of the test. No instructions were given regarding the device used to complete the task and users were free to use a desktop, mobile phone, or tablet. The user agent was also stored in the logs.

Forty-five participants were recruited among Master-level students at the Faculty of Translation at the University of Geneva. All have French as a working language, but not all are native French speakers. This roster allowed us to collect fifteen responses for each of the test versions.

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

5.4.2.2 Results

5.4.2.2.1 Comprehensibility of Pictographs Table 5.3 shows the number of correct pictograph selections for each question and the number of response choices. The proportion of correct responses by question varied between 2% and 91%, thus suggesting large differences in the difficulty of the questions and/or complexity of the response pictographs. According to Goodman and Kruskal's lambda, there is an association between the question (Q1–Q6) and correctness ($\lambda = 0.268$). We observed that the pain description questions (Q2 and Q5) obtained far fewer correct responses, suggesting either that the corresponding pictographs are less comprehensible, or that the pain qualifiers used to provide the written form of the "correct" response are more complex or difficult to understand for non-native French speakers.

5.4.2.2.2 Impact of the Number and Order of Pictographic Response Choices Looking at the combined results for all questions (see last column of Table 5.3), we observe that when the number of presented response choices is increased, the proportion of correct responses decreases. Although this is not the case for all of the individual questions, this tendency does suggest that increasing the number of response choices makes it harder for users to find the correct one.

The second variable analyzed is response time. Table 5.4 shows the median response time by question after removal of outliers.[1] We observe that the response time strongly varies between questions, with median response times

Table 5.3 *Correct responses by question and number of available response choices*

Response choices	Q1	Q2	Q3	Q4	Q5	Q6	all
5	14 (93%)	1 (7%)	14 (93%)	15 (100%)	10 (67%)	13 (87%)	67 (74%)
10	14 (93%)	0 (0%)	11 (73%)	15 (100%)	8 (53%)	13 (87%)	61 (68%)
15	13 (87%)	0 (0%)	10 (67%)	11 (73%)	9 (60%)	14 (93%)	57 (63%)
combined	41 (91%)	1 (2%)	35 (78%)	41 (91%)	27 (60%)	40 (89%)	185 (69%)

[1] The median time per question was 11.800ms, with eight very high values where participants took more than one minute to answer a single question. As the experiment was not carried out under controlled conditions, participants may have been distracted by influences external to the task. We have therefore excluded these eight extreme values from our analysis.

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

Table 5.4 *Median response time by question*

Question	Q1	Q2	Q3	Q4	Q5	Q6
Median response time [ms]	11,760	20,127	16,386	6673	10,360	7920

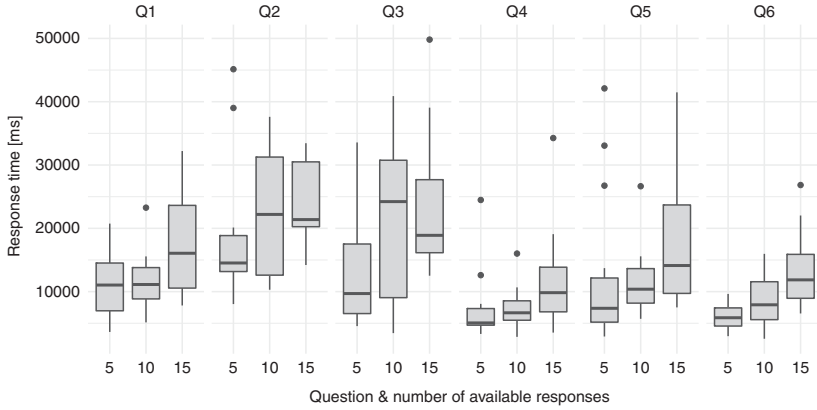

Figure 5.11 Response time in milliseconds for Q1-Q6, grouped by number of responses presented

ranging from 6 to 20 seconds. Moreover, response times are not normally distributed (Shapiro-Wilk test, $p<0.01$). A Kruskal-Wallis test showed that the question has a relatively strong, significant effect on the response time ($X^2(5, N=262) = 71.89$; $p<0.001$; $\varepsilon^2=0.275$).

As illustrated in Figure 5.11, response times are also influenced by the number of response pictographs presented: for most questions, the response time increases with the number of pictographs among which the participant had to find the correct response. The computation of Kendall's Tau ($\tau=0.293$) confirms that there is a medium-to-strong association between response time and the number of pictographs.

Regarding the order in which pictograph response choices are presented, in particular the position of the correct pictograph among the choices, we observed no impact on the correctness of the response according to Goodman and Kruskal's lambda ($\lambda = 0.05$). Response times appear equally unaffected, with median response times of 11,926, 11,152, and 10,410 milliseconds for correct pictographs positioned respectively at the beginning, middle, or end of the available choices. A Kruskal-Wallis test showed that the effect of the order on the response time is not significant ($X^2(2, N=262)=0.192$; $p=0.908$; $\varepsilon^2=0.0007$).

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

These results suggest that participants will look at all proposed options, even if they have already found a matching pictograph.

Finally, regarding the device used, seven of the forty-five participants completed the test on a mobile device, while the others used a desktop. The type of device did not have an impact on the correct selection of responses (Phi coefficient = -0.04).

5.5 Conclusion

To sum up, we assess the potential of using pictographs for medical dialogue and demonstrate the importance of evaluating their comprehensibility in a real context. We present two studies focused on the BabelDr system, a speech-enabled phraselator used to improve communication between doctors and allophone patients in emergency settings. The first study compared patient satisfaction with the bidirectional and unidirectional versions of BabelDr. Findings show that both versions are easy and convenient to use, even remotely, although most respondents prefer to use the interface with pictographs.

The second study aimed to evaluate the pictographs' usability in context. In a customized version of the bidirectional BabelDr system showing only the patient view, participants were presented with a doctor's question in French; a set of pictographic response choices; and the written form of the "correct" response that they should select. Results show that the pictographs are not equally comprehensible and that some – in particular, those used to describe pain types – present considerable difficulties, with as few as 2% of participants identifying the correct one. Regarding the number of pictographs presented, we observe that an increased number of response choices negatively affects participant's ability to select the correct answer and increases response time, thereby confirming our second hypothesis. Finally, regarding our third hypothesis, results do not show a notable impact of the order in which the pictographic responses are presented. Overall, this experiment has shown that multiple factors influence participants' ability to find a pictograph based on a written form, but that the comprehensibility of the individual pictographs is probably the most important.

These studies have some limitations. First, participants were not real patients in emergency situations, so factors such as stress or time constraints could not be considered. Second, we evaluated only a subset of the diagnostic questions available in BabelDr, with response pictographs extracted from only two open-source sets designed for AAC. A more extensive study using other pictograph

https://doi.org/10.1017/9781108938976.006 Published online by Cambridge University Press

sets, for example, domain-specific pictographs or illustrations aimed at different target audiences, would further our understanding of usability in this context. It would also be worthwhile to investigate whether the available pictographs cover all the symptoms and reasons for seeking consultation necessary for diagnosis in emergency settings.

Many studies have evaluated the usability of pictographs in the medical domain. However, to the best of our knowledge, our work contributes novel insights by focusing on the use of pictographs for diagnosis in a real-life system setting. Due to its flexible architecture, the BabelDr system is well suited to facilitate evaluation of various pictograph sets in a concrete and task-oriented manner. As an additional advantage of performing such evaluation directly in a medical translation tool, we can target varied language groups, such as the simulated CALD population of our first study.

5.6 Acknowledgments

This work is 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 the property of the Government of Aragón, which distributes them under a Creative Commons License (BY-NC-SA), and have been created by Sergio Palao for Arasaac (http://arasaac.org). The other pictographs used are the property of Sclera vzw (www.sclera.be/), which distributes them under a Creative Commons License 2.0.

References

Alvarez, J. 2014. "Visual Design: A Step Towards Multicultural Health Care," *Arch Argent Pediatr*, 112 (1) pages 33–40.

Bandeira, F. M., Faria, F. P. D., and Araujo, E. B. D. 2011. "Quality Assessment of Inhospital Patients Unable to Speak Who Use Alternative and Extended Communication." In *Einstein*, São Paulo, 9, pages 477–482.

Beukelman, D. R., and Mirenda, P. 1998. *Augmentative and Alternative Communication*. Baltimore: Paul H. Brookes.

Bouillon, P., Gerlach, J., Mutal, J. D., Tsourakis, N., and Spechbach, H. 2021. "A Speech-Enabled Fixed-Phrase Translator for Healthcare Accessibility." In *Proceedings of the 1st Workshop on NLP for Positive Impact*. Association for Computational Linguistics.

Bouillon, P., Gerlach, J., Spechbach, H., Tsourakis, N., and Halimi, S. 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.

Boujon, V., Bouillon, P., Spechbach, H., Gerlach, J., and Strasly, I. 2018. "Can Speech-Enabled Phraselators Improve Healthcare Accessibility? A Case Study Comparing BabelDr with MediBabble for Anamnesis in Emergency Settings." In *Proceedings of the 1st Swiss Conference on Barrier-free Communication (BfC)*, Winterthur, Switzerland.

Brooke, J. 1996. "SUS: A Quick Dirty Usability Scale," *Usability Evaluation in Industry*. 189 (3), pages 189–194.

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

Christensen, C., D. Redd, E. Lake, et al. 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*. American Medical Informatics Association, 585–594.

Eadie, K., Carlyon, M. J., Stephens, J., and Wilson, M. D. 2013. "Communicating in the Pre-Hospital Emergency Environment," *Australian Health Review*, 37 (2) pages 140–146.

Flores, G., Laws, M. B., Mayo, S. J., et al. 2003. "Errors in Medical Interpretation and Their Potential Clinical Consequences in Pediatric Encounters," *Pediatrics*, 111 (1) pages 6–14.

Halimi, S.A, Azari, R., Bouillon, P., and Spechbach, H. 2021. "A Corpus-Based Analysis of Medical Communication: Euphemism As a Communication Strategy for Context-Specific Responses." In *Corpus Exploration of Lexis and Discourse in Translation*. Routledge, 1–25.

Hill, B., Perri-Moore, S., Kuang, J., et al. 2016. "Automated Pictographic Illustration of Discharge Instructions with Glyph: Impact on Patient Recall and Satisfaction," *Journal of the American Medical Informatics Association*, *23* (6), 1136–1142.

Houts, P. S., Doak, C. C., Doak, L. G., and Loscalzo, M. J. 2006. "The Role of Pictures in Improving Health Communication: A Review of Research on Attention, Comprehension, Recall, and Adherence," *Patient education and counseling*, 61 (2),173–190.

Ji, M., Sørensen, K., and Bouillon, P. 2020. "User-Oriented Healthcare Translation and Communication." In *The Oxford Handbook of Translation and Social Practices*. Oxford University Press.

Kassam, R., Vaillancourt, R., and Collins, John B. 2004. "Pictographic Instructions for Medications: Do Different Cultures Interpret Them Accurately?" *International Journal of Pharmacy Practice*, 12(4),199–209.

Katz, M. G., Kripalani, S., and Weiss, B. D. 2006. "Use of Pictorial Aids in Medication Instructions: A Review of the Literature," *American Journal of Health-System Pharmacy*, 63(23), 2391–2397.

Khander, A., Farag, S., and Chen, K. T. 2018. "Identification and Evaluation of Medical Translator Mobile Applications Using an Adapted APPLICATIONS Scoring System," *Telemedicine and e-Health*, 24(8), 594–603.

Kerremans, K., De Ryck, L. P., De Tobel, V., et al. 2018. "Bridging the Communication Gap in Multilingual Service Encounters: A Brussels Case Study," *The European Legacy*, 23(7–8), 757–772.

Kim, H., Nakamura, C., and Zeng-Treitler, Q. 2009. "Assessment of Pictographs Developed Through a Participatory Design Process Using an Online Survey Tool," *Journal of Medical Internet Research*, 11(1), e5.

Lamy, J. B., Duclos, C., Bar-Hen, A., Ouvrard, P., and Venot, A. 2008. "An Iconic Language for the Graphical Representation of Medical Concepts," *BMC Medical Informatics and Decision Making*, 8(1), 16.

Meischke, H. W., Calhoun, R. E., Yip, M. P., Tu, S. P., and Painter, I. S. 2013. "The Effect of Language Barriers on Dispatching EMS Response," *Prehospital Emergency Care*, 17(4), 475.

Mutal, J. D., Bouillon, P., Gerlach, J., Estrella, P., and Spechbach, H. 2019. "Monolingual Backtranslation in a Medical Speech Translation System For Diagnostic Interviews: A NMT Approach." In *Proceedings of Machine Translation Summit XVII: Translator*, *Project and User Tracks*, 196–203.

Norré, M., Bouillon, P., Gerlach, J., and Spechbach, H. 2021. "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)*. ZHAW.

Norré, M., Bouillon, P., Gerlach, J., and Spechbach, H. 2020. Évaluation de la compréhension de pictogrammes Arasaac et Sclera pour améliorer l'accessibilité du système de traduction médicale BabelDr. Handicap 2020: Technologies pour l'autonomie et l'inclusion, 177–182.

Panayiotou, A., Gardner, A., Williams, S., et al. 2019. "Language Translation Apps in Health Care Settings: Expert Opinion," *JMIR Mhealth Uhealth*, 7(4):e11316.

Paolieri, D., and Marful, A. 2018. "Norms for a Pictographic System: The Aragonese Portal of Augmentative/Alternative Communication (ARASAAC) System," *Frontiers in Psychology*, 2538.

Seligman, M., and Dillinger, M. 2013. "Automatic Speech Translation for Healthcare: Some Internet and Interface Aspects." In *Proceedings of 10th International Conference on Terminology and Artificial Intelligence (TIA-13)*, Paris, France.

Sevens, L. 2018. *Words Divide, Pictographs Unite: Pictograph Communication Technologies for People with an Intellectual Disability.* Thesis, Katholieke Universiteit Leuven, Belgium.

Somers, H. 2007. "Theoretical and Methodological Issues Regarding the Use of Language Technologies for Patients with Limited English Proficiency." In *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-07)*, Skövde, 206–213.

Spechbach, H., Gerlach, J., Karker, S. M., et al. 2019. "A Speech-Enabled Fixed-Phrase Translator for Emergency Settings: Crossover Study," *JMIR Medical Informatics*, 7 (2), e13167.

Strasly, I., Sebaï, T., Rigot, E., et al. 2018. "Le projet BabelDr : rendre les informations médicales accessibles en Langue des Signes de Suisse Romande (LSF-SR)." In *Proceedings of the 2nd Swiss Conference on Barrier-free Communication*, Geneva, Switzerland, 92–96.

Turner, A. M., Choi, Y. K., Dew, K., et al. 2019. "Evaluating the Usefulness of Translation Technologies for Emergency Response Communication: A Scenario-Based Study," *JMIR Public Health and Surveillance*, 5(1), e11171.

Vandeghinste, V., and Schuurman, I. 2014. "Linking Pictographs to Synsets: Sclera2Cornetto." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 14)*. ELRA, Paris, 3404–3410.

Vaschalde, C., Trial, P., Esperança-Rodier, E., Schwab, D., and Lecouteux, B. 2018. "Automatic Pictogram Generation from Speech to Help the Implementation of a Mediated Communication." In *Proceedings of the 2nd Swiss Conference on Barrier-free Communication*, Geneva, Switzerland, 97–101.

Vaz, I. 2013. "Visual Symbols in Healthcare Settings for Children with Learning Disabilities and Autism Spectrum Disorder." *British Journal of Nursing*, 22 (3),156–159.

Wołk, K., Wołk, A., and Glinkowski, W. 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*. DOI: 10.1155/2017/4306416.

Yu, B., Willis, M., Sun, P., and Wang, J. 2013. "Crowdsourcing Participatory Evaluation of Medical Pictograms Using Amazon Mechanical Turk," *Journal of Medical Internet Research*, 15(6), e2513.

Zeng-Treitler, Q., Perri, S., Nakamura, C., et al. 2014. "Evaluation of a Pictograph Enhancement System for Patient Instruction: A Recall Study," *Journal of the American Medical Informatics Association*, *21*(6), 1026–1031.