



HAL
open science

What you need is what you get: adapting word prediction of Augmentative and Alternative Communication aids to youth language

Chérifa Ben Khelil, Frédéric Rayar, Jean-Yves Antoine, Lisa Hoiry, Mathieu Raynal, Anais Anais Lefeuvre-Halftermeyer

► To cite this version:

Chérifa Ben Khelil, Frédéric Rayar, Jean-Yves Antoine, Lisa Hoiry, Mathieu Raynal, et al.. What you need is what you get: adapting word prediction of Augmentative and Alternative Communication aids to youth language. 25th International Conference on Human-Computer Interaction (HCI 2023), Jul 2023, Copenhagen, Denmark. pp.240-247, 10.1007/978-3-031-35992-7_33 . hal-04190258

HAL Id: hal-04190258

<https://hal.science/hal-04190258>

Submitted on 29 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

What you need is what you get: adapting word prediction of Augmentative and Alternative Communication aids to youth language

Cherifa Ben Khelil¹, Frédéric Rayar¹, Jean-Yves Antoine¹, Lisa Hoiry²,
Mathieu Raynal², and Anaïs Halftermeyer³

¹ LIFAT - University of Tours, France cherifa.bk@gmail.com

² IRIT - University of Toulouse, France

³ LIFO - University of Orleans, France

Abstract. This paper focuses on the development of prediction models for Augmentative and Alternative Communication (AAC) that are tailored specifically to child and young users. A new native-speaking corpus for child and adolescent languages is introduced, which allows the training of models that better fit the needs of this user category. The experiments conducted on real texts show that adapting the prediction models for children and young users leads to an improvement in keystroke savings, which in turn reduces the number of predicted words to be displayed. This opens up new possibilities for rethinking the virtual keyboard organization and interaction styles, with the aim of reducing cognitive load during text entry.

Keywords: Augmentative and Alternative Communication · word prediction · corpus building · user adaptation · Keystroke Saving Rate.

1 Introduction

Augmentative and Alternative Communication (AAC) aims at supplementing or replacing speech and/or writing for people suffering from serious speech and motion impairments. It concerns a large diversity of physical and/or cognitive disabilities among which locked-in syndrome, cerebral palsy, amyotrophic lateral sclerosis and other neurodegenerative diseases. These disabilities are frequently accompanied by associated language disorders. Whatever the impairment, oral communication is impossible for these persons who also have serious difficulties in physically controlling their environment. In particular, most of them are not able to use the standard input devices of a computer or with strong difficulties. To answer such special needs, a typical AAC system consists of a virtual keyboard from which text is entered using a dedicated control device, and a speech synthesis. Although AAC systems restore or facilitate the communication abilities of their users, text entry remains slow and tiring. Two complementary approaches can be considered to speed up typing: fast key selection through optimized keyboards design, and keystroke saving through word prediction. Predicted words

are usually displayed in a dedicated prediction list, allowing word completion and hence faster text entry. Additionally, word prediction can help in writing correct messages for people with additional language disabilities or without sufficient language competences (children, for instance). Prediction models rely on automatic machine learning techniques. They are trained on large corpora that are extracted from web or newspapers resources and are supposed to be representative of the communication language. Word prediction has proven to provide a sensible assistance during communication, even if its benefits in terms of input speed will depend strongly on the disability and the context of use [4]. The issue of the adaptation of word prediction to any kind of user remains however, largely open. Some AAC systems incorporate a user language model which is combined with a general one and allows the system’s behaviour to adapt progressively to the user’s input. It is important to note that if the user’s language skills differ greatly from the general language model used for prediction, a considerable amount of text may be necessary for the adaptation to take effect.

In particular, standard AAC word predictions fail to adapt to child and adolescent language, mainly due to the lack of a large child corpus on which a prediction model can be trained. An effective adaptation of prediction has a double beneficial impact for the user:

1. Words that belong to the user’s lexicon will be predicted more frequently.
2. Words that the user does not know will be removed from the prediction list, avoiding confusing predictions and limiting his cognitive load.

In this paper, we investigate the benefits of developing prediction models that are adapted to child and young users. At first, we introduce in section 2 a new French-speaking corpus that is dedicated to child and adolescent languages which allows the training of specific prediction models that fit better the needs of child or young users. In section 3, we outline the evaluation process used to assess the performance of our adapted models, while in section 4 we discuss experimental results obtained from analyzing real-world texts. These results highlight the benefits of our models in terms of keystroke savings and how they can improve virtual keyboard organization and interaction styles.

2 Adaptation for children and adolescents needs

The user adaptation of word prediction meets two complementary challenges:

1. *Age adaptation* - Prediction should adapt to the overall language fluency of the user. From a language learning perspective with young users, this adaptation may be achieved by creating general prediction models adapted to several language age groups.
2. *Personalization* - The prediction should also be adapted to the language preferences of each individual. For example, an adolescent who is interested in Formula 1 will appreciate a prediction promoting Formula 1 related words and phrases such as "red flag" and "formation lap". This individual adaptation will be achieved through a dynamic user prediction model combined

with the general one as well as better consideration of the current communication context by the prediction [6].

The adaptation can be limited to the system lexicon or to its whole language model. Lexicon adaptation can be manually handled by the user or speech therapists. However, the practical application of manual adaptation is progressively becoming more limiting. The automatic adaptation of the lexicon can be envisaged by considering the average frequency of occurrence of words in the general language. In a language learning perspective, this solution is not optimal, as the list of most frequent words changes with the age of the user. Additionally, lexicon adaptation can not allow a fine-tuning concerning styles and/or language registers. For these reasons, this paper focuses on adapting the language model of the prediction, and not just lexicon customization.

The development of age-based language models requires the use of adapted training corpora. Large child corpora that meet the needs of machine learning techniques are lacking for the French language. Taking these constraints into consideration, we decided to collect a corpus of texts specifically for children and adolescents to train the language model on language patterns and vocabulary that are most relevant to them. This can result in a more accurate and effective word prediction system for children and adolescents who use AAC.

2.1 Corpus and language model building process

1. **Text collection:** We begin by collecting a large number of relevant texts for children (6-11 years) and adolescents (12- 17 years) using our scraping module. This includes news websites, Vikidia⁴, books, and other copyright-free stories that are available online. The texts are varied in terms of content, style, and genre, to ensure that the corpus is representative of the language patterns and vocabulary used by children and adolescents. As can be seen in Table 1, there were a total of 58 237 files extracted (18 601 471 words).

Table 1. Distribution of the collected texts according to source type.

Text source	Number of files	Number of words
Books and stories for Children	2 249	2 987 285
Children news sites	16 077	5 125 990
Adolescents news sites	20 422	976 777
Vikidia	17 240	7 403 234
Total	58 237	18 601 471

2. **Text preprocessing** - Once the texts were collected, we pre-processed them, using natural language processing (NLP) techniques, to remove unwanted characters and words.

⁴ Vikidia is a online encyclopedia adapted for children and adolescents.

3. **Text tokenization** - The next step was to split the text content into smaller units, namely words or multiword units, which are referred to as tokens. These latter are used to build our language models.
4. **Language model training** - The final step was to train age-based language models on the preprocessed and tokenized corpora. Our experiment was conducted on a stochastic N-gram prediction model, *Predict4all* (subsection 3.1). The fundamental premise of this approach is to use probabilistic analysis to determine the co-occurrence of words in a sequence of n tokens. This acquired knowledge is then applied to generate a ranking of likely following words for any sequence of n-1 tokens. This procedure can be followed to any machine learning techniques. We produced one model for adolescents *Adolescents_all* and three models for children *Children_stories*, *Children_info* and *Children_all*. *Adolescents_all* was trained with texts extracted from news sites for adolescents and Wikidia, resulting in a 4-gram model based on a vocabulary of 67 190 unique words. As for the models for children, they were trained as follows: *Children_stories* with texts collected from books and stories for children, *Children_info* with texts collected from news sites for children and *Children_all* with with all the texts collected for Children. These models' vocabulary sizes are 45 249, 42 982 and 67 075 distinct words.

3 Experimental evaluation

3.1 PREDICT4ALL

The experiments reported in this paper were conducted on the *Predict4All* prediction module. *Predict4All* is an open-source library⁵ that can be integrated into any AAC system (and actually any other application). In particular, the *LifeCompanion*⁶ AAC system developed in the *AAC4All*⁷ project of the French ANR research agency integrates the *Predict4All* prediction. As explained before, *Predict4All* is based on a standard 4-gram prediction model. It incorporates a general prediction language model combined with a dynamic user model trained on-the-fly on the messages entered by the user⁸. The general prediction model has been trained on more than 20 millions words from Wikipedia and subtitles (of films and series) corpus. It has a vocabulary of 111 880 unique words. *Predict4All* offers also the possibility of a semantic adaptation of the prediction [7] according to the current of communication. It was co-designed with speech therapists and occupational therapists and proposes an on-the-fly correction module that is fully integrated with the prediction. This correction module is particularly dedicated to young users and/or users with language disorders.

As our experiments focused specifically on the benefits of age-based prediction models, they were conducted with a minimal setting of *Predict4All*, e.g. with no user dynamic model, nor semantic adaptation and nor correction.

⁵ <https://github.com/mthebaud/predict4all>

⁶ <https://lifecompanionaac.org/>

⁷ <https://www.aac4all.org/en/home/>

⁸ This dynamic adaptation must be explicitly authorized by the user

3.2 Evaluation metric

The Keystroke Saving Rate (KSR) is a performance indicator used to assess the effectiveness of different text input methods, including those that use word prediction and/or auto-completion. It is usually measured by comparing the amount of keystrokes needed to accomplish a given task with and without the use of an assistive tool. The higher the keystroke saving rate, the more effective the method is considered. KSR is especially useful when it comes to evaluating text input tools for people who have trouble using a standard keyboard, but it is important to note that this metric does not take into account other factors such as typing speed, accuracy, user satisfaction, etc. To compute the KSR score, we calculate the ratio of keystrokes that were accurately predicted to those that were not predicted, based on the length of the list of suggested words. This formula is typically represented as follows:

$$\text{KSR} = \left(1 - \frac{\text{keystrokes with prediction}}{\text{keystrokes without prediction}}\right) \times 100 \quad (1)$$

When using the KSR metric to evaluate the effectiveness of word prediction in AAC software for people with disabilities, it is important to distinguish between theoretical KSR and real KSR. Theoretical KSR represents the highest achievable score if users were to consistently choose words from the predicted list whenever a correct prediction is made. In contrast, real KSR reflects the actual performance, which includes any inaccuracies or errors that may be caused by the user while entering text. In this paper, we calculated the theoretical KSR as a way of assessing the performance potential in using our models.

3.3 Evaluation procedure

We carried out an evaluation of our child- and adolescent-dedicated models for word prediction with the goal of assessing their performance and comparing them to the general prediction language model. To achieve this, we used a diverse range of test corpora that are catered to different age groups. Specifically, the test corpora included nine sets of texts, with the first three intended for adults (texts from online journal articles, lifestyle blog posts, and website threads), one for adolescents (texts from online newspaper articles specifically intended for adolescents), and the last five for children (a collection of children’s stories and fables, as well as essays written by children). We recorded KSR scores for these nine test corpora, with predicted word list lengths (N) ranging from 1 to 5. By analyzing the KSR scores at different predicted word list lengths, we can evaluate the impact of the prediction strategy on performance and determine the most effective approach for different age groups and types of text. This information can guide decisions about the design of virtual keyboards and improve the user experience for young users.

4 Results and discussion

4.1 Results

Table 2. Performance evaluation of our child- and teenager-friendly models and the general one on various test corpora: KSR analysis with varying prediction list lengths.

Test Corpus	Models	Word prediction list length				
		<i>N=1</i>	<i>N=2</i>	<i>N=3</i>	<i>N=4</i>	<i>N=5</i>
1- Newswire from the site "Le monde"	<i>Children_stories</i>	29.128%	37.265%	41.235%	44.009%	45.763%
	<i>Children_info</i>	33.658%	42.05%	46.116%	48.84%	50.42%
	<i>Children_all</i>	34.945%	43.241%	47.331%	49.913%	51.717%
	<i>Adolescents_all</i>	35.031%	43.533%	47.675%	50.334%	51.955%
	<i>General model</i>	38.329%	46.846%	51.015%	53.524%	54.968%
2- Text from a lifestyle blog post	<i>Children_stories</i>	35.338%	43.686%	48.154%	51.059%	52.692%
	<i>Children_info</i>	37.973%	45.713%	50.288%	52.877%	54.345%
	<i>Children_all</i>	39.301%	47.240%	51.601%	54.496%	56.120%
	<i>Adolescents_all</i>	39.782%	47.919%	51.991%	54.447%	56.074%
	<i>General model</i>	40.870%	49.308%	53.411%	56.036%	57.483%
3- Excerpt from thread on reddit	<i>Children_stories</i>	31.307%	38.872%	42.909%	45.741%	47.350%
	<i>Children_info</i>	34.932%	43.325%	47.608%	49.926%	51.732%
	<i>Children_all</i>	36.293%	44.849%	48.983%	51.418%	52.853%
	<i>Adolescents_all</i>	36.068%	44.982%	49.206%	51.726%	53.249%
	<i>General model</i>	39.482%	47.659%	51.864%	54.195%	55.926%
4- Excerpt from teenager's news sites	<i>Children_stories</i>	33.046%	41.269%	45.261%	47.955%	49.725%
	<i>Children_info</i>	35.544%	43.553%	47.721%	50.219%	51.788%
	<i>Children_all</i>	37.477%	45.700%	49.610%	52.132%	53.757%
	<i>Adolescents_all</i>	37.507%	45.586%	49.669%	52.081%	53.669%
	<i>General model</i>	38.922%	47.210%	51.285%	53.697%	55.253%
5- Excerpt from the children's story: "Sophie's Misfortunes"	<i>Children_stories</i>	38.301%	46.791%	51.029%	53.662%	55.546%
	<i>Children_info</i>	34.288%	42.442%	46.591%	49.211%	51.099%
	<i>Children_all</i>	38.577%	46.893%	50.985%	53.601%	55.426%
	<i>Adolescents_all</i>	34.764%	43.008%	47.296%	50.022%	51.881%
	<i>General model</i>	36.827%	45.175%	49.483%	52.069%	53.974%
6- Excerpt from The Fables of La Fontaine	<i>Children_stories</i>	33.097%	41.021%	45.36%	48.262%	50.134%
	<i>Children_info</i>	28.651%	37.001%	41.554%	44.575%	46.509%
	<i>Children_all</i>	34.925%	43.128%	47.128%	49.797%	51.567%
	<i>Adolescents_all</i>	29.745%	38.087%	42.646%	45.613%	47.532%
	<i>General model</i>	31.781%	40.305%	44.814%	47.800%	49.563%
7- Excerpt from children's books and stories	<i>Children_stories</i>	37.770%	46.018%	50.210%	52.762%	54.402%
	<i>Children_info</i>	32.49%	40.904%	45.056%	47.577%	49.215%
	<i>Children_all</i>	38.311%	46.630%	50.746%	53.182%	54.851%
	<i>Adolescents_all</i>	33.931%	42.199%	46.523%	48.954%	50.661%
	<i>General model</i>	36.161%	44.624%	49.062%	51.491%	53.249%
8- Collection of essays written by children (6-10 years)	<i>Children_stories</i>	40.859%	50.00%	54.365%	57.167%	58.885%
	<i>Children_info</i>	37.535%	46.65%	51.054%	53.692%	55.382%
	<i>Children_all</i>	41.235%	49.895%	54.285%	56.735%	58.687%
	<i>Adolescents_all</i>	38.031%	46.186%	50.192%	52.205%	55.095%
	<i>General model</i>	40.409%	48.595%	53.365%	55.359%	57.285%
9- A story written by a 11 year old child	<i>Children_stories</i>	35.885%	44.598%	48.693%	51.436%	53.227%
	<i>Children_info</i>	31.965%	39.917%	43.928%	46.927%	48.607%
	<i>Children_all</i>	36.159%	44.807%	48.75%	51.393%	53.129%
	<i>Adolescents_all</i>	33.25%	41.254%	45.678%	48.244%	50.239%
	<i>General model</i>	35.086%	43.799%	47.957%	50.621%	52.33%

Table 2 summarizes the recorded KSR scores for the nine test corpora. The scores are organized according to the length of the predicted word list (N). The results indicate that the general model offers the highest KSR scores for the first four corpora, which are intended for adults and adolescents. These scores, obtained from the general model and the adolescent model, differ by approximately 3%, while the difference between the scores obtained from the general model and the children’s models can be as high as 8%. On the other hand, when it comes to the test corpora intended for children, the general model typically yields slightly lower scores than the dedicated children’s models. Specifically, the results show that the *Children_stories* and *Children_all* models offer the best scores for the last five test corpora. The difference between these models and the general model ranges from 1% to 3%.

4.2 Discussion

The experiments conducted on real texts show the potential of prediction models adapted for young users and that a general model trained on more than 20 millions words from a variety of texts may not be the best fit for specific audience groups. These findings offer a promising approach to facilitate the adaptation of AAC systems for young users. Instead of relying on a general model with a large vocabulary and combining it with a dynamic user model, it is now possible to start with a more specific general model that takes into account the user’s age, and then merge it with the user’s dynamic model. This approach may offer greater flexibility as it enables the AAC system to adapt more quickly and accurately to the user’s unique vocabulary and language usage patterns. As the specific general model has a more constrained vocabulary, the combination with the dynamic user model is expected to yield better results, allowing for a more rapid and effective adaptation to the user’s language use patterns. Additionally, the reduction in vocabulary size can have a positive impact on the memory requirements of prediction models, making them more practical to deploy on low-resource devices, such as mobile phones or tablets.

Another important finding from this study is that the adaptation of these models improves keystroke saving rates, which can reduce the number of predicted words displayed on the virtual keyboard. This finding opens up new possibilities for the organization of virtual keyboards. Traditionally, recommendations for the number of words in the prediction list are between 5 and 7 words [1,2,3]. However, recent research [5] has shown that users of pointing keyboards, a type of virtual keyboard that relies on eye gaze or other pointing methods for text entry, only look at the top of the prediction list and tend to choose the predicted word when it appears in the first or second position. Based on the observed improvements, we can consider reducing the size of this list to 3 words. This would make the interface lighter and thus reduce the cognitive load during the text input. There is also another type of virtual keyboard, scanning keyboards. It uses a scanning method to cycle through the keyboard keys, allowing users to select letters or other input options at the desired row and/or column. It is reasonable to expect that reducing the number of predicted words displayed

would also improve the efficiency and speed of this navigation method. Limiting the number of displayed options can lead to an increase in words per minute (WPM) for users, as it reduces the scanning step.

5 Conclusion

In this paper, we investigated the effectiveness of customizing word prediction models for youth. Four models were trained on different types of text intended for children and adolescents. Our results suggest that the accuracy of these models can vary significantly depending on the intended audience and the specific text they are trained on. Especially, for texts intended for children or written by children, the specialized children's models produced better results. These findings highlight the importance of developing specialized models for specific audience groups, as this can improve the accuracy of word prediction systems. Our first experiment primarily focused on written texts and it remains to be seen whether the findings can be extended to spoken communication scenarios. To address this, other ongoing work is in progress to test the models on transcripts of discussions. Furthermore, a clinical study will be planned to further investigate the potential of these models to improve communication with an AAC system for the young users in real-life situations.

Acknowledgments

This research was funded by the French National Research Agency as part of the AAC4ALL project (ANR-21-CE19-0051).

References

1. Heinisch B, Hecht J. Predictive word processors: a comparison of six programs. *Tam News* 8:4–9, 1993.
2. Norman D.A, Fisher D. Why alphabetic keyboards are not easy to use: Keyboard layout doesn't much matter. *Human factor*, 24: 509-519, 1982.
3. Swiffin A, Arnott J, Pickering J, Newell A. Adaptive and predictive techniques in communication prosthesis, *Augmentative Alt. Communication* 3:181–191, 1987.
4. Pouplin S, Roche N, Antoine JY, Vaugier I, Pottier S, Figere M, Bensmail D. The effect of word prediction settings on text input speed in persons with cervical spinal cord injury: a prospective study. *Disabil Rehabil.* 2017 Jun;39(12):1215-1220.
5. Raynal M, and Badr G. Study of User Behavior When Using a List of Predicted Words. In *Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022, Lecco, Italy, July 11–15, 2022*, 331–337.
6. Wandmacher T, Antoine J-Y, Poirier F. Sibylle : a system for alternative communication adapting to the context and its user. *Actes ACM Conference on Assistive Technologies. ASSETS'2007, Phoenix. Arizona. 2007.203-210*
7. Wandmacher T, Ovchinnikova E, Alexandrov F. Does latent semantic analysis reflect human associations?. *20th European Summer School in Logic, Language and Information, 04.08.-15.08.2008.Hamburg, Germany.2008 p.63-70*