

Proposition d'un cadre méthodologique d'évaluation des systèmes d'assistance à la saisie de textes : Applications aux systèmes Sibylle et VITIPI

Philippe Boissière, Igor Schadle¹

Université Paul Sabatier IRIT, 118, Route de Narbonne, 31062 Toulouse Cedex 9, France

{boissier,schadle}@irit.fr

Résumé : Après un rappel sur les besoins des systèmes d'aide à la saisie de texte pour les personnes handicapées, cet article pose le problème de leur évaluation. En effet, si plusieurs de ces systèmes existent au niveau de la recherche, tous sont évalués dans des cadres, sur des textes et avec des métriques différentes. C'est pourquoi, après un bref passage en revue des systèmes existants, nous proposons une méthodologie et un cadre d'évaluation commun pour de tels systèmes. Cette évaluation se base sur une métrique facile à calculer et à interpréter, un recueil de textes « écologiques » (i.e. écrits par une personne handicapée), et un protocole d'évaluation. Cette méthodologie est ensuite appliquée à deux de ces systèmes : VITIPI et Sibylle. Cet article se termine sur les suites à donner à cette méthodologie et sur les apports de ce cadre commun aux deux systèmes VITIPI et Sibylle.

I. INTRODUCTION

Le cadre dans lequel va se situer notre article est celui de l'assistance à l'écriture pour les personnes handicapées. Certaines d'entre elles ne peuvent pas utiliser le clavier ordinaire, et sont obligées d'avoir recours à des dispositifs spécialisés qui peuvent être une adaptation de matériels (claviers élargis, guides doigt), ou des programmes spécialisés qui simulent le fonctionnement du clavier à l'écran [29]. Selon les capacités motrices de la personne, la sélection de la touche se fera par un dispositif de pointage piloté par un système plus ou moins complexe (souris ordinaire, track ball, joystick, suivi du regard). Si elle ne peut pas utiliser de dispositif de pointage, il lui restera la possibilité d'utiliser un système à défilement. Dans ce cas, les lettres vont s'afficher l'une après l'autre, et quand la lettre désirée est mise en saillance, l'utilisateur appuie sur un contacteur pour sélectionner la lettre. Il existe différentes méthodes de balayage plus ou moins optimisées [9]. Il va de soi que quel que soit le moyen d'écriture utilisé par la personne handicapée, sa vitesse d'écriture restera très lente et il sera nécessaire d'utiliser des systèmes d'aide à la saisie de textes pour compenser en partie cette lenteur et prévenir une fatigue musculaire et/ou oculaire dans le cas des maladies neuromusculaires [1], [2], [30].

Les systèmes d'aide à la saisie de textes sont nombreux et doivent être évalués. Or cette évaluation pose des questions méthodologiques que nous exposerons. Nous présenterons une typologie de ces systèmes et décrirons les deux systèmes proposés à l'évaluation : Sibylle et VITIPI. Nous proposerons ensuite une méthodologie d'évaluation que nous appliquerons à ces deux systèmes.

II. PROBLÉMATIQUE DE L'ÉVALUATION DES SYSTÈMES D'ASSISTANCE À LA SAISIE.

Les systèmes d'aide à la saisie de textes sont relativement nombreux. Nous en citerons quelques-uns dans le paragraphe suivant. Leurs performances sont variables. Évalués le plus souvent par les concepteurs eux-mêmes, ils sont fonction de plusieurs paramètres.

Le premier d'entre eux concerne le type de textes utilisés pour mesurer les performances. Les tests peuvent être effectués sur des textes totalement hétéroclites. Certains utilisent des textes de *Time Magazine* [13], du British National Corpus [28], *Le Livre de la Jungle* de R. Kipling [16], des recueils de prévisions météorologiques [5] ou sur des articles du journal *Le Monde* [22]. Tous ces corpus représentent des volumes de textes relativement volumineux. Par conséquent, les résultats sont fiables sur un plan quantitatif, mais ces corpus sont-ils représentatifs des écrits réellement produits par les personnes handicapées auxquels ces systèmes s'adressent ?

Le deuxième paramètre concerne la métrique utilisée pour mesurer la performance du système. Le gain de saisie peut se mesurer par l'entropie en comptant le nombre de caractères par bit [32]. On peut également estimer le nombre de mots écrits par minute de façon théorique [27], de façon empirique par estimation de paramètres [28], ou par des tests effectués auprès d'utilisateurs [30], [31]. D'autres auteurs préfèrent calculer le gain de caractères économisés pour 100 caractères écrits [1] ou le pourcentage de lettres écrites par le système [33], [6], [8]. [27] propose une métrique pour déterminer le taux d'erreurs, nous la détaillerons dans le chapitre V.

¹ Pendant son post-doctorat à l'IRIT

L'applicabilité de ces diverses métriques rend par conséquent difficilement comparable les systèmes d'assistance à la saisie de textes entre eux. En effet, il est difficile pour une personne handicapée de se représenter l'économie de x bits par seconde. Une autre limite concerne la vitesse de saisie dépendante de l'interface de présentation des items prédits. De plus [31] ont montré que ce paramètre n'est pas significatif de la difficulté de saisie pour les personnes handicapées motrices. Comme dans le cadre du programme TECHNOLANGUE² pour les outils du TALN (Traitement Automatique du Langage Naturel) la question se pose de définir un cadre méthodologique de référence adapté à l'évaluation des systèmes de prédiction.

Le troisième paramètre à considérer est celui de l'Interaction Homme Machine (IHM). Comment celle-ci est elle conçue ? Va-t-elle demander à l'utilisateur de lui proposer des abréviations [20], [28] ? Va-t-elle lui renvoyer des listes de mots [2] ou ne proposera-t-elle qu'une suite de caractères [4] ? D'autres caractéristiques sont aussi à prendre en compte : le système prédit-il les mots de façon isolée ou tient-il compte des mots qui précèdent ? Corrige-t-il les fautes de frappes, d'orthographe ? Fonctionne-t-il indépendamment du traitement de textes ? etc. Les lecteurs trouveront une liste de ces caractéristiques dans [8].

L'évaluation des systèmes d'aide à la saisie soulève au moins deux types de problèmes. Le premier, est le choix d'une base de textes unique sur laquelle tous les systèmes d'une langue donnée pourront être évalués. À cette base de textes pourra être associée une base de connaissance lexicale ayant une couverture linguistique plus ou moins représentative. Cette Base de Connaissances (BdC) est composé d'un lexique et de règles d'attributs syntaxiques voire même de traits sémantiques selon les systèmes. Si pour une application finalisée le lexique peut être réduit, en revanche, le nombre de contraintes entre chacun de ses mots peut devenir combinatoire et difficile à maîtriser. Le second consiste à déterminer une métrique permettant de mesurer objectivement le rendement de tous les systèmes qui puisse être compréhensible par les utilisateurs.

L'objectif de cette publication est de proposer un cadre commun d'évaluation pour les systèmes d'assistance à la saisie. Nous appliquerons cette méthode aux systèmes Sibylle [22], [24] et VITIPI [3], [4]. Regardons à présent un bref aperçu des systèmes existants.

III. TYPOLOGIE DES SYSTÈMES EXISTANTS

D'un point de vue de la prédiction des mots, les systèmes d'aide à la saisie peuvent se répartir en deux classes : Les systèmes *descendants* et les systèmes *ascendants*.

Dans les systèmes *descendants*, le linguiste élabore manuellement une BdC à partir des règles grammaticales connues de façon explicite. On peut citer les systèmes comme Kombe [11], [21] pour une discussion entre le médecin et le patient, Illico [17], [18], HandiAs [12], [14], [15].

À l'opposé, les systèmes *ascendants* sont fondés sur l'hypothèse que les règles de grammaire sont implicitement encodées dans le texte. Pour les générer automatiquement, le système analyse de grosses quantités de textes. Parmi eux, on peut citer les systèmes Sibylle [22] et VITIPI [3] et [4]. La limite de cette approche est que ces systèmes sont très dépendants des tailles des corpus de textes soumis pour l'apprentissage et peuvent être également trop spécifiques de l'usage de la langue pour un utilisateur déterminé.

L'évaluation des systèmes ne peut porter que sur les domaines pour lesquels une BdC a pu être construite (modélisée). Ceci rend difficile les comparaisons inter-systèmes, difficiles en raison de leur dépendance vis à vis des connaissances linguistiques modélisées. Une autre contrainte est la modalité de la communication engendrée : communication textuelle, saisie de texte, Short Message Service (SMS) etc.

Ces deux facteurs imposent que les BdC doivent être adaptées à la situation et à la modalité de communication.

IV. DESCRIPTION DES SYSTÈMES DE L'ÉTUDE

A. Le système Sibylle

Sibylle a été conçu en collaboration avec le CMRRF de Kerpape pour des personnes IMOC (Infirme Moteur d'Origine Cérébrale) utilisant un bouton poussoir comme interface d'accès à l'ordinateur [24]. L'interface de l'application présente principalement une zone d'édition de texte un clavier de lettres et une liste de mots (figure 1).



Figure 1 : Interface principale de Sibylle.

L'originalité de Sibylle est de présenter à la fois une liste de mots et un clavier de lettres dynamiques [23]. En ce qui concerne la prédiction de mot, Sibylle utilise un modèle de

² <http://www.technolangue.net/>

langage statistique intégrant la notion de *chunks* [25]. C'est ce module, seul, que nous évaluerons dans le cadre de cet article.

B. Le système VITIPI

VITIPI est un système d'assistance à la saisie de textes. Il repose sur un modèle de prédiction de parties de mots qui les affiche dès qu'il n'y a plus d'ambiguïté lexicale. VITIPI prend en compte les fautes de frappes, certaines fautes d'orthographe au fur et à mesure de la saisie, ainsi que des mots n'appartenant pas à son vocabulaire de base tout en continuant de prédire des lettres.

VITIPI ne possède aucune information syntactico-sémantique. Il utilise un modèle assimilable à des n-grammes, sans probabilité, représenté sous la forme d'un transducteur unique qui permet à la fois d'afficher automatiquement les lettres en fonction d'un lexique donné, et de gérer l'enchaînement des mots. Ce transducteur et son lexique associé constituent la BdC du système. L'utilisateur a la possibilité de définir sa propre BdC à partir de textes déjà écrits ; celle-ci se génère automatiquement. Le système VITIPI fonctionne, soit avec une BdC vide, soit avec le lexique des mots les plus courants de la langue française extraits de N. Catach [10]. Un lexique où figure 5 930 mots permet de prédire 26 % de lettres.

C. Caractéristique des systèmes

Un modèle de comparaison basé sur des caractéristiques des différents systèmes a été proposé par Ph. Boissière [5], [8] pour évaluer les systèmes d'aides à la saisie. Les Tableaux 1 et 2 reprennent la plupart de celles-ci en séparant les caractères propres à l'IHM et au TALN.

Caractéristiques IHM	Sibylle	VITIPI
Méthodes d'abréviations	NON	NON
Présentation de liste de mots	OUI	option
Complétion automatique du mot sans présentation de liste	NON	OUI
Indépendance du traitement de textes	NON	OUI

Tableau 1 : Comparaison des caractéristiques IHM de Sibylle et VITIPI.

Caractéristiques TALN	Sibylle	VITIPI
Insertion automatique d'espace	OUI	OUI
Majuscule automatique en début de phrase	OUI	OUI
Accentuation automatique	NON	NON
Prise en compte des fautes de frappe à la saisie	NON	OUI
Prise en compte des fautes d'orthographe à la saisie	NON	OUI option
Fonctionnement en mots isolés	OUI	OUI option
Fonctionnement avec segments de phrases	OUI	OUI
Prise en compte de nouvelles entités	OUI	OUI
Mise à jour immédiate de la BdC	OUI	OUI
Génération automatique de la BdC pour une langue donnée	NON	OUI

Tableau 2 : Comparaison des caractéristiques TALN de Sibylle et VITIPI.

Nous constatons au travers de la description des capacités de ces deux systèmes des fonctionnalités différentes. L'objectif d'une méthodologie d'évaluation est de tenir compte de ces différences fonctionnelles et linguistiques.

V. MÉTHODOLOGIE D'ÉVALUATION

La méthodologie d'évaluation que nous proposons est principalement axée sur une évaluation linguistique. Notre objectif est de quantifier de façon la plus objective possible les performances du système de prédiction en mots. Pour cela, nous définirons une métrique. Nous décrirons ensuite les caractéristiques des corpus textuels employés pour comparer les deux systèmes. Nous expliquerons enfin le protocole utilisé.

A. La métrique utilisée

Pour calculer son taux d'erreur, Soukoreff [27] raisonne ainsi : Lorsque l'utilisateur s'aperçoit qu'il fait une erreur, il revient en arrière pour l'effacer, puis frappe la lettre correcte. Pour chaque erreur, cela fait deux caractères saisis en plus qui ne vont pas apparaître en sortie. En conséquence, le nombre de caractères réellement tapés par l'utilisateur (*Flux d'entrée*) sera supérieur ou égal au nombre de caractères affichés (*Flux de sortie*). Soukoreff [27], définit le KSPC (KeyStroke Per Characters) par :

$$KSPC = \frac{\text{Flux.d'entrée}}{\text{Flux.de.sortie}} \quad [1]$$

L'objectif d'un système de prédiction, c'est que le message écrit par l'utilisateur (flux d'entrée) doit être plus court que celui qui s'affiche (flux de sortie). Or, l'équation [1] indique en fait le pourcentage de lettres écrites par l'utilisateur. Pour connaître le rendement du système ou taux d'économie de saisies il suffit donc de prendre le complément de [1]. Nous obtenons ainsi la métrique suivante :

$$KSPC = 1 - \frac{\text{Nb car. saisis} + \text{Nb touches de fonction}}{\text{Nb car. du texte}} \quad [2]$$

avec :

Nb car. saisis: Nombre de caractères saisis par l'utilisateur ;

Nb touches de fonctions: Nombre de fois où une touche de fonction a été tapée. Une touche de fonction peut être une touche de correction (par exemple, pour une proposition erronée dans VITIPI) ou une touche correspondant à un mot de la liste de mots dans Sibylle (à une liste de X mots sera associée X touches de fonction) ;

Nb car. du texte: Nombre total de caractères du texte.

C'est cette équation [2] que nous nous référerons pour la suite du texte.

En considérant que l'utilisateur ne fera qu'une *sélection* par caractère et par touche de fonction, ce taux exprime le taux d'économie de saisie. Il vaudra 0 % dans le cas d'une saisie sans système de prédiction et 100 % si le texte est entièrement prédit sans aucune saisie. Cette métrique a pour avantage d'être indépendante de l'interface utilisateur. Elle peut s'appliquer à tous les systèmes et à toutes les langues. Elle est plus facilement interprétable pour un utilisateur (pas forcément informaticien ou ergothérapeute) que le nombre de caractères par bit ou l'entropie, plus communément utilisée pour les claviers logiciels.

Au-delà de cette proposition du taux de rendement, nous proposons de prendre en compte l'activité même de saisie des personnes handicapées. Celle-ci peut concerner aussi bien une rédaction formelle d'un article scientifique ou une communication textuelle avec un style plus relâché (courriels, communication MSN qui a son propre style de langage par exemple, etc.).

B. Le recueil de textes

Nous avons tenu à ce que notre évaluation soit la plus « écologique » possible, c'est-à-dire la plus proche d'une situation réellement rencontrée par une personne handicapée. Pour cela nous avons constitué un corpus de textes (courriels) écrits par une personne handicapée. Celle-ci est âgée d'une quarantaine d'années ; elle a une Infirmité Motrice Cérébrale, sans trouble de l'écriture, et dispose d'un niveau d'étude universitaire. Ce corpus est constitué de 205 courriels issus de sa correspondance personnelle, soit : 1 400 phrases composées de 19 640 mots produisant un vocabulaire de 2 712 mots. Le nombre total de caractères est de 106 813. Ce corpus a été scindé en vingt-six sous-corpus d'à peu près égale taille pour permettre de tester les capacités d'apprentissage des systèmes avec l'évolution du taux d'économie de saisie au cours de ces vingt-six sous corpus. Chaque sous corpus contient 7,8 messages en moyenne, ce qui correspond à environ 4 100 caractères. Tous les noms propres ont été changés afin de rendre ce corpus anonyme et utilisable par d'autres. Nous avons ensuite procédé à la suppression des entêtes et des signatures des courriels. Puis nous avons effectué des corrections comme les erreurs de frappe sur les espaces (doublons, espace avant une virgule, etc.). Dans le même ordre d'idée, nous avons également corrigé des fautes d'orthographe comme « je souhaiterait ». Par contre, nous avons conservé tout ce qui concerne les usages propres à l'utilisateur comme la mise en relief avec des lettres capitales, l'utilisation de *smiles* : -), de mots comme @+ (à plus tard) ou encore l'usage de répétition de points d'exclamations ou d'interrogations.

Notre objectif est de mesurer la qualité du modèle à prédire et non à reproduire des fautes. Comme le mentionne [23] l'utilisation de système d'aide réduit le nombre de fautes de frappes : le système VITIPI [3] corrige en moyenne 72% de fautes.

C. Procédé d'apprentissage

Le premier des 26 sous corpus a été testé avec la BdC initiale propre à chacun des deux systèmes d'étude. Chaque BdC a ensuite été enrichie à partir des textes de ce premier sous corpus. Celles-ci sont ensuite utilisées pour simuler la saisie du deuxième sous corpus, qui lui-même enrichira les versions ultérieures des BdC. Ce procédé d'apprentissage sera reproduit jusqu'au 26^{ème} sous corpus. Nous réalisons ainsi un apprentissage incrémental.

VI. RÉSULTATS

Nous allons discuter des résultats obtenus selon deux paramètres : la taille d'apprentissage des corpus et les modèles n-grammes. La Figure 2 et la Figure 4 présentent respectivement les résultats des deux systèmes étudiés (VITIPI

et Sibylle) : le taux d'économie de saisies est en ordonnées et les sous corpus en abscisses.

A. Évaluation de VITIPI

Le graphique (Figure 2) donne les résultats en pourcentage de prédiction de l'évaluation de VITIPI dans trois conditions d'expérimentation différentes : sans lexique initial et sans tenir compte des mots précédents (cas des uni-grammes, $N=1$ -Lexique), sans lexique initial mais en tenant compte des deux mots précédents (cas tri-grammes $N=3$ -Lexique) et enfin avec lexique (cas tri-grammes $N=3$ + Lexique). On observe qu'au début pour les premiers sous corpus, le lexique apporte une aide intéressante. Plus le nombre de messages augmente, plus le système apprend le vocabulaire propre à l'utilisateur. Dans les derniers sous corpus, le rôle du lexique n'a plus d'importance. Par contre, la prise en compte des mots précédents permet d'augmenter le rendement (près de 14 %). Rappelons que le système VITIPI ne propose pas une liste de mots contrairement au système Sibylle.

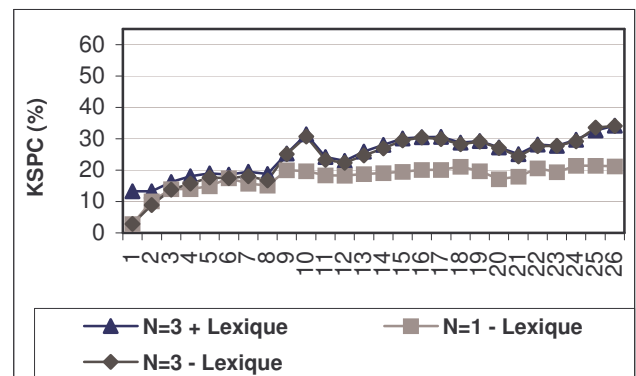


Figure 2 : Résultats de VITIPI

On retrouve des résultats analogues dans la Figure 3 où nous avons fait figurer sur l'axe des ordonnées, à la fois le taux de prédiction du système pour les tri-grammes (avec ou sans lexique) et le pourcentage de mots nouveaux rencontrés pour chaque sous corpus. Le nombre de mots nouveaux décroît jusqu'à 3 % pour les tests faits avec le vocabulaire de base (6 % sans ce vocabulaire) : ceci montre l'effet de l'apprentissage du système à l'utilisateur.

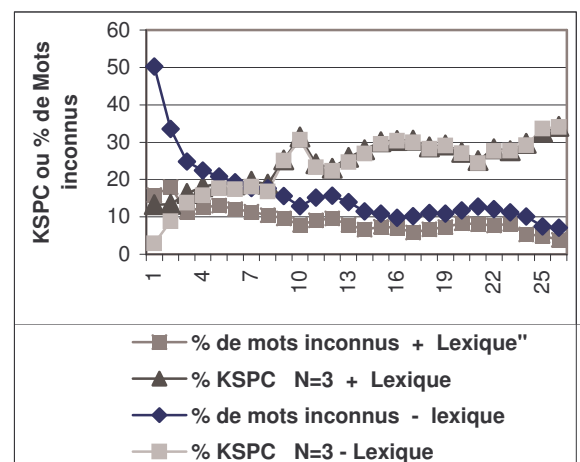


Figure 3 : Pourcentage des mots nouveaux dans chaque sous-corpus

B. Évaluation de Sibylle

Nous avons voulu évaluer les performances du système Sibylle en les comparant à trois modèles de référence en modélisation statistique du langage. Ces modèles sont le tri-gramme, le bi-gramme et l'uni-gramme qui prennent respectivement en compte les deux derniers mots, le dernier mot et aucun mot. Le graphique (Figure 4) permet de comparer le taux de rendement de Sibylle (en haut sur le graphique,) avec ces trois modèles de référence. À l'inverse de Sibylle (qui possède des connaissances grammaticales), ces modèles ne bénéficient cependant pas d'un apprentissage préliminaire. Pour ces quatre séries, nous avons considéré la présentation d'une liste de cinq mots. Les taux d'économie de saisies (KSPC) sont très proches pour le bi-gramme et le tri-gramme (quasi-confondues sur la Figure 4). Ce résultat s'explique compte tenu de la taille du corpus : le tri-gramme n'apporte un réel bénéfice qu'à partir de centaines de milliers de mots.

Quel que soit le modèle n-gramme, on constate l'intérêt de l'apprentissage. Par exemple, le système Sibylle démarre à un taux de rendement avoisinant les 50 %, augmente régulièrement pour terminer à un taux supérieur à 60 %. Enfin si le modèle bi-gramme et tri-gramme (qui n'ont pas de connaissances initiales) semblent rattraper le système Sibylle au début, l'écart se maintient constant autour de 7 % à partir du sous-corpus 9.

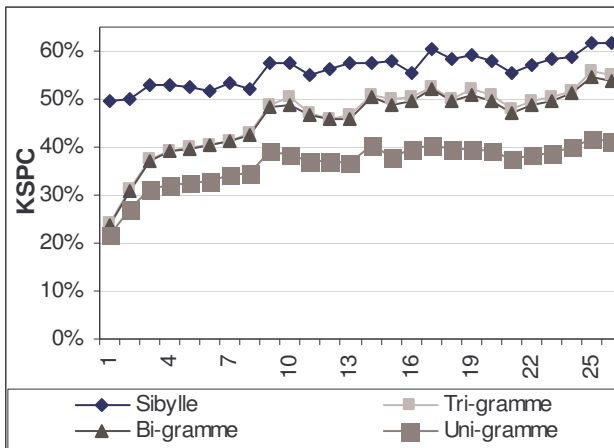


Figure 4 : Résultats de Sibylle

Nous avons ensuite testé l'effet de la taille de la liste de mots. Le système Sibylle a été évalué avec trois tailles différentes : des listes de 1, 5 et 10 mots. Les résultats (Figure 5) montrent que l'écart entre une liste de 1 mot et 5 mots, est de l'ordre de 12 à 13 % ; cet écart est cependant moindre (environ 4 %), entre les listes de 5 à 10 mots. Ces résultats pourront être analysés de façon plus fine par la mise en œuvre d'un protocole centré sur l'évaluation de l'interface homme machine. En effet, même si la charge cognitive s'accroît quand on passe d'une liste de 1 à 5 mots, la présentation d'une liste à 5 mots peut être envisageable vu le gain apporté. A contrario, le gain relativement faible apporté par le passage de 5 à 10 mots risque de ne pas contrebalancer le surcoût de charge cognitive

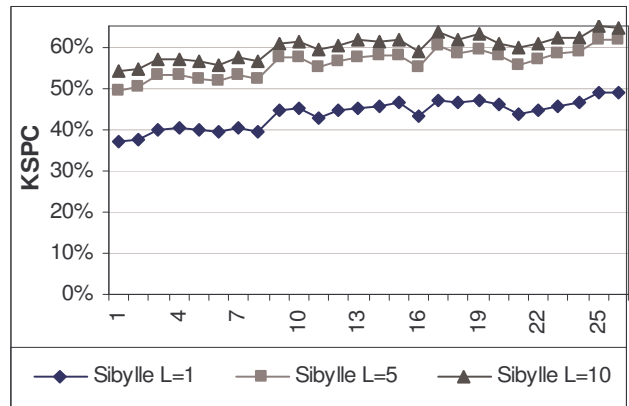


Figure 5: Influence du nombre de mots proposés par Sibylle sur le taux d'économie

VII. CONCLUSION

Cet article poursuit les travaux initiés par [5] sur la proposition d'un cadre d'évaluation commun aux systèmes d'aide à la saisie de texte. Pour ce qui concerne la partie purement informatique, ces systèmes reposent sur deux modules distincts d'importance égale.

- Le module qui assure la prédiction des mots, et éventuellement des phrases. Cette partie relève du TALN ;
- Le module qui assure la présentation des propositions et assure l'interface entre l'utilisateur et le système, c'est la partie IHM.

Cependant, nous pensons qu'il existe un troisième niveau dépendant du mode d'accès à l'ordinateur (souris, bouton poussoir, guide doigt, etc.) et du degré de liberté qu'il permet. Ici s'arrête les compétences de l'informaticien, là commencent celle de l'ergothérapeute qui doit assurer le lien entre l'outil informatique – aussi adapté soit-il –, et les capacités motrices et cognitives de l'utilisateur. La conception de l'interface du clavier virtuel est étroitement liée à cette modalité et deux systèmes ne peuvent être comparés sans prendre en compte cet aspect.

Cet article s'est plus précisément intéressé à l'évaluation du module linguistique. Nous avons proposé une métrique dont le but est d'être générale (applicable à tout système), indépendante de l'interface et facilement interprétable. Celle-ci a été appliquée à l'évaluation de deux systèmes d'assistance à la saisie. Ainsi, ce cadre commun a nécessité au système Sibylle une évolution de son modèle de langage pour permettre un apprentissage « rapide » des données utilisateur (le système avait été évalué uniquement sur un gros corpus du journal Le Monde). De même, cette évaluation commune incite à introduire des connaissances de niveau supérieur pour VITIPI.

Parallèlement, nous avons utilisé un corpus de test représentatif d'une situation de communication correspondant à des besoins de sujets handicapés, comme on peut en rencontrer en situation réelle³.

Nous souhaitons poursuivre ce travail de définition d'un cadre méthodologique d'évaluation des systèmes de prédiction en invitant les concepteurs francophones à se joindre à ce travail. Nous envisageons également de recueillir des

³ (<http://www.irit.fr/chatcom/Questionnaire.html>)

corpus relatifs à d'autres situations de communication (communication textuelle, rédaction libre, etc.). Le corpus utilisé est relativement exempt de *grosses erreurs linguistiques*, même si nous avons apporté quelques corrections typographiques et orthographiques. Concernant les perspectives au niveau de la partie TALN, nous aimerions tester les différents modules de prédiction avec des productions fortement agrammaticales provenant de sujets dyslexiques ou aphasiques. Pour la partie interface Homme-Machine, l'évaluation est beaucoup plus difficile à réaliser du fait de la très grande variabilité inter et intra handicap. Ce qui peut être utilisable pour un individu peut très bien se révéler inutilisable pour un autre, même s'ils ont à peu près la même pathologie. Seule une observation « in vivo », sur le terrain avec les aménagements indispensables dont la personne a besoin, permet de valider ou pas l'utilisabilité d'une interface pour une personne donnée. Pour ce faire, la plate forme E-Assiste [19] conçue à l'IRIT se révèle être un outil très efficace pour ce type d'évaluation.

VIII. REMERCIEMENTS

Merci à Nadine Vigouroux et à Frédéric Vella pour avoir relu cet article et pour leurs suggestions.

IX. BIBLIOGRAPHIE

- [1] Berard C. Clavier-écran : concevoir avec les utilisateurs. IFRATH Handicap 2004, Paris, 17-18 Juin 2004, pp. 83-88.
- [2] Berard C., Neimeijer D. Evaluating effort reduction through different word prediction systems. IEEESMC, La Haye 2004 0-7803-8566-7/04/S20, 00.
- [3] Boissière, P., Dours, D. VITIPI : Versatile Interpretation of Text Input by Persons with Impairments. 5th International Conference on Computers for Handicapped Persons, Linz July 1996, pp.165-172.
- [4] Boissière, P., Dours, D. Un système d'aide à l'écriture basé sur un principe d'auto apprentissage et adapté à tous les handicaps moteurs. Handicap 2000, IFRATH, Paris, Juin 2000, pp. 81-86.
- [5] Boissière Ph., Dours D. Vers un modèle d'aide à l'évaluation de systèmes d'assistance à l'écriture : application à VITIPI. Handicap 2002, IFRATH, Paris, Juin 2002, pp. 159-164.
- [6] Boissière Ph., Dours D. A Proposal of an Evaluation Framework for Writing Assistance Systems: Application to VITIPI. Proceedings of the 8th International Conference, ICCHP'2002, Linz, Austria, July 2002. (LNCS 2398), pp 276-278.
- [7] Boissière Ph., Dours D. Preliminary results about VITIPI evaluation efficiency. 7th ERCIM Workshop "User Interfaces for All", Special Theme: "Universal Access", Chantilly. ERCIM – INRIA, 24 – 25 octobre 2002, pp. 187-188.
- [8] Boissière Ph., Dours D. An evaluation framework for writing assistance systems: Application to VITIPI ». Modelling, Measurement & Control, Série C, (bioengineering). (AMSE) pp. 119 –128.
- [9] Cantegril, B., Toulotte, J.-M. Réflexions sur l'aide à la communication des personnes présentant un handicap moteur. Atelier Thématique TALN 2001, Tours, 2-5 Juillet 2001, pp. 193-202.
- [10] Catach N., Les listes orthographiques de base du Français (LOB). Nathan 1984.
- [11] Guenther, F., Krüger-Thielmann, K., Pasero, R. and Sabatier, P. Communication aids for handicapped persons. In 2nd European conference on the advancement of rehabilitation technology, ECART'93 Stockholm May 1993.
- [12] Le Pévédic, B. Prédiction Morphosyntaxique évolutive dans un système d'aide à la saisie de textes pour des personnes handicapées physiques. Thèse de Doctorat IRIN, 1997.
- [13] Leshner, G. W., Moulton, B. J., Higginbotham, D. J. Effects of n-gram order and training text size on word prediction. In Proceedings of the RESNA '99 Annual Conference, Arlington, VA: RESNA Press, pp. 52-54.
- [14] Maurel, D., Fourche., B., Briffault. HandiAS : Aider la communication en facilitant la saisie rapide de textes. In IFRATH Handicap 2000, Paris 15-16 Juin 2000, pp. 87-92.
- [15] Maurel, D., Rossi, N., Thibault, R. Handias : un système multilingue pour l'aide à la communication de personnes handicapées. In Atelier Thématique TALN 2001, Tours, 2-5 Juillet 2001, pp. 203-212.
- [16] Menier, G., Poirier, F. Système adaptatif de prédiction de texte. In Atelier Thématique TALN 2001, Tours, 2-5 Juillet 2001, pp. 213-222.
- [17] Pasero, R., Sabatier, P. Linguistic games for language Learning and tests, an ILLICO application. In Computer-Assisted Language Learning.
- [18] Pasero, R., Sabatier, P. Concurrent Processing for Sentences Analysis, Synthesis and Guided Composition. Natural Language Understanding and computational Logic, Lecture Notes in Computer Science, Springer 1998.
- [19] Raynal, R. Maubert, S. Vigouroux, N. Vella, F. Magnien, L.. *E-Assiste: A platform Allowing Evaluation of Text Input System*. Dans 3rd Int. Conf. on Universal Access in Human-Computer Interaction (UAHCI 2005), Las Vegas, USA, 22 juillet 27 juillet 2005. Lawrence Erlbaum Associates (LEA), ISBN 0-8058-5807-5, p. CDRom.
- [20] Ricco, X., Dutoit, T. Vers un logiciel multilingue et gratuit pour l'aide aux personnes handicapées de la parole : HOOK. Atelier Thématique TALN 2001, Tours, 2-5 Juillet 2001, pp. 223-232.
- [21] Richardet, N. Composition de phrases assistée - Un système d'aide à la communication pour handicapés. Thèse de doctorat, Université de la Méditerranée, Marseille, France, 1998.
- [22] Schadle, I. Sibylle : Système linguistique d'aide à la communication pour les personnes handicapées. Thèse de doctorat, Université de Bretagne Sud, 2003.
- [23] Schadle, I., Antoine J.-Y., Le Pévédic B., Poirier F. Sibylle : Prédiction de lettre pour la communication assistée. Revue RIHM, 3(2), 2003., pp. 115-133.
- [24] Schadle, I., Poirier F. Sibylle : un système d'aide à la saisie de texte. IHM'04, ACM Press, Namur, 2004., pp. 141-146.
- [25] Schadle I., Antoine J.-Y., Le Pévédic B., Poirier F. Sibylle - Modélisation stochastique du langage intégrant la notion de chunks. TALN 2004, Fès.
- [26] Sperandio J.-C. Critères ergonomiques d'assistance technologique aux opérateurs. JIM'2001 : Interaction Homme-Machine & Assistance pour GrandPublic / Systèmes Complexes / Handicap, Metz, 4-6 juillet 2001, pp. 30-37.
- [27] Soukoreff, R. W., & MacKenzie, I. S. Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI 2003, pp. 113-120. New York: ACM (2003) (<http://www.yorku.ca/mack/chi03.html>)
- [28] Väyrynen P. Perspectives on the Utility of Linguistic Knowledge in English Word Prediction. Thèse de l'Université d'OULU, 19 Novembre 2005 (<http://herkules oulu.fi/isbn951427850X/isbn951427850X.pdf>)
- [29] Vella F. Vigouroux N., Truillet Ph. SOKEYTO: a design and simulation environment of software keyboards. Dans : Assistive technology: from virtuality to reality - 8th European conference for the advancement of assistive technology in europe (AAATE 2005), Lille, France, 6 septembre 9 septembre 2005. A Pruski, H Knops (Eds.), IOS Press, ISBN 1-58603-543-6, p. 723-727.
- [30] Vigouroux N., Vella F., Truillet Ph., Mathieu Raynal. Evaluation of AAC for text input by two groups of subjects: able-bodied subjects and disabled motor subjects. Dans : 8th ERCIM Workshop "User Interfaces for All", Vienne, Autriche, 28 juin 29 juin 2004. (<http://www.ui4all.gr/workshop2004/publications/adjunct-proceedings.html>)
- [31] Vigouroux N., Vella F., Raynal M., Boissière Ph. Solutions et défis pour une meilleure accessibilité et utilisabilité des communicateurs - Optimisation de la saisie de texte. Dans : Handicap et Environnement : de l'adaptation du logement à l'accessibilité de la cité - Entretiens de l'Institut de Garches, Nanterre, 24 novembre 25 novembre 2005. -, p. 209-222.
- [32] Ward D. J. Adaptive Computer Interfaces, in Doctoral Thesis of Philosophy, Cambridge University, November 2001.
- [33] Zagler, W., Seisenbacher, G. German language predictive typing : Results from a feasibility investigation. Proceedings of the ICCHP'2000, Karlsruhe, 17-21 July 2000, pp.771-779.