Évaluation de la prédiction de mots dans un système de communication améliorée et alternative pour les personnes en situation de handicap

Presage et ACAT

Magali Norré
Institut Langage et Communication
Université catholique de Louvain
Louvain-la-Neuve, Belgique
magali.norre@uclouvain.be

***Résumé*—Cet article présente une évaluation d'un système de prédiction de mots dans un logiciel de Communication Améliorée et Alternative (CAA) pour les personnes en situation de handicap. En plus d'avoir une mobilité réduite, ces utilisateurs ont un usage de la parole plus ou moins altéré qui doit être compensé par une aide technologique proposant des modalités de saisie adaptées à leurs capacités. Afin d'accélérer leur vitesse de communication, différentes techniques de prédiction et de modélisation du langage sont utilisées. Ces approches sont étudiées par le biais de divers systèmes existants. Le paramétrage de prédicteurs est analysé, leur configuration en français étant évaluée par un simulateur et testée par une personne handicapée. Les résultats montrent qu'un modèle de langage construit à partir d'un gros corpus permet d'économiser plus d'une saisie sur deux, les performances de ces systèmes variant en fonction de plusieurs paramètres.**

Mots-clés—prédiction de mots, communication améliorée et alternative, caa, technologie d'assistance, traitement automatique du langage, interaction homme-machine, évaluation, handicap

I. 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 d'importantes 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 dont la communication est plus ou moins altérée ont désormais la possibilité de se procurer ces systèmes de Communication Améliorée et Alternative (CAA). Ces méthodes sont dites *améliorées* lorsqu'elles permettent de suppléer les capacités déjà présentes, pouvant même aider à l'émergence de l'oralisation ou de la graphie pour certaines personnes ; et *alternatives* lorsqu'elles remplacent totalement les moyens d'expression orale [1]. Les systèmes que nous évaluerons sont davantage adaptés aux personnes paralysées ayant un trouble sévère de la parole, mais ne présentant pas une déficience intellectuelle, comme le locked-in syndrome.

Disponibles sur ordinateur ou tablette, la plupart intègrent des technologies comme la prédiction de textes, les claviers virtuels et la synthèse vocale destinée à oraliser le message. Cet article vise à présenter le fonctionnement des moteurs de prédiction, ces systèmes utilisés dans les téléphones portables qui suggèrent des mots en fonction des lettres déjà entrées. Cette technique bien connue du grand public permet à l'utilisateur d'économiser les saisies requises pour composer un mot et accélérer la vitesse de saisie. Étant donné que certains utilisateurs en situation de handicap ne peuvent se servir d'un clavier ou d'une souris, le recours à ces systèmes est alors nécessaire et se traduit par leur omniprésence dans de nombreuses communications alternatives modernes. L'état de l'art (section II) permettra de passer en revue leurs caractéristiques. Nous détaillerons ensuite les deux systèmes constituant notre étude de cas : le moteur de prédiction Presage [2] intégré au logiciel de CAA ACAT [3] (section III).

II. État de l'art

Tout d'abord, nous introduisons quelques systèmes de prédiction de mots et de CAA existants (A), ainsi que certaines limites concernant leur évaluation (B).

A. Systèmes de prédiction de mots et de CAA

Avant d'être utilisés par le grand public dans les interfaces limitées comme les dispositifs mobiles, les systèmes d'optimisation de la saisie comme la prédiction de mots ont été initialement conçus pour les personnes en situation de handicap [4]. Il faut savoir que si l'on compare le débit de la communication orale estimé généralement à 150-200 mots par minute [5] [6] [7], les outils de CAA ne dépassent pas 10 à 15 mots par minute. Afin d'accélérer la vitesse de saisie de textes et diminuer la fatigue des utilisateurs, il existe diverses techniques et de nombreuses recherches sur la prédiction de textes [8] [9].

Les systèmes de prédiction se basent sur certaines informations contextuelles : statistiques ou linguistiques. Afin de modéliser le langage et l'adapter à l'utilisateur, il existe différents types d'approches qui peuvent néanmoins être

complémentaires et concerner à la fois la prédiction de lettres ainsi que la prédiction de mots. Les informations statistiques et linguistiques sont stockées dans des tables appelées dictionnaires ou lexiques [9]. Leur contenu peut soit être fixé préalablement par le concepteur et donc identique pour tous les utilisateurs, soit adapté automatiquement en fonction de ceux-ci.

Les systèmes statistiques modélisent le langage à l'aide de fréquences et de modèles markoviens basés sur les n-grammes – bigramme : séquence de deux éléments (mots, lettres ou caractères), trigramme : trois éléments, etc. – qui sont parfois associés à des techniques stochastiques plus avancées comme le lissage ou certaines interpolations. Il est possible d'adapter ce type de systèmes d'une langue à l'autre, mais également à un vocabulaire ou à un registre de langage plus spécifique. De nombreux claviers virtuels gratuits ou payants existent. Beaucoup de ces prédicteurs utilisent une approche statistique. C'est le cas de PolyPredix commercialisé par AssistiveWare [10] [11]. La prédiction statistique est utilisée dans le système Sibylle pour prédire les mots [12] [13] et les lettres [7].

Les outils de prédiction permettent de compléter un mot en ne proposant qu'une suggestion commençant par les caractères saisis, le choix s'affinant naturellement à chaque saisie. Lorsque l'utilisateur décide de retenir le mot proposé en le validant, celui-ci est directement intégré au texte permettant ainsi « d'économiser les dernières saisies qui auraient été nécessaires à la composition du mot » [14]. Les systèmes de prédiction affichent une liste de mots présentée au-dessus du clavier ou à côté. Le système de complétion VITIPI [15] affiche les fins de mots lorsqu'il n'y a plus d'ambiguïté possible. Les auteurs font remarquer que vu qu'il n'existe pas de liste de prédictions, la question de la surcharge cognitive ne se poserait plus [16] [17].

Les logiciels de prédiction et de complétion peuvent être intégrés à un clavier virtuel. L'utilisateur accepte les mots prédits en cliquant dessus à l'aide d'un dispositif de pointage (ex. : un contacteur) combiné à un système de défilement automatique du curseur (souvent par bloc, ligne et colonne). La sélection de mots se fait via un seul clic ou avec des raccourcis clavier. Le prototype de complétion néerlandais Soothsayer [18] fonctionne par exemple avec la touche espace, tandis que l'éditeur de texte intégré au moteur de prédiction Presage [2] nécessite la souris ou les touches F1 à F6 pour sélectionner les propositions. L'interface utilisateur ACAT [3] peut afficher la liste de mots à la fois horizontalement et verticalement.

B. Limites linguistiques, cognitives et ergonomiques

L'aide à la communication est un domaine pluridisciplinaire lié à la linguistique(-informatique), la psychologie cognitive, ainsi que l'ergonomie [19]. En ce qui concerne l'évaluation des systèmes de prédiction de mots, il est important de tenir compte de certaines limites concernant plusieurs aspects comme la modélisation du langage en Traitement Automatique des Langues (TAL), l'usage réel de la prédiction par les personnes en situation de handicap relatif à l'Interaction Homme-Machine (IHM), mais aussi à l'apport de différentes métriques. Il existe deux types d'évaluation : une méthode subjective qui prend en compte l'avis de potentiels utilisateurs et une méthode objective centrée sur des métriques qui peut concerner soit le système end-to-end avec l'interface, soit seulement la prédiction.

Bien que l'on constate des écarts entre une évaluation objective et une évaluation subjective avec des participants, ce « décalage avec l'estimation des performances de la prédiction ne doit pas remettre pas en cause l'utilité des métriques utilisées » [8]. Concernant les mesures subjectives, il n'existe pas de méthode pour évaluer la satisfaction des utilisateurs dans le cas spécifique d'un système de prédiction de mots. L'Évaluation de la Satisfaction envers une Aide Technique (ESAT) est un questionnaire général qui comporte 12 énoncés de satisfaction sur les ajustements, le confort, l'efficacité, etc. pour tous les systèmes de CAA. Cette échelle est très rarement citée dans les études sur la prédiction de textes.

Pour évaluer la vitesse de saisie, les chercheurs calculent parfois le nombre moyen d'appuis par caractère saisi ou le nombre de mots par minute. Les résultats des différentes métriques dépendent souvent de l'utilisateur et de l'ergonomie puisqu'une mauvaise interface peut occasionner une charge cognitive supplémentaire qui pourrait, par exemple, ralentir la production de texte [20]. Il existe en réalité peu de recherches sur l'utilisation des systèmes de prédiction sur le long terme. L'IHM a un impact sur l'utilisation des aides à la saisie de texte. Si le système end-to-end comporte des sous-claviers pour accéder à certains caractères, il faudra compter plus de saisies si on évalue par exemple le KSR.

Les métriques classiques en TAL (par exemple, la perplexité ou l'entropie d'un modèle) sont rarement utilisées dans le domaine de la CAA. Issues de la théorie de l'information, elles « ne constituent pas des indicateurs explicites de l'aide fournie par la prédiction » [8]. Afin de l'évaluer et adapter ce système, les chercheurs et concepteurs implémentent souvent un module qui permet de simuler automatiquement la meilleure utilisation, c'est-à-dire ne calculant pas les erreurs que l'utilisateur produirait et sélectionnant directement la bonne prédiction dès qu'elle est susceptible d'apparaître dans la liste. Lors de tests en situation réelle, il ne faut pas oublier que les systèmes de prédiction imposent une charge cognitive aux utilisateurs [21].

C'est pourquoi, nous avons choisi d'effectuer une évaluation objective, ainsi qu'une évaluation subjective.

III. ÉTUDE DE CAS

L'objectif de cette étude était d'évaluer objectivement le moteur de prédiction de mots Presage et son paramétrage, avant de réaliser une évaluation subjective en situation réelle de cet outil intégré dans le logiciel de CAA complet ACAT. Nous présentons donc les deux systèmes que nous avons configurés et évalués (A), avant d'aborder notre méthodologie (B), ainsi que nos résultats (C) qui feront l'objet de la discussion et nous évoquerons certaines questions éthiques (D).

A. Presage intégré à ACAT

Les deux systèmes configurés et évalués ont été conçus de manière indépendante et sont tous les deux open source (Fig. 1). Nous avons d'abord testé la version 0.9.1 (2015) de Presage et ensuite la version 1.0.0 (2016) d'ACAT, qui intègre Presage.

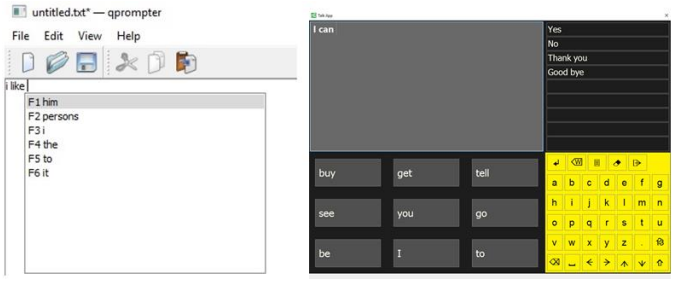


Fig. 1. Interface de Presage et ACAT.

Presage (anciennement Soothsayer) est un système de prédiction de mots développé par un chercheur dans le cadre de sa thèse [2]. Il est constitué de plusieurs prédicteurs statistiques et modèles de langage n-grammes fonctionnant avec des corpus. 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 Corporation et initialement conçue pour une seule personne : le physicien Stephen Hawking atteint d'une sclérose latérale amyotrophique [3]. 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 avec un détecteur de mouvements des sourcils, des joues ou de la bouche via une webcam. En raison du handicap de notre utilisateur de test, nous n'avons pas utilisé cette modalité de saisie. Le logiciel étant libre, des packs de langue sont disponibles sur GitHub.

B. Méthodologie

Nous avons étudié l'impact de plusieurs paramètres : le corpus, le modèle de langage n-gramme, le nombre de prédictions lexicales à afficher (évaluation objective), mais aussi l'interaction et la satisfaction de l'utilisateur concernant le système complet (évaluation subjective).

Par défaut, Presage est basé sur un modèle de langage trigramme lissé, entrainé à partir d'un livre libre de droit. Nous avons donc d'abord constitué des corpus d'entrainement et un corpus de test. Nous avons employé la version française de *Google Books Ngram* [22] de 2009 que nous avons tokenisée et traitée (46 786 461 uni/bi/trigrammes). Nous avons testé deux autres corpus d'entrainement : l'un était constitué du livre *À se tordre* d'Alphonse Allais (38 123 mots) et l'autre de textes *Perso* représentant l'idiolecte de potentiels utilisateurs (18 337 mots), c'est-à-dire constitué d'articles de *Blog* (14 071 mots) et d'une *Interview* (4 306 mots), écrits par deux personnes handicapées.

Le corpus de test relevait d'un registre de langue familier. Il consistait en une centaine de phrases simples ne faisant pas partie des corpus d'entrainement, mais issues de diverses sources et destinées à refléter différents usages quotidiens. Nous avons pris une quarantaine de phrases générales issues d'un corpus destiné à évaluer des techniques de saisie de textes [23], que nous avons traduites en français. Nous avons ajouté des phrases utilisées dans le cadre de l'évaluation d'autres systèmes de CAA.

1) Évaluation objective

Afin d'évaluer le système de prédiction Presage, nous avons réalisé un simulateur automatique en adaptant et en étendant celui de l'auteur [2]. Nous avons ainsi calculé trois métriques : le Keystroke Saving Rate ou KSR (1), le taux d'économie de saisie qui ne prend pas en compte l'interface et le hit ratio (2).

$$KSR = \left(1 - \frac{ki + ks}{kn}\right) * 100 \qquad (1)$$

En (1), ki est le nombre d'appuis réalisés, ks le nombre d'appuis nécessaires pour sélectionner la suggestion dans la liste de Presage (ks = 1 par mot), kn le nombre d'appuis qui auraient été nécessaires pour composer le texte sans utiliser le système. Le taux d'économie est similaire au KSR, mais ks n'est pas calculé. Le hit ratio (2) est le taux d'utilisation de la prédiction.

$$Hit\ ratio = \frac{nombre\ de\ mots\ prédits}{nombre\ total\ de\ mots} * 100 \qquad (2)$$

2) Évaluation subjective

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 notre questionnaire, elle utilisait 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.

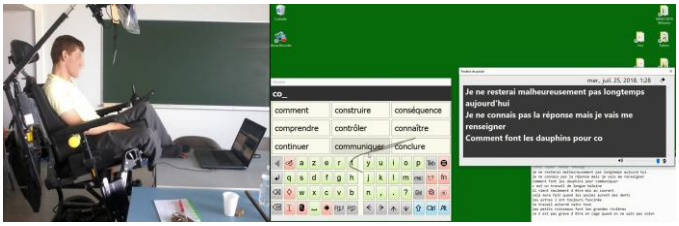


Fig. 2. Presage dans ACAT : test utilisateur.

Il avait aussi la possibilité d'utiliser certaines abréviations grâce à un mécanisme de désabréviation automatique. L'une des consignes était d'y recourir en fonction des possibilités du texte. Pour sélectionner les caractères et les prédictions, il a utilisé la souris avec ses pieds (Fig. 2) car la crispation de son visage ne permettait pas d'utiliser le système de reconnaissance de mouvements. Après l'expérience que nous avons enregistrée avec l'outil Morae, il a répondu à un questionnaire et nous avons utilisé la System Usability Scale ou SUS [24].

Les paramètres de Presage ont été déterminé en fonction des résultats théoriques obtenus avec notre simulateur. L'option d'apprentissage automatique n'a pas été testée et est restée désactivée afin de conserver les mêmes probabilités tout au long de l'expérience. Le test utilisateur a duré une heure : une séance d'installation et d'explications de 10 minutes suivie de deux sessions (avec une pause), chacune destinée à retranscrire 10 phrases. La première a duré 28 minutes et la deuxième 20.

C. Résultats

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

TABLE I. RÉSULTATS PAR CORPUS

	Google Books Ngram	**À se tordre**	**Perso**
KSR	57,9 %	39,6 %	39 %
Taux d'économie	75 %	54,7 %	53,1 %
Hit ratio	90 %	79,4 %	74,5 %

Nos résultats ont également montré que le modèle trigramme entrainé avec *Google Books Ngram* était meilleur qu'un modèle unigramme et qu'un modèle bigramme (Table II).

TABLE II. RÉSULTATS PAR NOMBRE DE N-GRAMMES

	Unigramme	**Bigramme**	**Trigramme**
KSR	43,8 %	54,1 %	57,9 %
Taux d'économie	60,3 %	71,2 %	75 %
Hit ratio	87,2 %	90,4 %	90 %

Au modèle de langage général entrainé avec le corpus *Google Books Ngram*, nous avons ajouté un modèle de langage utilisateur (Table III). Nous supposons donc que l'utilisateur ait saisi manuellement les textes du corpus *Perso*, du *Blog* ou de l'*Interview* lorsque le mode d'apprentissage était activé. Nous notons une sensible amélioration par rapport à la Table I.

TABLE III. RÉSULTATS AVEC UN MODÈLE DE LANGAGE UTILISATEUR

	Perso	**Blog**	**Interview**
KSR	59 %	58,7 %	58,5 %
Taux d'économie	76,4 %	76,1 %	75,9 %
Hit ratio	91,8 %	91,5 %	91,6 %

Toujours entrainé avec *Google Books Ngram*, nous montrons que plus le nombre de prédictions proposées est grand, plus les résultats sont élevés (Table IV). Il faut néanmoins noter que dans un système de CAA, il est nécessaire de tenir compte des aspects ergonomiques et cognitifs des utilisateurs.

TABLE IV. RÉSULTATS PAR NOMBRE DE PRÉDICTIONS

	1	**3**	**5**	**7**	**9**
KSR	45,3 %	51,2 %	57,9 %	60,1 %	61,4 %
Taux d'économie	60,5 %	67,5 %	75 %	77,5 %	78,9 %
Hit ratio	79,7 %	85,6 %	90 %	91,6 %	92,2 %

Ces différents tests effectués sur le même corpus de test ont permis d'identifier les paramètres à implémenter dans ACAT pour le test utilisateur, c'est-à-dire le modèle de langage trigramme lissé, entrainé à partir du corpus *Google Books Ngram*, associé à un modèle de langage utilisateur (le corpus *Perso*) et proposant une liste de 9 prédictions.

Les enregistrements du test utilisateur 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 (seuil à 73 %) [24]. Durant le test, nous avons observé que plusieurs propositions étaient mal contextualisées, que des mots étrangers, inexistants ou des sigles (qui ne seront probablement jamais utilisés) étaient proposés. Le participant a aussi noté avoir été confronté à quelques bugs, mais a reconnu avoir été satisfait des systèmes et des prédictions.

D. Discussion

Tout d'abord, il faut souligner le fait qu'il est très difficile de comparer les études entre elles, notamment en raison de méthodologies hétérogènes. Les conditions de test ne sont pas toujours explicites et certaines ne font pas de distinction entre deux métriques souvent confondues : le KSR et le taux d'économie de saisie. Ils varient selon que l'on évalue l'interface complète ou seulement la prédiction. Les résultats du KSR sont effectivement toujours moins bons que ceux obtenus avec le taux d'économie, l'écart variant de 15 à 20 %. Bien qu'il ne donne pas d'information sur le nombre d'appuis effectués ou économisés, le hit ratio ou taux d'utilisation de la prédiction est généralement resté élevé, c'est-à-dire dépassant les 70 %.

Nos premiers tests ont montré que le corpus *Google Books Ngram* permettait d'économiser plus d'une saisie sur deux que ce soit via le KSR ou avec le taux d'économie. Nos résultats corroborent ceux de plusieurs auteurs qui obtiennent des KSR aux alentours des 50 % en fonction de divers corpus et paramètres [4] [7] [12] [13] [19] [20]. Nous avons constaté un très grand écart – de plus ou moins 20 % – dans les résultats obtenus avec *Google Books Ngram* et les deux autres corpus d'entrainement. Les données d'apprentissage ont donc un impact important sur les résultats. Si nous comparons la taille du corpus d'entrainement, nous avons vu qu'un corpus de plusieurs millions de n-grammes était plus performant qu'un corpus constitué d'un seul livre comme *À se tordre*. Nous avons aussi montré qu'un modèle de langage trigramme qui, pour chaque mot, prend en compte les deux précédents est plus performant qu'un modèle d'ordre inférieur comme mentionné dans différentes études [7] [25]. Quant à la combinaison d'un modèle de langage général avec un modèle de langage utilisateur, nous avons vu que plus le deuxième corpus d'entrainement est gros, meilleurs seront les résultats. L'efficacité est toutefois limitée. Il faudrait en effet que le système soit utilisé régulièrement.

En ce qui concerne les prédictions, nous supposions que plus leur nombre augmente, plus il y a de chances que le mot prédit soit le bon et donc celui recherché par l'utilisateur. Les résultats théoriques du hit ratio confirment bien cette hypothèse. Nous constatons une augmentation de plus de 10 % du taux d'utilisation selon que le système propose une ou (plus de) cinq prédictions. L'amélioration étant moins importante à partir de cinq, nous pensons que le choix doit se faire en fonction des préférences de l'utilisateur. Une étude [6] a par exemple testé jusqu'à vingt suggestions alors que d'autres [26] [27] avaient constaté une amélioration des résultats, c'est-à-dire une diminution du pourcentage du nombre d'appuis, en augmentant le nombre de mots prédits de un à dix.

Le temps et l'effort cognitif nécessaires pour sélectionner une proposition augmentent si le nombre de prédiction est élevé [9].

Un ergothérapeute insiste sur le fait que plus le nombre de prédictions est important, plus le KSR augmente et par conséquent, plus le temps de recherche d'un mot dans la liste augmente, ce qui diminuerait la vitesse de communication [14]. Comme il a déjà été mentionné dans la littérature [28], il faut donc toujours trouver le juste équilibre entre la maximisation de la performance et la minimisation de la charge cognitive ressentie par l'utilisateur, c'est-à-dire faire le choix entre un système plus performant ou moins fatiguant.

L'évaluation que nous avons conduite suggère les bénéfices que l'on peut attendre de l'utilisation de la CAA. Avant de conclure, il nous semble important de discuter de ces systèmes d'aide d'un point vue éthique afin de mettre en avant certains risques et leurs implications. Des ateliers comme ETeRNAL (Ethique et Traitement Automatique des Langues) ainsi que plusieurs études [29] [30] [31] commencent à aborder la question des risques éthiques liés aux technologies numériques. Une démarche conséquentialiste consiste alors à caractériser les facteurs de risque liés à l'utilisation des outils numériques.

Les concepteurs du système de prédiction Sibylle ont montré que celui-ci pouvait conduire à la production de plus de textes et une diminution de fautes d'orthographe chez des enfants avec une IMC [12]. Les auteurs s'interrogent toutefois sur la modification des compétences langagières due à une certaine dépendance, en se demandant « si l'aide favorise une maitrise plus rapide du système de la langue, ou si l'amélioration de la qualité des productions ne masque pas un abandon de cette capacité au profit du système… » [29].

Autrement dit, un risque attenant à l'usage de ces technologies comme outil d'assistance ou de compensation serait la perte à terme d'une compétence supplée par la technologie. Les utilisateurs pourraient en effet avoir tendance à se fier davantage aux propositions de la machine plutôt qu'à leurs propres analyses. Ainsi, bien que ces systèmes soient conçus pour réduire leur handicap en leur permettant d'accélérer leur communication, ils peuvent aussi l'augmenter en les aidant à tout moment dans leurs choix orthographiques comme le font les correcteurs automatiques. Ainsi, ils se substituent aux réelles connaissances des utilisateurs.

Si nous poursuivons cette réflexion, le choix du corpus d'apprentissage et donc des mots que le système va proposer peut également être considéré comme un facteur de risque. On limite en effet à ce seul lexique appris par le système les possibilités d'interaction de l'utilisateur, ceci le plus souvent avec un but justifiable : saisir plus vite et ne pas le troubler avec des prédictions qu'il ne comprendrait pas. Les chercheurs sont confrontés à un dilemme. Si la prédiction dépasse les compétences langagières de l'utilisateur, elle peut le perturber, ce qui pourrait empêcher l'apprentissage chez des enfants. Mais si elle est trop simple, elle réduit ses possibilités d'expression en deçà de ses capacités et augmente son handicap. C'est pourquoi, le rôle des logopèdes et des ergothérapeutes est d'ajuster au mieux le lexique du système aux capacités du patient.

D'autres facteurs de risque peuvent également être envisagés. Des prédictions inappropriées pourraient avoir « une influence négative sur l'état psychologique du patient (énervement, fatigue cognitive) » [29] [31]. En testant *Google Books Ngram* avec un réel utilisateur, nous avons noté plusieurs propositions lexicales « mal contextualisées ». Pour un usage en situation écologique, nous pensons que nos traitements ne sont pas suffisants vu le nombre de mots qui ne seront probablement jamais utilisés.

Il existerait donc plusieurs risques relatifs à la prédiction. Nous avons étudié un risque cognitif (régression) se traduisant par une perte d'autonomie ainsi qu'un risque psychologique lié au ressenti par rapport à l'usage de la prédiction. Ce type de réflexion nous semble important à relever, car les concepteurs de systèmes de prédiction ont tendance à penser en termes de gains immédiats, observables et mesurables objectivement (comme le taux d'économie de saisie ou la vitesse de saisie) en ne parlant pas des risques sur le long terme ou des éventuels « effets plus globaux sur la rééducation et l'évolution de la pathologie » [29].

IV. Conclusion

Cet article a montré qu'entrainé avec la version française du gros corpus *Google Books Ngram*, Presage permettait d'économiser plus d'une saisie sur deux. Nos résultats ont confirmé les précédentes études sur la prédiction, à savoir qu'un modèle de langage trigramme est plus efficace qu'un modèle n-gramme d'ordre inférieur et que plus le nombre de prédictions est grand, meilleurs seront le KSR, le taux d'économie de saisie et le hit ratio. Il est aussi possible d'obtenir des améliorations avec un modèle de langage utilisateur, mais cela prend du temps car le système de prédiction de mots doit s'adapter et donc être fréquemment utilisé. Le test de Presage dans ACAT en situation réelle par une personne handicapée était centré sur l'étude de l'interaction et de la satisfaction d'un potentiel utilisateur. Néanmoins, la grande variabilité au sein d'un même handicap (ici IMC) ne permet pas d'obtenir des conclusions facilement généralisables à l'ensemble des personnes qui seraient dans une situation de handicap plus ou moins similaire.

Notre test utilisateur a montré que certaines prédictions avaient peu de chance d'être utilisées et qu'elles n'étaient pas systématiquement choisies dès qu'elles apparaissaient dans la liste. Nous avons aussi vu que l'utilisation des prédictions dépend de la charge cognitive et visuelle du patient. En plus de l'effort physique, l'utilisateur doit se concentrer sur l'écriture de son texte, éviter de faire des erreurs, lire les prédictions et faire appel à ses capacités de mémorisation lorsqu'il pense pouvoir utiliser une abréviation. Ainsi, malgré l'implémentation d'un mécanisme permettant d'économiser davantage de saisies (et donc de pallier la lenteur de la communication), la personne testée a préféré ne pas les utiliser. D'un point de vue ergonomique, l'interface peut également avoir un impact sur les résultats de certaines métriques utilisées dans le cadre de la prédiction. En testant ACAT, nous avons par exemple constaté que le simple ajout d'un sous-clavier contenant les lettres accentuées est susceptible d'influencer le KSR.

Bien que nous n'ayons pas pu décrire tous nos résultats [32], nous pouvons conclure que l'intégration de Presage dans ACAT offre des possibilités intéressantes. Notre étude s'est limitée à un test utilisateur avec un seule personne en situation de handicap.

Elle a toutefois permis de relever une charge cognitive élevée, certains problèmes ergonomiques, mais aussi de compléter l'évaluation du système de prédiction. L'adaptation utilisateur reste un point central quand on évalue ce type de technologies. Il serait intéressant d'évaluer d'autres personnes en situation de handicap et de contribuer à l'amélioration ainsi qu'à l'accessibilité de ces systèmes open source.

Remerciements

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. Le corpus sous forme de n-grammes est disponible (http://storage.googleapis.com/books/ngrams/books/datasetsv2.html) sous la Creative Commons Attribution 3.0 Unported License. Enfin, certains fichiers relatifs à cette étude sont disponibles (https://github.com/Galami/WordPrediction_AAC).

Références

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

[2] Vescovi, M. *Soothsayer: Un sistema multi-sorgente per la predizione del testo*. Thèse, Politecnico di Milano, 2004.

[3] Denman, P., Nachman, L. et Prasad, S. « Designing for "a" user: Stephen Hawking's UI », dans *Proceedings of the 14th Participatory Design Conference: Short Papers, Interactive Exhibitions, Workshops*, ACM, 2016, p. 94-95.

[4] Antoine, J.-Y. « Prédiction de mots et saisie de requêtes sur interfaces limitées : Dispositifs mobiles et aide au handicap », dans *Recherche d'information contextuelle, assistée et personnalisée*, Recherche d'information et web, Hermès-Lavoisier, 2011, p. 273-298.

[5] Wood, M. E. J. *Syntactic Pre-Processing in Single-Word Prediction for Disabled People*. Thèse, University of Bristol, 1996.

[6] Copestake, A. « Augmented and alternative NLP techniques for augmentative and alternative communication », dans *Natural Language Processing for Communication Aids*, 1997, p. 37-42.

[7] Schadle, I. *Sibylle : Système linguistique d'aide à la communication pour les personnes handicapées*. Thèse, Université de Bretagne Sud, 2003.

[8] Antoine, J.-Y. et Maurel, D. « Aide à la communication pour personnes handicapées et prédiction de texte : Problématique, état des lieux et retour sur trente ans de recherche en communication augmentée », dans *TAL Traitement Automatique des Langues*, 48(2), 2007, p. 9-46.

[9] Garay-Vitoria, N. et Abascal, J. « Text prediction systems: A survey », dans *Universal Access in the Information Society*, 4(3), 2006, p. 188-203.

[10] Bérard, C. et Niemeijer, D. « Evaluating Effort Reduction through Different Word Prediction Systems », dans *International Conference on Systems, Man and Cybernetics*, 3, IEE, 2004, p. 2658-2663.

[11] Niemeijer, D. « In memoriam of Christian Bérard : Striving for effort reduction through on-screen keyboard word prediction », dans *Assistive technology : From virtuality to reality*, Lille, France, 2005.

[12] Wandmacher, T. et Antoine, J.-Y. « Modèle adaptatif pour la prédiction de mots : Adaptation à l'utilisateur et au contexte dans le cadre de la communication assistée pour personnes handicapées », dans *TAL Traitement Automatique des Langues*, 48(2), 2007, p. 71-95.

[13] Wandmacher, T. *Adaptive word prediction and its application in an assistive communication system*. Thèse, Université François Rabelais de Tours, 2008.

[14] Pouplin, S. *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*. Thèse, Université Paris-Saclay, 2016.

[15] Boissière, P. et Dours, D. « Comment VITIPI un système d'assistance à l'écriture pour les personnes handicapées peut offrir des propriétés intéressantes pour le TALN ? », dans *Proceedings of TALN*, 1, 2001.

[16] Boissière, P., Vigoureux, N., Mojahid, M., Sauzin, D. et Vella, F. « TALN et IHM : Une approche transdisciplinaire pour la saisie de textes de personnes en situation de handicaps », dans *22eme Conference sur le Traitement Automatique des Langues Naturelles (TALN 2015)*, 2015.

[17] Boissière, P., Vigoureux, N., Mojahid, M., Sauzin, D. et Vella, F. « Adaptation of AAC to the context communication : A real improvement for the user illustration through the VITIPI word completion », dans *International Conference on Computers for Handicapped Persons*, Springer, 2012, p. 451-458.

[18] Stoop, W. et Van Den Bosch, A. « Improving word prediction for augmentative communication by using idiolects and sociolects », dans *Dutch Journal of Applied Linguistics*, 3(2), 2014, p. 136-153.

[19] Blache, P. et Rauzy, S., « Le moteur de prédiction de mots de la plateforme de communication alternative », dans *TAL Traitement Automatique des Langues*, 48(2), 2007, p. 47-70.

[20] Trost, H., Matiasek, J. et Baroni, M. « The Language Component of the FASTY Text Prediction System », dans *Applied Artificial Intelligence*, 19(8), 2005, p. 743-781.

[21] Renaud, A., Shein, F. et Tsang, V. « Grammaticality judgement in a word completion task », dans *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing : Writing Processes and Authoring Aids*, ACL, 2010, p. 15-23.

[22] 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. et Aiden, E. L. « Quantitative analysis of culture using millions of digitized books », dans *Science*, 331(6014), 2011, p. 176-182.

[23] MacKenzie, I. S. et Soukoreff, R. W. « Phrase sets for evaluating text entry techniques », dans *CHI'03 extended abstracts on Human factors in computing systems*, 2003, p. 754-755.

[24] Brooke, J. « SUS: A quick and dirty usability scale », dans *Usability evaluation in industry*, 189(194), 1996.

[25] Lesher, G. W., Moulton, B. J. et Higginbotham, D. J, « Effects of ngram order and training text size on word prediction », dans *Proceedings of the RESNA'99 Annual Conference*, Citeseer, 1999, p. 52-54.

[26] Garay-Vitoria, N. et Abascal, J. « Intelligent Word-Prediction to Enhance Text Input Rate (A Syntactic Analysis-Based Word-Prediction Aid for People with Severe Motor and Speech Disability) », dans *Proceedings of the 2nd international conference on Intelligent user interfaces*, ACM, 1997, p. 241-244.

[27] Hunnicutt, S. et Carlberger, J. « Improving Word Prediction Using Markov Models and Heuristic Methods », dans *AAC Augmentative and Alternative Communication*, 17(4), 2001, p. 255-264.

[28] Ghedira, S. « Optimisation de la communication pour personnes handicapées », dans *22eme Conference sur le Traitement Automatique des Langues Naturelles (TALN 2015)*, 2015.

[29] Antoine, J.-Y., Lefeuvre-Halftermeyer, A. et Allegre, W. « Pour une réflexion éthique sur les conséquences de l'usage des NTIC : Le cas des aides techniques (à composante langagière ou non) aux personnes handicapées », dans *Journée ATALA Ethique et TAL*, 2014.

[30] Lefeuvre-Halftermeyer, A., Goevaere, V., Antoine, J.-Y., Allegre, W., Pouplin, S., Departe, J.-P., Slimani, S. et Spagnulo, A., « Typologie des risques pour une analyse éthique de l'impact des technologies du TAL », dans *TAL et Ethique*, 57(2), 2016, p. 47-71.

[31] Antoine, J.-Y., Labat, M.-E., Lefeuvre-Halftermeyer, A. et Toinard, C., « Vers une méthode de maîtrise des risques dans l'informatisation de l'aide au handicap », dans *Envirorisk'2014, Le forum de la gestion des risques technologiques, naturels et sanitaires*, Cépaduès, 2014.

[32] Norré, M. *Configuration et évaluation d'un système de prédiction de mots au sein d'un logiciel de Communication Améliorée et Alternative (CAA) pour personnes handicapées : Étude de Presage et ACAT*. Mémoire, Université libre de Bruxelles, 2018.