Automatic Pictogram Generation from Speech to Help the Implementation of a Mediated Communication

Céline Vaschalde
Université d'Orléans La Source
Univ. Grenoble Alpes, LIG, Grenoble, France
vaschalde.c@gmail.com

Pauline Trial, Emmanuelle Esperanca-Rodier, Didier Schwab and Benjamin Lecouteux
Univ. Grenoble Alpes, LIG, Grenoble, France
pauline.trial@etu.univ-grenoble-alpes.fr
emmanuelle.esperanca-rodier|didier.schwab|benjamin.lecouteux @univ-grenoble-alpes.fr

Abstract

The goal of our research is to develop an automatic pictogram generation tool from speech to help the social circle of users of Alternative and Augmentative Communication to communicate among themselves. We describe here the issues of such a tool, we then detail our development methodology and finally we describe our evaluation protocol.

1 Introduction

When the use of speech or sign language to communicate is impossible because of aphasia, dysarthria and aggravating physical disorders, people are not able to express their feelings or needs and can't create any social link, which is central to the proper development of a human being.

Using Alternative and Augmentative Communication (AAC) methods could be a way to help these people. These methods replace or support a speaker's speech abilities. They often use visual encoding of the information, especially pictograms which are more iconic than words due to their likeness to the referent. (Duboisdindien, 2014)

The pictogram can be defined, in AAC, as a schematic graphic sign whose signifier has a more or less strong similarity with the signified, unlike phonic or graphic linguistic signs whose stimulus form is arbitrary and independent of that of the referent. It therefore allows a more iconic representation of the information and is therefore more easily interpretable. Nevertheless, the way that people interpret a pictogram can be extremely variable because of the set of pictograms used, the cultural background, and the meaning of the pictogram (the grammatical ones are more complex to understand because they are less iconic).

Pictograms, thanks to their iconicity, can help people to communicate in a foreign country when they do not speak the local language and do not share any linguistic background with local inhabitants. As Rada Mihalcea and Chee Wee Leong have shown in 2009 in *"Toward communicating simple sentences using pictorial representations"*, pictogram translations can help people who do not share the same language to communicate.

However, in order to learn how to build sentences using pictograms and to increase the size of the speaker's vocabulary, it is necessary to have a rich input of pictogram sentences from the family (Beukelman and Mirenda, 2017).

Communication boards, paper-based or electronic medium, are used to encode these sentences. Finding the required pictogram in a communication board is an uneasy task. The family has to learn how to use the communication tool and, if it is a physical communication board, they have to spend time to look for the relevant pictogram. Because of this complicated navigation, interaction is not spontaneous and can even be perceived as really negative.

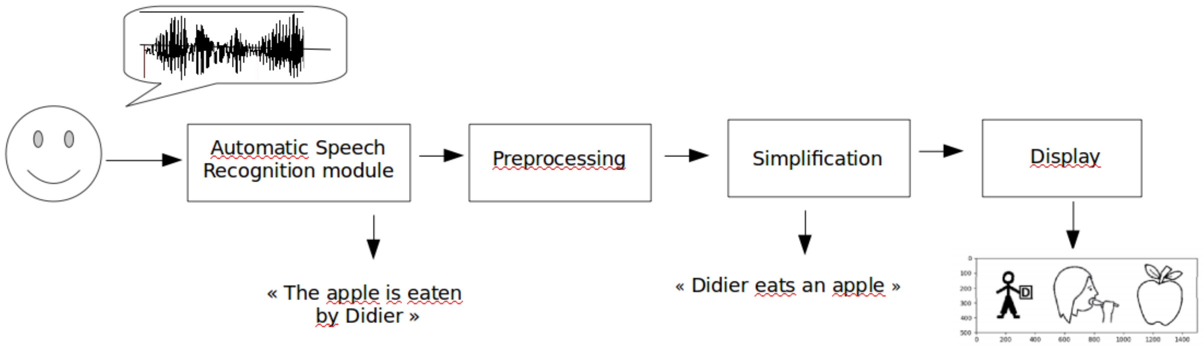
Automatic Speech Recognition module
Preprocessing
Simplification
Display
« The apple is eaten by Didier »
« Didier eats an apple »

Figure 1. Method of automatic pictogram generation from speech.

2 An automatic pictogram generation tool

An automatic pictogram generation tool working with everyday speech is a good way to solve this problem. Such a tool allows people close to the user of an AAC method to speak with their own language without necessarily having to learn how to encode pictogram sentences and without losing time to find pictogram in a communication board. It gives a better access to school for AAC users. If a text-to-speech tool is also used, it becomes possible for students using AAC to communicate with teachers or other students. A social bond can be created, with possibilities of mutual help, which leads to a positive learning environment.

Pictogram generation allows to overcome the language barrier between people and can allow people to join a school or a training course more easily than before.

3 Methodology

Our methodology (Figure 1) is based on the work of Vandeghinste and al (2017) and their Text2picto project. We propose 2 modules in order to generate pictograms from speech. The first one is the Automatic Speech Recognition (ASR) system and the second one generates a simplified message.

These modules result from the studies of the translation strategies of text into pictograms in a corpus collected on the web. In this corpus, grammatical words are often removed, as well as adverbs. Translating every word does not improve the comprehension of the text (maybe except for the mild disabilities or for people who knows already the structures of oral language) .

The granularity level of the translation must be adjustable to be adapted to each situation and each disability. Besides the syntactical structure can been changed to clarify the role of each phrase: when the passive voice is used, or when a sentence is included in another one, syntactical roles are not always easy to define.

The next sections detail the 2 modules.

3.1 Automatic Speech Recognition

We propose to use an Automatic Speech Recognition (ASR) module allowing to work directly with the voice. It takes a speech signal and transforms it into an orthographic transcription. The ASR model is based on a hybrid HMM-DNN model, developed by (Elloumi and al, 2018) with KALDI toolkit (a free ASR system) (Povey and al., 2011).

3.2 Simplification module

The next step consists in analyzing the syntactical structure of the sentences and to get the lemma of each words (the canonical form of a word). After this preprocessing of the transcriptions, the sentence is simplified. A simplification is necessary because a literal translation of a sentence into pictograms might be unintelligible for people with cognitive or mental impairments. This simplification can also help foreign people who do not master the language of the country they lived in. Two different simplification methods are proposed.

The first method is a syntactical simplification: as recommended by the Pathways project which have developed European Easy-to-Read, it is easier to understand simple sentences, in active voice, than long sentences and passive voice. We have implemented a passive-to-active sentence transformer which finds passive sentences and simplify them into active voice to be sure that everyone understands “who does what”.

Our second simplification method defines two levels of translation, one which translates every word and the other one which does not translate determiners and adverbs. It will be easier for people with symbolisation problems to understand the sentences without these linguistic units as they are not part of the core meaning of the sentence. For foreigners, using grammatical units which work quite differently in their native language can be difficult. Hence, keeping only the semantically relevant units to encode a sentence into pictograms seems to be better.

4 Evaluation

To evaluate the performances of our system we have created 2 evaluation tracks: one assessing the quality of text-to-pictogram, and the other one assessing speech-to-pictogram. For the first corpus, we have gathered six children stories copyright-free that we have manually translated in pictograms following strict guidelines built from our study of translation strategies. We have also manually created a simplified version of these stories by deleting articles, the verb be and some adverbs as we saw in our study of translation strategies.

The second corpus created to evaluate speech-to-picto contains twenty sentences extracted from audio recordings taken from the “Books for children” (a module of the ESLO corpus). These sentences are directly translated into pictograms, without preliminary orthographic transcription.

The choice of creating our own evaluation data was motivated by the fact that texts already translated in pictograms are hard to find and difficult to process. Besides most of these resources use proprietary sets of pictograms. Thus, we had to build our own evaluation corpus, with comparable data (same syntactical structures and vocabulary) such as poems and lullabies.

The evaluation of translation performances will be both qualitative with human judgments and automatic (BLEU [Papineni & al, 2002], WER...).

Nevertheless, using ASR implies some problematics that are important to consider such as noise in data, impact of the ASR errors on the other modules… The evaluation will measure how the performances of the ASR can affect the other ones.

First results for text-to-picto

We are able to present here only our first results for text-to-picto. Only BLEU score of text-to-picto have been calculated. Indeed, our first experimentation with speech recognition has obtained a Word Error Rate of 70%.

We can explain these results by the fact that spontaneous speech has many characteristics that complicate speech recognition (superimposed speech, disfluencies, poor acoustic conditions, etc.).The best speech recognition systems today get 40% WER on semi-prepared speech but the state-of-the-art performances on spontaneous speech is well below. Because of this preliminary result on ASR, we did not evaluate our prototype from speech because the results would have been catastrophic.

For text-to-picto, our prototype obtains a BLEU score of 26,65 when all the words are translated and 19,91 when the text is simplified (some grammatical words are deleted). This evaluation highlights the difficulties encountered in the task of simplifying text. Indeed, it is particularly difficult to identify grammatical words that can be deleted from those whose deletion may significantly change the meaning of the message. When we have built our simplified evaluation corpus, many complex cases in the deletion of adverbs caused us problems, and many adverbs had to be kept in order not to modify too much the meaning of the text.

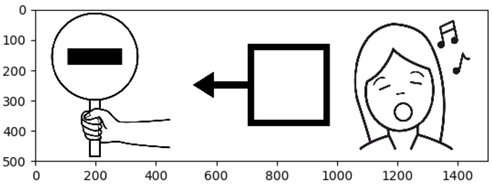

Figure 2.“It’s forbidden to sing here”: Simplified translation with our prototype.

In this example we can see that when the prototype simplifies the sentence, it deletes the adverb “here”. But without this adverb, the meaning of the sentences changes a lot. In our evaluation corpus, in this case we have chosen not to delete the adverb to keep the meaning.

The removal of all adverbs by our level 1 prototype therefore explains the lower results.

5 Research opportunities

5.1 Word Sense Disambiguation

To improve the results obtained by our pictogram translation model, we have identified several possibilities: first, the addition of a lexical disambiguation model, to determine the meaning of a word in context, could prevent the display of an irrelevant homonym.

To avoid generating the pictogram of a mouse (the animal) instead of a computer mouse in sentences like “The mouse of my computer is broken”, we will annotate the input sentences with a neural model that assigns a WordNet ID to every word (Vial & al, 2018). WordNet is a free lexical database in which words are “grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept” (George A. Miller, 1995). After this annotation step, our tool will query a database, developed by our team, in which each pictogram has been assigned a WordNet ID. When the most relevant sequence of pictograms is found, they are then displayed to the user.

5.2 Re-training of ASR system

The improvement of the performance of our recognition system will involve a re-training of the system from an oral corpus presenting spontaneous speech in daily interaction. Indeed, to better fit to our use case, we will retrain the ASR system with the ESLO [Eshkol-Taravella and al, 2011], which contains spontaneous speech (more linguistic facts of spontaneous speech like disfluencies, bad acoustic conditions and speakers overlapping (Dufour, 2010)).

5.3 Lexical simplification

Another possible improvement would be the addition of lexical simplification to our syntax simplification module. Indeed, the complex vocabulary, often absent from pictogram sets, is not translated by our prototype. A process of simplifying these words would therefore be really relevant, both to help the complete display of the user's sentences and to help individuals in situations of language disability understanding them.

6 Conclusion

In this paper we proposed a tool allowing to translate speech into pictograms. We address several issues that are linked to this technology: disambiguation, simplification and evaluation. Finally, this tool, developed in French and with the Arasaac set of pictograms, might improve the quality and the frequency of the input in pictograms which accelerate the acquisition of pictogram encoding, allows to break language barriers and can facilitate the access to school or work. It can be adapted for other languages and other set of pictograms. We plan to test our tools with real users and gather their reviews to highlight what we have to improve. These tests will measure the impact of such a technology on the acquisition of language of the AAC users.

References

Beukelman D. R. and Mirenda P. 2017. *Communication alternative et améliorée, Aider les enfants et les adultes avec des difficultés de communication*, 1er édition, DE BOECK SUPERIEUR, 400 p, collection Apprendre et Réapprendre

Cataix-Nègre E. 2017. *Communiquer autrement, Accompagner les personnes avec des troubles de la parole ou du langage*, 2ème édition, DE BOECK SUPERIEUR, 336 p, collection Pratiques en réeducation

Duboisdindien G. 2014. *L'interprétation des pictogrammes. Statut linguistique et limites de l’utilisation des pictogrammes dans la réhabilitation langagière. - Étude de deux groupes d’enfants âgés de 5 à 6 ans – entraînés Versus non entraînés*, mémoire de recherche de Master de Linguistique Générale et Appliquée Spécialité Fonctionnements Linguistiques et Dysfonctionnements langagiers, sous la direction de BOGLIOTTI Caroline, Université Paris-Ouest Nanterre La Défense, 100 p.

Dufour R. 2010. *Transcription automatique de la parole spontanée,* Thèse de doctorat en informatique sous la direction de DELÉGLISE P. , Soutenue à l’UFR de sciences exactes et naturelles du Mans

Elloumi Z., Besacier L., Galibert O., Kahn J. & Lecouteux B. 2018. *ASR Performance Prediction on Unseen Broadcast Programs using Convolutional*

Eshkol-Taravella I., Baude O., Maurel D., Hriba L., Dugua C. & Tellier I. 2012. Un grand corpus oral « disponible » : le corpus d’Orléans 1968-2012., in Ressources linguistiques libres, TAL. Volume 52 – n° 3/2011, 17-46

Miller G., 1995. Wordnet: A lexical database. Actes de Acm 38, pp. 39-41.

Papineni, K., Roukos, S., Ward, T., Zhu, W. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation.

Povey D., Ghoshal A., Boulianne G., Burget L., Glembek O., Goel N., Hannemann M., Motlicek P., Qian Y., Schwarz P., Silovsky J., Stemmer G. & Vesely K. 2011. *The Kaldi speech recognition toolkit,* IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, Hilton Waikoloa Village, Big Island, Hawaii, US

Vandeghinste, V., Schuurman, I., Sevens, L. & Van Eynde, F. 2017. Translating Text into Pictographs. Natural Language Engineering 23 (2):217-244

Vial L., Lecouteux B. & Schwab D. 2018. Approche supervisée à base de cellules LSTM bidirectionnelles pour la désambiguïsation lexicale. 25e conférence sur le Traitement Automatique des Langues Naturelles.