

Configuration et évaluation d'un système de prédiction de mots au sein d'un logiciel de Communication Améliorée et Alternative (CAA)

Magali Norré et Max De Wilde
Faculté de Lettres, Traduction et Communication
Université Libre de Bruxelles, Belgique
Mnorre | madewild@ulb.ac.be

Résumé

Cet article présente la configuration et l'évaluation d'un système de prédiction de mots ainsi que d'un logiciel de Communication Améliorée et Alternative (CAA) pour les personnes handicapées. En plus d'avoir une mobilité réduite, ces utilisateurs ont un usage de la parole altéré qui doit être compensé par une aide technologique proposant des modalités de saisie adaptées à leurs capacités et permettant d'améliorer leur vitesse de communication. Le paramétrage de prédicteurs statistiques a été évalué par un simulateur et testé par une personne handicapée. Les résultats montrent qu'un modèle de langage lissé trigramme construit à partir du corpus Google Books Ngram permet d'économiser plus d'une saisie sur deux, les performances de ces systèmes variant en fonction de plusieurs paramètres.

1 Introduction

Dans une société où la circulation de l'information et de la communication est devenue un réel enjeu, l'écart avec les personnes en situation de handicap qui connaissent à la fois de sévères difficultés pour se déplacer et communiquer pourrait s'avérer davantage problématique. De nombreuses recherches sont néanmoins menées depuis des années dans le domaine du handicap et des aides logicielles. Certaines d'entre elles aboutissent à des prototypes opérationnels qui sont parfois commercialisés. Les personnes paralysées dont la communication est altérée ont désormais la possibilité de se procurer ces Communications

Améliorées et Alternatives (CAA). Disponibles sur ordinateur ou tablette, elles permettent à l'utilisateur de communiquer en utilisant un clavier virtuel associé à une prédiction de mots (Antoine and Maurel 2007; Garay-Vitoria and Abascal 2006) afin d'améliorer leur vitesse de saisie souvent réduite en raison de leurs capacités motrices. Le message peut alors être synthétisé.

Nous présenterons d'abord le système de prédiction Presage et l'interface ACAT (section 2), nous décrirons ensuite notre méthodologie (section 3), nos résultats (section 4) ainsi que certaines perspectives (section 5).

2 Presage et ACAT

Les deux outils que nous avons configurés et évalués ont été conçus de manière indépendante et sont tous les deux open source.³⁹ Nous avons testé la version 0.9.1 (2015) de Presage et 1.0.0 (2016) d'ACAT.

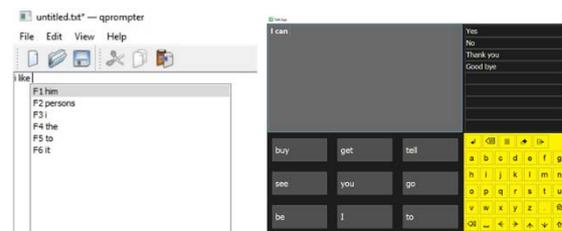


Figure 1. Interface de Presage et ACAT.

Presage (anciennement Soothsayer) est un système de prédiction de mots développé par un chercheur dans le cadre de sa thèse (Vescovi 2004). Il est constitué de plusieurs prédicteurs statistiques ou modèles de langage n-grammes. Presage fonctionne à partir de corpus encodés sous la forme de base de données. Il propose une liste de mots les plus probables à l'utilisateur.

Assistive Context-Aware Toolkit ou ACAT est l'interface utilisateur développée par Intel

³⁹ Presage est disponible en ligne sur SourceForge (<http://presage.sourceforge.net/>) sous la *GNU General Public License version 2.0* et ACAT (<https://01.org/acat>)

sur GitHub (<https://github.com/01org/acat>) sous l'*Apache License version 2.0*.

Corporation et initialement conçue pour le physicien Stephen Hawking atteint d'une sclérose latérale amyotrophique (Denman et al. 2016). Presage est inclus par défaut dans la version adaptée au public. ACAT est un système de CAA en anglais qui comprend deux modes (*Talk* ou *App*) ainsi que plusieurs claviers virtuels accessibles via différentes modalités de saisie. Le système de défilement permet à l'utilisateur qui ne peut pas se servir d'une souris de sélectionner les caractères et les mots à l'aide d'un détecteur de mouvements des sourcils, des joues ou de la bouche via une webcam. En raison des caractéristiques du handicap de notre utilisateur de test, nous ne l'avons pas utilisé. Le logiciel étant libre, des packs de langue sont disponibles sur GitHub.

3 Méthodologie

L'objectif était d'évaluer diverses configurations de Presage afin d'analyser leur efficacité et leurs performances en termes d'amélioration de la communication. Nous avons étudié l'impact de plusieurs paramètres : le corpus, le modèle de langage n-gramme, le nombre de prédictions lexicales à afficher, mais aussi l'interaction et la satisfaction de l'utilisateur concernant l'interface. Pour ce faire, nous avons procédé à des évaluations quantitatives et qualitatives.

3.1 Configuration

Par défaut, Presage est basé sur un modèle de langage trigramme lissé, entraîné à partir d'un livre libre de droit. Nous avons donc d'abord constitué des corpus d'entraînement et de test. Nous avons employé la version française de Google Books Ngram⁴⁰ (Michel et al. 2011) de 2009 que nous avons tokenisée et traitée (46 786 461 uni/bi/trigrammes). Nous avons testé deux autres corpus d'entraînement : l'un était constitué du livre *À se tordre d'A. Allais* (38 123 mots) et l'autre de textes Perso représentant l'idiolecte de potentiels utilisateurs (18 337 mots), c'est-à-dire composés d'une interview et de plusieurs articles de blog écrits par deux personnes handicapées. Le corpus de test consistait en une centaine de phrases simples ne faisant pas partie des corpus d'entraînement, mais issues de diverses sources et destinées à refléter différents usages.

⁴⁰ Le corpus sous forme de n-grammes est disponible (<http://storage.googleapis.com/books/ngrams/books/datas>

3.2 Simulations

Afin d'évaluer le système de prédiction de mots, trois métriques ont été utilisées : le *Keystroke Saving Rate*, le taux d'économie de saisie qui ne prend pas en compte l'interface et le *hit ratio*.

$$KSR = 1 - \frac{ki + ks}{kn} * 100$$

où ki est le nombre de saisies réalisées, ks le nombre de saisies nécessaires pour sélectionner la suggestion (dans notre étude $ks = 1$ par mot) et kn le nombre de saisies qui auraient été nécessaires pour composer le texte sans utiliser le système prédictif. Le taux d'économie est similaire au KSR, mais ks n'est pas calculé.

$$Hit\ ratio = \frac{\text{nombre de mots prédits}}{\text{nombre total de mots}} * 100$$

Le *hit ratio* désigne le taux d'utilisation de la prédiction.

3.3 Test utilisateur

Afin d'évaluer Presage intégré à ACAT, nous avons effectué un test utilisateur avec une personne handicapée ayant une Infirmité Motrice Cérébrale (IMC) athétosique. D'après un premier questionnaire, elle utilise régulièrement des systèmes de CAA intégrant la prédiction de mots. Le participant était un adulte qui ne parle pas, ne présente pas de déficience intellectuelle et a un bon niveau de français. Nous lui avons demandé de recopier une vingtaine de phrases issues du corpus de test et d'utiliser la prédiction dès que le mot correct était affiché dans la liste.



Figure 2. User experience.

Il avait aussi la possibilité d'utiliser certaines abréviations grâce à un mécanisme de désabréviation automatique. Pour sélectionner les caractères et les prédictions, il a utilisé la souris avec ses pieds (Figure 2) car la crispation de son visage ne permettait pas d'utiliser la reconnaissance de mouvements. Après l'expérience que nous avons enregistré avec

(etsv2.html) sous la *Creative Commons Attribution 3.0 Unported License*.

l'outil *Morae*, nous lui avons envoyé un questionnaire et l'échelle du *System Usability Scale* ou *SUS* (Brooke 1996) a été utilisée. Les paramètres de *Presage* ont été déterminés en fonction des résultats théoriques obtenus avec le simulateur.

4 Résultats

Les simulations ont montré que le modèle de langage trigramme lissé avec interpolation linéaire entraîné avec le corpus *Google Books Ngram* obtenait les meilleures performances et permettait d'économiser plus d'une saisie sur deux en proposant 5 prédictions (Table 1).

	<i>Google Books</i>	<i>À se tordre</i>	<i>Perso</i>
KSR	57.932%	39.617%	39.020 %
T. d'économie	75.052%	54.705%	53.181 %
Hit ratio	90.094%	79.402%	74.528 %

Table 1. Résultat par corpus.

Nos résultats ont aussi montré que le modèle trigramme entraîné avec *Google Books Ngram* était meilleur que le modèle unigramme ou bigramme (Table 2). En l'associant avec un modèle trigramme utilisateur, nous avons également observé une amélioration de 1 à 2 %.

	<i>Unigramme</i>	<i>Bigramme</i>	<i>Trigramme</i>
KSR	43.800%	54.108%	57.932%
T. d'économie	60.382%	71.287%	75.052%
Hit ratio	87.264%	90.408%	90.094%

Table 2. Résultats par nombre de *n*-grammes.

Toujours entraîné avec *Google Books Ngram*, nous montrons que plus le nombre de prédictions est grand, plus les résultats sont élevés (Table 3).

	1	5	9
KSR	45.354%	57.932%	61.458%
T. d'économie	60.501%	75.052%	78.996%
Hit ratio	79.716%	90.094%	92.295%

Table 3. Résultats par nombre de prédictions.

Ces différents tests ont permis d'identifier les paramètres à implémenter dans *ACAT* pour le test utilisateur, c'est-à-dire le modèle de langage trigramme lissé, entraîné à partir du

corpus *Google Books Ngram*, associé à un modèle utilisateur et proposant une liste de 9 prédictions.

Les vidéos du test utilisateur⁴¹ ont montré que la personne évaluée ne sélectionnait pas toujours la prédiction correcte dès qu'elle apparaissait dans la liste. Le participant qui avait l'occasion d'utiliser les abréviations préalablement encodées dans *ACAT* n'en a employées que quatre sur les dix occasions. Le score obtenu avec le *SUS* était de 70 % donc presque bon (73 %).

Durant l'*user experience*, nous avons observé que plusieurs propositions étaient mal contextualisées, que des mots étrangers, inexistantes ou des sigles (qui ne seront probablement jamais utilisés) étaient proposés. L'utilisateur a reporté avoir été confronté à quelques bugs, mais a reconnu avoir été satisfait des systèmes et des prédictions.

5 Conclusion

Cet article a montré qu'entraîné avec un modèle trigramme et le corpus *Google Books Ngram*, *Presage* permettait d'économiser plus d'une saisie sur deux. Nos résultats liés au nombre de *n*-grammes et de prédictions ont confirmé ceux trouvés dans la littérature (Wandmacher 2008). Le test d'*ACAT* par une personne handicapée était centré sur l'étude de l'interaction et de la satisfaction d'un potentiel utilisateur.

Nous pouvons conclure que ces deux systèmes offrent des possibilités intéressantes. Notre étude s'est limitée à un test utilisateur. Néanmoins, elle a permis de relever une charge cognitive élevée, certains problèmes ergonomiques, mais aussi de compléter l'évaluation du système de prédiction. Il serait intéressant d'évaluer d'autres utilisateurs handicapés et de contribuer à l'amélioration ainsi qu'à l'accessibilité de ces systèmes *open source*.

Références

Antoine Jean-Yves, and Maurel Denis. 2007. Aide à la communication pour personnes handicapées et prédiction de texte : Problématique, état des lieux et retour sur trente ans de recherche en

⁴¹ Tous les fichiers relatifs à cette étude sont disponibles en ligne (https://github.com/Galami/WordPrediction_AAC).

- communication augmentée. *TAL Traitement Automatique des Langues*, 48(2):9–46.
- Brooke John. 1996. SUS: A quick and dirty usability scale. *Usability evaluation in industry*, 189(194).
- Denman Pete, Nachman Lama, and Prasad Sai. 2016. Designing for "a" user: Stephen Hawking's UI. *Proceedings of the 14th Participatory Design Conference: Short Papers, Interactive Exhibitions, Workshops*, pages 94–95. ACM.
- Garay-Vitoria Nestor, and Abascal Julio. 2006. Text prediction systems: A survey. *Universal Access in the Information Society*, 4(3):188–203.
- Michel Jean-Baptiste, Shen Yuan Kui, Aiden Aviva Presser, Veres Adrian, Gray Matthew K., Pickett Joseph P., Hoiberg Dale, Clancy Dan, Norvig Peter, Orwant Jon, et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.
- Vescovi Matteo. 2004. *Soothsayer: Un sistema multi-sorgente per la predizione del testo*. Thèse, Politecnico di Milano.
- Wandmacher Tonio. 2008. *Adaptive word prediction and its application in an assistive communication system*. Thèse, Université François – Rabelais.