

# **Evaluation of a Word Prediction System in an Augmentative and Alternative Communication for Disabled People**

M. Norré

Cental/ILC, Catholic University of Louvain (UCLouvain), Louvain-la-Neuve, Belgium

Corresponding Author Email: magali.norre@uclouvain.be

https://doi.org/10.18280/mmc\_c.811-409

ABSTRACT

Received: 15 July 2020 Accepted: 22 November 2020

#### Keywords:

word prediction, language modeling, augmentative and alternative communication, AAC, usability, disability This article investigates the evaluation of a word prediction system in an Augmentative and Alternative Communication (AAC) software for disabled people. In addition to having a reduced mobility, these users have an altered use of speech that must be compensated by a technological aid offering input methods adapted to their capabilities. To improve their communication speed, different prediction and language modeling techniques are used. We present the parameterization of statistical predictors. Their configuration in French is evaluated by a simulator and tested by a disabled person. The results show that a language model built from a large literary corpus saves more than one keystroke out of two, the performance of these systems varying according to several parameters.

### **1. INTRODUCTION**

In a society where the circulation of information and communication has become a real issue, the gap with disabled people who have both significant difficulties in moving and communicating could prove to be more problematic. Nevertheless, much research has been carried out for years in the field of assistive technologies. Some of them lead to operational prototypes that are sometimes marketed. People whose communication is impaired now have the possibility to obtain these systems of Augmentative and Alternative Communication (AAC). These methods are called augmentative when they make it possible to supplement the skills already present, which may even help the emergence of oralization for certain individuals; and called alternative when they completely replace the means of oral expression [1]. The systems we will evaluate are more suitable for paralyzed people with severe speech impairment, but without an intellectual disability, such as locked-in syndrome.

Available on computer or tablet, most of them integrate technologies such as text prediction, virtual keyboards and speech synthesis to communicate. This article aims to present the functioning of prediction engines, these systems used in mobile phones that suggest words based on letters already entered. This technique, well known to the general public, allows the user to save required keystrokes to input a word and speed up the typing. Since some disabled users cannot use a keyboard or mouse, the use of these systems is necessary and ubiquitous in many modern alternative communications. The state of the art (section 2) will review their characteristics. We will then detail the two systems constituting our case study: the Presage prediction engine [2] integrated into the ACAT AAC software [3] (section 3).

# 2. STATE OF THE ART

First, we introduce some existing word prediction and AAC

systems (section 2.1), as well as several limitations regarding their evaluation (section 2.2).

# 2.1 Word prediction and AAC

Before being used by the general public in limited interfaces such as mobile devices, word prediction systems have been initially designed for disabled people [4]. Compared to speech rates in oral communication generally estimated at 150 to 200 words per minute [5, 6], AAC tools do not exceed 10 to 15 words per minute. In order to improve the speed of text entry and reduce user fatigue, there are various techniques and a lot of research on text prediction [7, 8]. Prediction systems are based on contextual information: statistical or linguistic. To model the language and adapt it to the user, there are different types of approaches which can nevertheless be complementary and concern both letter prediction and word prediction.

Statistical systems model language using frequencies and Markovian n-gram models – bigram: sequence of two elements (words, letters or characters), trigram: three elements, etc. – which are sometimes associated with more advanced stochastic techniques such as smoothing or interpolations. It is possible to adapt this type of systems from one language to another, but also to a more specific vocabulary or language register. There are several free or paid virtual keyboards available. Many of these predictors use a statistical approach. This is the case of PolyPredix marketed by AssistiveWare [9, 10]. Statistical prediction is used in the Sibylle system to predict words [11, 12] and letters [6].

The completion tools allow you to complete a word by offering only a suggestion based on the characters already entered, the choice being refined after each entry. When the user decides to select the proposed word by validating it, it is directly integrated into the text, which makes it possible to save the last keystrokes that would have been required to type the word [13]. Prediction systems display a list of words above or next to the keyboard. The VITIPI completion system [14]



displays word endings when there is no longer any ambiguity.

Prediction and completion systems can be integrated into a virtual keyboard. The user accepts the predicted words by clicking on them using a pointing device (e.g. a switch) combined with an iterative scanning system (often by block, row, and column). The word selection is done with a single click or with keyboard shortcuts. For example, the Soothsayer completion prototype [15] works with the space key, while the text editor of the Presage prediction engine [2] requires the mouse or the F1 to F6 keys to select the predicted words. The ACAT user interface [3] can display the word list both horizontally and vertically.

#### 2.2 Linguistic, cognitive and ergonomic limitations

Communication aid is a multidisciplinary field related to (computational) linguistics, cognitive psychology, and ergonomics [16]. Regarding the evaluation of word prediction systems, it is important to consider certain limitations concerning several aspects, such as language modeling in Natural Language Processing (NLP), the actual use of prediction by disabled people related to Human Computer Interaction (HCI), but also the contribution of different metrics. There are two types of evaluation: a subjective method which considers the opinion of potential users and an objective method focused on metrics which can concern either the endto-end system with the interface or only the prediction.

To evaluate typing speed, researchers sometimes calculate the average number of keystrokes or the number of words per minute. The results of different metrics often depend on user and ergonomics since a bad interface can cause an additional cognitive load that could, for example, slow down text production [17]. There is little research on the use of prediction systems in the long term. The HCI has an impact on the use of text entry aids. If the end-to-end system has sub-keyboards to access certain characters, more keystrokes will be required if the KSR is evaluated.

Classical metrics in NLP (e.g. perplexity or entropy of a model) are rarely used in AAC. Based on information theory, they are not explicit indicators of the help provided by the prediction [7]. To evaluate and adapt this system, researchers often implement a module that automatically simulates the best use (a perfect user), i.e. does not calculate the errors that the user would produce, but directly selects the correct prediction as soon as it is likely to appear in the list. It is an objective and theorical evaluation that shows the maximum capabilities of the system. It should be noted that in real-world testing, prediction systems impose a cognitive load on users [18]. We have chosen to carry out an objective evaluation and a subjective evaluation.

# **3. USE CASE**

The aim of this study was to objectively evaluate the Presage word prediction engine and its settings, before carrying out a subjective evaluation of this tool integrated into the ACAT AAC software. We therefore present the two systems that we have configured and evaluated (section 3.1), before explaining our methodology (section 3.2), our results (section 3.3) which will be the subject of discussion and we will consider certain ethical questions (section 3.4).

#### 3.1 Presage integrated into ACAT

Presage and ACAT systems were designed independently and are open source (Figure 1). We tested version 0.9.1 of Presage (2015) and version 1.0.0 of ACAT (2016).

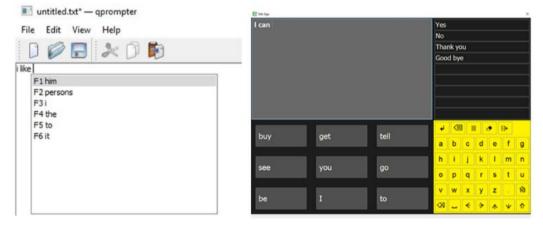


Figure 1. Presage and ACAT interfaces

Presage (formerly Soothsayer) is a word prediction system developed by Vescovi in the context of his thesis [2]. It consists of several statistical predictors and n-gram language models built from corpora. Presage proposes a list of the most probable words to the user.

Assistive Context-Aware Toolkit or ACAT is the user interface developed by Intel Corporation and initially designed for a person: physicist Stephen Hawking, who had amyotrophic lateral sclerosis [3]. Presage is included by default in the open source version. ACAT is an AAC software in English that includes two modes (Talk or App) with several virtual keyboards accessible via different input modalities. The scanning system allows the user who cannot use a mouse to select characters and words with an eyebrow, cheek or mouth movement detector via a webcam. Due to the disability of our test user, we did not use this input modality. Since the system is open source, language packs are available on GitHub.

### 3.2 Methodology

By default, Presage is based on a smoothed trigram language model trained on a copyright-free book. We have therefore built a training corpus and a test corpus. We used the French version of *Google Books Ngram* [19] from 2009 which we "tokenized" and processed (46,786,461 uni/bi/trigrams). We tested two other training corpora: one built from  $\hat{A}$  se tordre book by Alphonse Allais (38,123 words) and the other from *Perso*(nal) texts representing the idiolect of potential users (18,337 words), i.e. consisting of *Blog* articles (14,071 words) and an *Interview* (4,306 words), written by two disabled adults with an athetoid cerebral palsy.

The test corpus consisted of 100 simple sentences from familiar register that were not part of the training corpus, but from several sources to reflect different everyday uses. We used 40 general phrases from a corpus built for evaluating text entry techniques [20], which we translated manually into French. We have added sentences used for the evaluation of other AAC systems.

#### A) Objective evaluation

To evaluate the Presage prediction system, we designed an automatic simulator by adapting and extending the author's. We used three metrics: (1) the Keystroke Saving Rate or KSR, (2) the character saving rate that does not take into account the interface, (3) and the hit ratio.

$$KSR = \left(1 - \frac{k_i + k_s}{k_n}\right) \cdot 100 \tag{1}$$

In KSR (1), ki is the number of actual keystrokes, ks the number of keystrokes required to select suggestion from the Presage list (ks = 1 per word), kn the number of keystrokes required to compose the text with no prediction enabled. In this study, the character saving rate (2) is the percentage of characters that could be saved by using the system. It is similar to (1), but ks is not calculated. The hit ratio (3) is the rate of use of the prediction.

#### **B)** Subjective evaluation

To evaluate Presage integrated into ACAT, we conducted a user test with a disabled person with an athetoid cerebral palsy. According to questionnaire, he regularly used AAC systems incorporating word prediction. The participant was an adult who does not speak, has no intellectual disability and has a good level of French. We asked him to copy 20 sentences from the test corpus and to use the prediction as soon as the correct word was displayed in the list.



Figure 2. Presage integrated into ACAT: user test

The participant could also use certain abbreviations via an expansion mechanism. One of the instructions was to use them according to the possibilities of the text. To select characters and predictions, he used the mouse with his feet (Figure 2), because his face twitching did not allow him to use the movement detector. After the experiment we recorded with the Morae tool, he completed questionnaires and we used the System Usability Scale or SUS [21].

Presage parameters have been determined according to the theoretical results obtained with our simulator. The user test lasted one hour: an installation and explanation session that lasted 10 minutes, followed by two sessions (with a break), each to transcribe 10 sentences. The first lasted 28 minutes and the second 20 minutes.

#### 3.3 Results

Automatic simulations showed that the smoothed trigram language model with linear interpolation trained on the *Google Books Ngram* corpus obtained the best performance and saved more than one keystroke out of two by proposing five predictions (Table 1).

Our results also showed that the trigram model trained on the *Google Books Ngram* corpus was better than a unigram model and a bigram model (Table 2).

In addition to the general language model trained on the *Google Books Ngram* corpus, we have added a user language model (Table 3). We can therefore assume that the user has manually typed the texts of the *Perso* corpus, the *Blog* articles, or the *Interview* when the learning mode was activated. We note a slight improvement compared to Table 1.

Still trained on the *Google Books Ngram* corpus, we show that the higher the number of predictions proposed, the higher the results (Table 4). However, it should be noted that in an AAC system, it is necessary to consider the ergonomic and cognitive aspects of the users.

These different tests carried out on the same test corpus made it possible to identify the parameters to be implemented in ACAT for the user test, i.e. the smoothed trigram language model, trained on the *Google Books Ngram* corpus, associated with a user language model trained on the *Perso* corpus and displaying a list of 9 predictions.

User test recordings showed that the participant did not always select the correct prediction as soon as the word appeared in the list. Although he had the opportunity to use the abbreviations previously encoded in ACAT, he used them on four out of the ten occasions. The SUS score was 70% and therefore almost 'good' (threshold at 73%) [21]. During the

test, we observed that several predicted words were badly contextualized. The participant also noted that he had encountered bugs, but he admitted to being satisfied with the systems and the predictions.

Fable	1.	Result	s by	corpus
-------	----	--------	------	--------

	Google Books Ngram	A se tordre	Perso
KSR	57.9%	39.6%	39%
Character saving rate	75%	54.7%	53.1%
Hit ratio	90%	79.4%	74.5%

**Table 2.** Results by number of n-grams

	Unigram	Bigram	Trigram
KSR	43.8%	54.1%	57.9%
Character saving rate	60.3%	71.2%	75%
Hit ratio	87.2%	90.4%	90%

Table 3. Results with a user language model

	Perso	Blog	Interview
KSR	59%	58.7%	58.5%
Character saving rate	76.4%	76.1%	75.9%
Hit ratio	91.8%	91.5%	91.6%

Table 4. Results by number of predictions

	1	3	5	7	9
KSR	45.3%	51.2%	57.9%	60.1%	61.4%
Character saving rate	60.5%	67.5%	75%	77.5%	78.9%
Hit ratio	79.7%%	85.6%	90%	91.6%	92.2%

# **3.4 Discussion**

First of all, it should be noted that it is difficult to compare studies with each other, particularly because of heterogeneous methodologies. The test conditions are not always explicit, and some authors do not distinguish between two often confused metrics: the KSR and character saving rate. They provide different measures, whether the complete interface is evaluated or not. In our own experiments, the KSR is always worse than those obtained with the character saving rate, the difference varying from 15 to 20%. The hit ratio was generally high, i.e. over 70%.

On the Google Books Ngram corpus, we observed KSR and character saving rate around 50% and higher. The results corroborate those of several authors who obtain a KSR around 50% according to different corpora and parameters [4, 6, 11] [12, 17, 18]. We found a very large discrepancy of around 20% in the results obtained with the Google Books Ngram and the two other training corpora. This shows the need of a contextual adaptation of the prediction on the current situation of use of the system. We did not study deeply the question of the minimal amount of training data required to reach satisfactory performances. However, it is observed that the corpus of several million n-grams was better than the  $\hat{A}$  se tordre corpus consisting of a single book. We have also shown that a trigram language model is more efficient than a lower-order n-gram model as mentioned in studies [6, 22]. As for the combination of a general language model with a user language model, the larger the training corpus of user, the better the results. The improvements that one can expect from a user prediction model require however a frequent use of the system in order to obtain a sufficient amount of personal training data.

As for the number of word predictions displayed, we assumed that the higher the number of predictions displayed, the more likely the predicted word is the user's word. The theoretical results of the hit ratio confirm this hypothesis. We observed an increase of more than 10% of the hit ratio when the number of predictions displayed increases from one to five. As the improvement is less important with higher values (> 5), the choice should be made according to the user's preferences. Some authors [23] found an improvement in the theoretical performance, i.e. the percentage of keystrokes decreased by increasing the number of predicted words from one to ten. However, it is important to distinguish the results of an objective and theoretical evaluation from those of a real-life evaluation, as we must consider that the time and cognitive effort needed by disabled people to select a prediction increase if the number of word predictions displayed is high [8]. As mentioned in the literature [24], it is always necessary to find the right balance between maximizing performance and minimizing the cognitive load felt by the user, i.e. choosing between a more efficient system or a less tiring system.

The evaluation we have conducted suggests the benefits that can be expected from the use of AAC. Before concluding, it seems important to discuss these aid systems from an ethical point of view to highlight certain risks and their implications. Several studies [25, 26] are starting to address the issue of ethical risks related to digital technologies. They follow a consequentialist approach consisting in characterizing the risk factors associated with the use of digital tools. The developers of the Sibylle prediction system have shown that it can lead to the production of more texts and a decrease in spelling errors in children [11]. However, the authors question the impact on the user language skills of a constant use of their system. They wonder whether the aid promotes a faster mastery of the language system or whether the improvement in the quality of productions does not mask a loss of this capacity for the benefit of the system [26].

In other words, a risk associated with the use of these technologies as an assistance or compensation tool would be the loss of a skill supplemented by the technology over time. Users might indeed tend to rely more on machine's suggestions rather than their own analyses. Thus, although these systems are designed to reduce their handicap by allowing them to speed up their communication, they can favor a disabling language dependency and a loss of autonomy by helping them at all times in their spelling choices, such as automated spelling correctors. In this way, they replace the real knowledge of users. If we continue this reflection, the choice of the corpus and the words that the system will propose can also be considered as a risk factor. The possibilities of linguistic interaction offered by AAC systems are constrained by their lexicon, often with a justifiable goal: typing faster and not confusing the user with predictions that he would not understand. Researchers are facing a dilemma. If the prediction is beyond the user's language skills, it may confuse the user, which could prevent learning in children. But if it is too simple, it reduces her possibilities of expression below her ability and increases her handicap.

Other risk factors can also be considered. Inappropriate predictions could have a negative influence on the psychological state of the patient (nervousness, cognitive fatigue) [25, 26]. By testing *Google Books Ngram* corpus with a real user, we noted several badly contextualized lexical

proposals. For use in an ecological situation, our preprocessing phase was not sufficient given the number of words that will probably never be used. There would therefore be several risks associated with prediction. We studied a cognitive risk (regression) resulting in a loss of autonomy as well as a psychological risk linked to the feeling in relation to the use of prediction. It is important to note that the developers of prediction systems tend to think in terms of immediate, observable and objectively measurable gains (e.g. KSR), not to mention the long-term risks or possible more global effects on rehabilitation and the development of the disability [25].

## 4. CONCLUSION

This article has shown that Presage saves more than one keystroke out of two when the system is trained on the French version of the large *Google Books Ngram* corpus. Our results confirmed previous studies on prediction, i.e. a trigram language model is more efficient than a lower-order n-gram model and the more word predictions displayed, the better the KSR, the character saving rate and the hit ratio. It is also possible to achieve improvements with a user language model, but it takes time as the word prediction system must adapt and therefore be used frequently. The test of Presage integrated into ACAT with a disabled person focused on the interaction and satisfaction of a potential user. Nevertheless, the high diversity of disabilities addressed by AAC systems does not make it possible to draw conclusions that can be generalized to all disabled person.

Our user test showed that some predictions were unlikely to be used and they were not systematically chosen as soon as they appear in the word list. It is important to note that the actual use of predictions depends on the cognitive and visual load of the patient. In addition to physical effort, the user must concentrate on writing his text, avoid spelling errors, read the predictions, and use of his memory skills when he thinks he can use an abbreviation. Thus, despite the implementation of a mechanism to save more keystrokes (and therefore to compensate for the slowness of communication), the participant preferred not to use it. From an ergonomic point of view, the interface may also have an impact on results of some metrics used for prediction. In testing ACAT, for example, we found that the simple addition of a sub-keyboard containing the accented characters is likely to influence the KSR.

Finally, we can conclude that the integration of Presage in ACAT offers interesting possibilities. Our study was limited to a user test with a single person with a motor disability. However, we have identified a tremendous cognitive load and some ergonomic problems in carrying out the evaluation of the prediction system. User adaptation remains a central point when evaluating these technologies. It would be interesting to evaluate other disabled people and make these open source systems more accessible.

## ACKNOWLEDGMENTS

Presage is available online at SourceForge (http://presage.sourceforge.net/) under a GNU General Public License version 2.0 (GPLv2) and ACAT (https://01.org/acat) on GitHub (https://github.com/01org/acat) under an Apache License Version 2.0.

#### REFERENCES

- [1] Beukelman, D.R., Light, J.C. (2020). Augmentative & alternative communication: Supporting children and adults with complex communication needs. Paul H. Brookes Publishing Company.
- [2] Vescovi, M. (2004). Soothsayer: un sistema multisorgente per la predizione del testo. Master's thesis. Politecnico di Milano, Dipartimento di elettronica e informazione.
- [3] Denman, P., Nachman, L., Prasad, S. (2016). Designing for "a" user: Stephen Hawking's UI. In Proceedings of the 14th Participatory Design Conference: Short Papers, Interactive Exhibitions, Workshops-Volume 2, pp. 94-95. https://doi.org/10.1145/2948076.2948112
- [4] Antoine, J.Y. (2011). Prédiction de mots et saisie de requêtes sur interfaces limitées: dispositifs mobiles et aide au handicap. In Recherche d'information contextuelle, assistée et personnalisée, Recherche d'information et web, Hermès-Lavoisier, pp. 273-298. https://hal.archives-ouvertes.fr/hal-01016521/
- [5] Copestake, A. (1997). Augmented and alternative NLP techniques for augmentative and alternative communication. In Natural Language Processing for Communication Aids, pp. 37-42. https://www.aclweb.org/anthology/W97-0506.pdf
- [6] Schadle, I. (2003). Sibylle: système linguistique d'aide à la communication pour les personnes handicapées (Doctoral dissertation, Lorient). https://www.theses.fr/2003LORIS028
- [7] Antoine, J.Y., Maurel, D. (2007). Aide à la communication pour personnes handicapées et prédiction de texte. TAL, 48(2): 9-46.
- [8] Garay-Vitoria, N., Abascal, J. (2006). Text prediction systems: A survey. Universal Access in the Information Society, 4(3): 188-203. https://doi.org/10.1007/s10209-005-0005-9
- [9] Bérard, C., Niemeijer, D. (2004). Evaluating effort reduction through different word prediction systems. In 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE cat. no. 04ch37583), pp. 2658-2663. https://doi.org/10.1109/ICSMC.2004.1400732
- [10] Niemeijer D. (2005). In memoriam of Christian Bérard: Striving for effort reduction through on-screen keyboard word prediction. In Assistive technology: From virtuality to reality.
- [11] Wandmacher, T., Antoine, J.Y. (2007). Modèle adaptatif pour la prédiction de mots. TAL, 48(2): 71-95.
- [12] Wandmacher, T. (2009). Adaptive word prediction and its application in an assistive communication system. Thesis, Université François Rabelais de Tours.
- [13] Pouplin, S. (2016). Evaluation de l'efficacité des logiciels de prédiction de mots sur la vitesse de saisie de texte sur l'outil informatique pour les personnes blessées médullaires cervicaux (Doctoral dissertation, Université Paris-Saclay).
- [14] Boissière, P., Vigouroux, N., Mojahid, M., Vella, F. (2012). Adaptation of AAC to the context communication: A real improvement for the user illustration through the VITIPI word completion. In International Conference on Computers for Handicapped Persons, pp. 451-458. https://doi.org/10.1007/978-3-642-31534-3\_67
- [15] Stoop, W., van den Bosch, A. (2014). Improving word

prediction for augmentative communication by using idiolects and sociolects. Dutch Journal of Applied Linguistics, 3(2): 136-153. https://doi.org/10.1075/dujal.3.2.03sto

- [16] Blache, P., Rauzy, S. (2007). Le moteur de prédiction de mots de la Plateforme de Communication Alternative. Traitement Automatique des Langues, 48(2): 47-70. https://hal.archives-ouvertes.fr/hal-00285527/
- [17] Trost, H., Matiasek, J., Baroni, M. (2005). The language component of the FASTY text prediction system. Applied Artificial Intelligence, 19(8): 743-781. https://doi.org/10.1080/08839510500191901
- [18] Renaud, A., Shein, F., Tsang, V. (2010). Grammaticality judgement in a word completion task. In Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids, pp. 15-23. https://www.aclweb.org/anthology/W10-0403.pdf
- [19] Michel J.B., Shen Y.K., Aiden A.P., Veres A., Gray M.K., Pickett J.P., Hoiberg D., Clancy D., Norvig P., Orwant, J., Pinker, S., Nowak, M.A., Aiden, E.L. (2011). Quantitative analysis of culture using millions of digitized books. In Science, 331(6014), 176-182. https://doi.org/10.1126/science.1199644
- [20] MacKenzie, I.S., Soukoreff, R.W. (2003). Phrase sets for evaluating text entry techniques. In CHI'03 extended abstracts on Human factors in computing systems, pp. 754-755. https://doi.org/10.1145/765891.765971

- [21] Brooke J. (1996). SUS: A quick and dirty usability scale. In Usability Evaluation in Industry, 189(194).
- [22] Lesher, G.W., Moulton, B.J., Higginbotham, D.J. (1999). Effects of ngram order and training text size on word prediction. In Proceedings of the RESNA'99 Annual Conference, pp. 52-54.
- [23] Hunnicutt, S., Carlberger, J. (2001). Improving word prediction using markov models and heuristic methods. Augmentative and Alternative Communication, 17(4): 255-264. https://doi.org/10.1080/aac.17.4.255.264
- [24] Ghedira S. (2015). Optimisation de la communication pour personnes handicapées. In 22eme Conférence sur le Traitement Automatique des Langues Naturelles (TALN).
- [25] Antoine, J.Y., Lefeuvre, A., Allegre, W. (2014). Pour une réflexion éthique sur les conséquences de l'usage des ntic: le cas des aides techniques (a composante langagiere ou non) aux personnes handicapées. Journée ATALA" Ethique et TAL. https://www.schplaf.org/kf/pdf/2014\_ATALA\_ETHIQ UE\_TAL\_ANTOINE\_LEFEUVRE\_ALLEGRE\_VFIN ALE.pdf
- [26] Antoine, J.Y., Labat, M.E., Lefeuvre, A., Toinard, C. (2014). Vers une méthode de maîtrise des risques dans l'informatisation de l'aide au handicap. In Envirorisk'2014, Le Forum de la Gestion des Risques Technologiques, Naturels et Sanitaires, pp. 9. https://hal.archives-ouvertes.fr/hal-01213198/