



**HAL**  
open science

## Prédiction de mots et saisie de requêtes sur interfaces limitées : dispositifs mobiles et aide au handicap

Jean-Yves Antoine

► **To cite this version:**

Jean-Yves Antoine. Prédiction de mots et saisie de requêtes sur interfaces limitées : dispositifs mobiles et aide au handicap. Patrice BELLOT. Recherche d'information contextuelle, assistée et personnalisée, Hermès-Lavoisier, pp.273-298, 2011, Recherche d'information et web, 978-2-7462-2583-1. hal-01016521

**HAL Id: hal-01016521**

**<https://hal.science/hal-01016521>**

Submitted on 3 Jul 2014

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

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

## Chapitre 10

# Prédiction de mots et saisie de requêtes sur interfaces limitées : dispositifs mobiles et aide au handicap

### 10.1. Introduction : saisie de message sur interfaces limitées

Le développement d'Internet au cours de la dernière décennie a bousculé les usages de l'informatique en entraînant la démocratisation d'un outil jusqu'alors peu utilisé dans le cadre domestique. L'exemple de la France est ainsi édifiant : alors que seulement 16 % des foyers étaient équipés d'un ordinateur en 1997, cette proportion atteignait déjà 45 % en 2004. Et si 74 % des foyers français disposent en 2009 d'un ordinateur à domicile, cette proportion atteint 92 % dans les familles avec enfants.

La révolution Internet est donc juste derrière nous qu'une nouvelle ère se profile avec autant de fulgurance : celle de l'informatique mobile et ubiquitaire. A l'opposé de l'informatique de bureau ou à domicile, l'informatique ubiquitaire (ou ambiante) met en jeu de multiples systèmes à tout moment et dans n'importe quel lieu de votre vie quotidienne. On pense bien entendu aux téléphones intelligents (*Smartphones*) qui permettent à leur utilisateur mobile de communiquer, naviguer sur la Toile, prendre des notes, gérer un agenda ou encore lire des données multimédias. L'informatique ubiquitaire peut toutefois être transparente à l'usager, comme par exemple dans le cas des étiquettes d'identification par radiofréquence (RFID) utilisées dans les transports en commun.

La recherche d'information est directement concernée par cette évolution. Un des usages les plus répandus des téléphones mobiles intelligents (au premier titre desquels l'*IPhone*) est en effet la recherche d'une information ou d'un service sur la Toile. Si cette recherche est initiée par une requête à base de mots-clés ou d'un énoncé en langue naturelle, on se retrouve dans une problématique plus large : la saisie de texte sur interface limitée. On entend par là que l'utilisateur ne dispose pas d'un clavier standard du fait des dimensions réduites du dispositif utilisé : il peut s'agir par exemple d'un clavier de téléphone à nombre de touches réduites, ou d'un clavier virtuel affiché sur un écran tactile. Dans tous les cas, la vitesse de composition des messages est ralentie par le caractère limité du dispositif d'entrée. On observe également souvent une augmentation des erreurs de saisie.

L'ingénierie des langues peut proposer des outils à même de compenser ces insuffisances. C'est en particulier le cas de la prédiction linguistique, qui fait l'objet de ce chapitre : si le système est capable de prédire correctement les prochaines lettres ou mots que l'utilisateur souhaite saisir, la sélection des hypothèses correspondantes va accélérer la composition des messages et éviter certaines erreurs. Les téléphones mobiles possèdent déjà des modules de prédiction relativement basiques. Dans ce chapitre, nous allons étudier plusieurs modèles plus ou moins évolués et comparer leurs performances respectives.

Avant d'aller plus avant, notons toutefois que la problématique de la saisie de texte sur interface limitée n'est pas née avec l'informatique mobile. Cela fait en effet près de trois décennies qu'elle a retenu l'attention des chercheurs travaillant sur l'aide à la communication pour personnes handicapées (AAC pour *Alternative and Augmentative Communication* en anglais). Ces systèmes ont pour objectif de restaurer les capacités de communication de personnes souffrant d'un handicap moteur très sévère se traduisant par une tétraplégie ou une athétose accompagnée souvent d'une perte de l'usage de la parole. Les capacités motrices très limitées de la personne handicapée empêchent toute saisie directe de message sur un clavier d'ordinateur : le plus souvent, celle-ci ne va pouvoir piloter son ordinateur que par une interface adaptée ; tout ou rien n'autorisant que l'équivalent d'un simple clic. Ici, le caractère limité de l'interface provient donc des capacités motrices limitées des utilisateurs et non pas du système informatique utilisé (un ordinateur portable fixé sur le fauteuil électrique de la personne). La miniaturisation continue des systèmes informatiques permet toutefois une convergence remarquable des besoins de l'aide à la communication et de l'informatique mobile : de plus en plus, les personnes handicapées demandent à utiliser non plus un ordinateur encore assez lourd mais un micro-portable (*Netbook*) ou un téléphone intelligent, bien plus discrets. Dans une perspective de *design for all* cher à la communauté du handicap,

les modèles de prédiction que nous étudierons dans ce texte concerneront donc tout type d'utilisateur mobile, qu'il soit valide ou installé sur un fauteuil électrique.

Dans un premier temps, nous allons situer la problématique de l'aide à la saisie de message en décrivant les différents dispositifs d'entrée qui peuvent être utilisés dans ces usages mobiles. Cette étude nous permettra de saisir l'importance de la prédiction linguistique pour l'aide à la composition de message. Nous présenterons ensuite en détail différents modèles de prédiction, en insistant plus particulièrement sur les techniques les plus avancées en matière d'adaptation contextuelle de la prédiction. Notre propos s'appuiera sur des résultats d'évaluation expérimentale afin de situer l'intérêt de chaque technique étudiée.

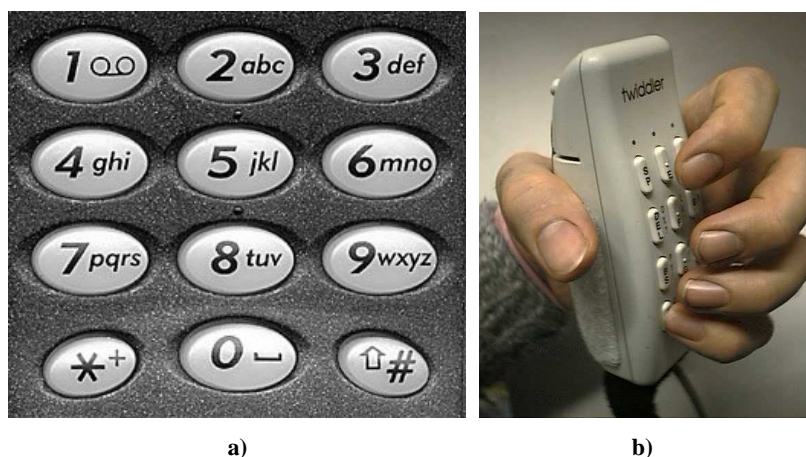
## **10.2. Interfaces limitées : dispositifs d'entrée**

### **10.2.1. Dispositifs d'entrée de texte pour systèmes informatiques mobiles**

De par la diversité des systèmes, l'informatique ubiquitaire met en jeu une grande diversité de dispositifs d'entrée de texte. Si l'on se limite aux *Smartphones*, quatre grandes classes de méthodes d'entrée de texte peuvent être distinguées.

En premier lieu, on trouve les claviers physiques à nombre limité de touches, directement issus du clavier à 12 touches des téléphones classiques. Pour la saisie de texte, ce type de clavier tient lieu de dispositif d'entrée ambigu. En effet, plusieurs caractères sont associés à chacune des touches (figure 10.1a).

Pour entrer un message sur un tel dispositif ambigu, deux solutions sont envisageables. La première consiste à laisser à l'utilisateur la charge de la commande de désambiguïsation. Le paradigme multi-appuis demande ainsi à l'utilisateur d'effectuer un nombre d'appuis variable sur la même touche pour distinguer les différents caractères qui lui sont associées (par exemple, un appui sur la touche « 2 » pour saisir la lettre *a*, deux pour la lettre *b* etc.). L'autre approche consiste à confier la désambiguïsation au système, l'utilisateur n'effectuant qu'un appui par caractère saisi. Cette désambiguïsation à la volée a été popularisée par la société Tegic Communication et son mode de saisie T9. Dans ce cas, le système consulte un dictionnaire interne pour filtrer les séquences de mots correspondant à l'entrée ambiguë. En cas d'hypothèses multiples subsistantes, l'affichage des différentes hypothèses permet à l'utilisateur de préciser son choix. Dans cet exemple, la désambiguïsation ne fait appel qu'à un dictionnaire statique.



**Figure 10.1.** *Claviers physiques limités a) clavier ambigu téléphonique standard (12 touches), b) clavier à accords Twiddler*

Elle peut également s'appuyer sur une analyse du contexte de saisie, le filtrage des hypothèses dépendant des mots ou caractères déjà saisis. On se retrouve alors dans la problématique de la prédiction linguistique qui va nous intéresser dans ce chapitre. Notons que désambiguïsation statique et prédiction lexicale peuvent consister en des solutions complémentaires.

Beaucoup moins répandus, d'autres claviers physiques à nombre de touches limitées permettent une entrée de texte non ambiguë. Il s'agit des claviers à accords (*chording keyboard* en anglais) qui s'utilisent en réalisant un appui simultané sur plusieurs touches par caractère saisi (figure 10.1b). Contrairement aux apparences, la prise en main d'un tel clavier est rapide. Peu connu du grand public, les claviers à accord sont utilisés par les télétypistes (prise de texte lors de débats parlementaires, par exemple) et pour tout usage avec une seule main libre. La démocratisation de ces dispositifs d'entrée est parfois annoncée avec celle de l'informatique vestimentaire. Dans l'immédiat, on notera que la saisie n'est pas ambiguë et ne nécessite donc pas de techniques de prédiction linguistique.

Au vu des dernières évolutions du domaine, il est à prévoir que les systèmes mobiles sans clavier physique sont amenés à se généraliser à l'avenir. La figure 10.2 montre ainsi l'interface utilisateur de l'*iPhone* d'*Apple* en mode saisie de message. La solution retenue ici consiste à afficher à l'écran un clavier virtuel utilisable par appui tactile, avec le doigt ou un stylet.



Figure 10.2. Clavier virtuel sur téléphone intelligent : exemple de l'iPhone (Apple)

En dépit de la ressemblance avec le clavier standard d'un ordinateur de bureau, la saisie de message sur de tels dispositifs est ralentie par la taille réduite des touches et l'absence d'un *feedback* proprio-réceptif qui favorise également les erreurs de saisie. Ces dispositifs s'accompagnent donc souvent d'une prédiction linguistique embarquée permettant là encore d'accélérer la saisie et de réduire les erreurs. Sur la figure 10.3a, on voit ainsi la liste de trois mots (*over*, *ogre*, *overall*) proposé par le système. Un appui sur les touches de défilement permet à l'utilisateur de choisir la bonne proposition.

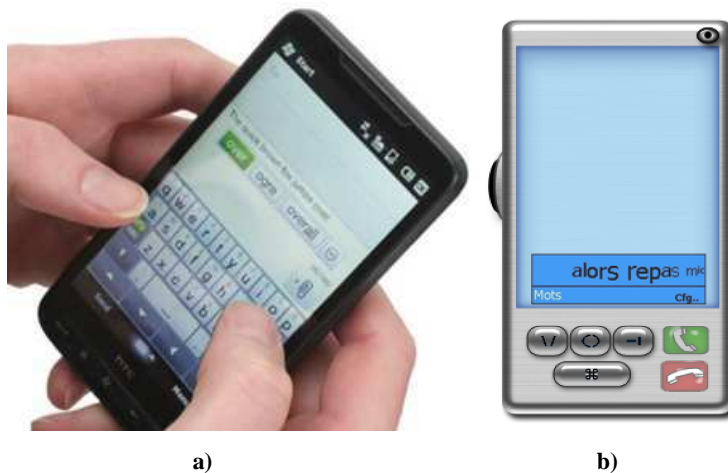


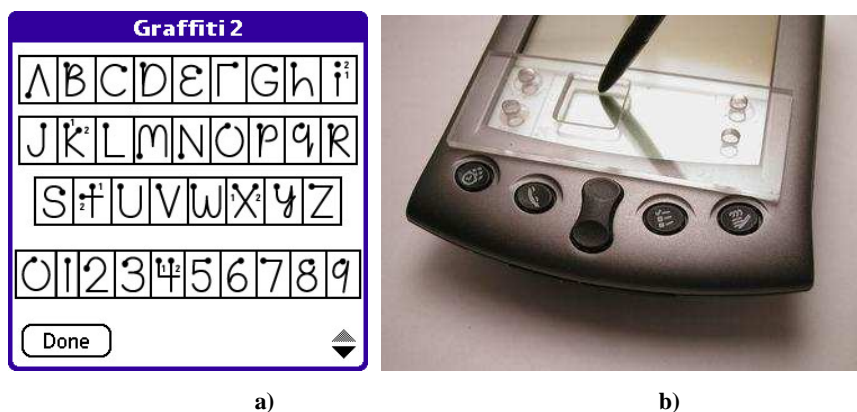
Figure 10.3. Claviers virtuel avec prédiction de mots, a) clavier alphabétique classique, b) clavier à trois primitives graphiques UniGlyph

L'usage de la prédiction n'empêche toutefois pas que la taille des touches du clavier virtuel reste très réduite. Le concepteur du système doit alors faire face à des contraintes antinomiques : augmenter la taille des touches tout en faisant en sorte que le clavier laisse le maximum d'espace libre pour la zone de saisie.

Une solution à ce dilemme est de limiter le nombre de touches sur le clavier : on se retrouve alors avec un clavier virtuel ambigu. Si cette approche est assez peu développée sur des systèmes commercialisés, elle a fait l'objet de plusieurs propositions de recherche. Parmi les plus originales, citons le clavier UniGlyph [BEL 07] qui ne comporte que trois couches, ce qui permet de bien dégager l'interface.

Ici, chaque touche correspond à des primitives graphiques regroupant en moyenne 12 caractères (figure 10.3b) : une touche correspond aux caractères comportant un trait incliné, comme la lettre V, une autre aux caractères arrondis et une dernière aux caractères avec barres verticales ou horizontales.

Quel que soit le nombre de touches retenues et leur disposition, le recours à la prédiction linguistique est incontournable puisque la séquence d'appuis doit être désambiguïsée. On aperçoit la liste de mots prédits (*alors, repas*) par le système sur la figure 10.3b consacrée au système UniGlyph. Ceux-ci sont présentés à l'aide d'un menu zoomable (*fish eye*).



**Figure 10.4.** a) *alphabet Graffiti2 (PalmOS) et b) système de guidage du stylet Edgewrite combiné à une version adaptée de l'alphabet Unistroke*

La dernière approche ignore le clavier et propose directement à l'utilisateur une saisie manuscrite sur écran tactile, en utilisant un stylet. Développée initialement sur

les organisateurs personnels (PDA), cette approche semble en perte de vitesse avec la généralisation des téléphones intelligents englobant *de facto* un PDA mais utilisant une interface à clavier. Compte tenu du manque de fiabilité de la reconnaissance d'écriture, ce paradigme de saisie de texte requiert l'utilisation d'un alphabet dédié quelque peu artificiel. Une des premières propositions en la matière fut l'alphabet Unistroke [GOL 93] repris ensuite dans EdgeWrite [WOB 03] (figure 10.4b). L'alphabet le plus utilisé à l'heure actuelle est certainement Graffiti2 que l'on trouve sur les PDA utilisant le système d'exploitation PalmOS (figure 10.4a).

Suivant le système considéré, la reconnaissance des caractères peut se limiter à une reconnaissance des formes ou intégrer également un modèle de langage qui guide les traitements d'images. Cette modélisation est équivalente à celle mise en œuvre dans le cadre de la prédiction de mots.

L'écriture manuscrite sur écran tactile relève du geste bidimensionnel (2D). De nombreux dispositifs mobiles intègrent désormais des accéléromètres qui permettent de passer facilement au geste 3D. Ce type de dispositif, qui a été popularisé par la console de jeux Wii de Nintendo, se retrouve désormais dans de nombreux téléphones intelligents.

L'utilisation des données tridimensionnelles comme support à l'entrée d'information reste limitée sur les *Smartphones* : elle est cantonnée au parcours de menus déroulants ou à la simulation d'un clic souris. On se retrouve alors dans une problématique proche de l'aide à la communication pour les personnes handicapées.

### **10.2.2. Dispositifs d'entrée de texte pour systèmes d'aide à la communication**

Le recours à un clavier virtuel est systématique dans le cadre de l'aide à la communication pour personnes handicapées. Les capacités motrices de l'utilisateur peuvent lui permettre de contrôler le système informatique à l'aide d'un pointeur. C'est par exemple le cas des personnes myopathes utilisant un joystick microgravité.

Cette situation d'usage est proche de celle des téléphones à clavier virtuel. On notera toutefois que la disposition des touches est optimisée afin de minimiser, pour une langue donnée, les déplacements du pointeur souris.

Cette optimisation ne concerne seulement que les personnes handicapées, comme le montre la figure 10.5 où le clavier optimisé FITALY [MAC 99] a été implémenté sur un organisateur personnel PalmV afin de limiter les déplacements du stylet.





**Figure 10.5.** Clavier optimisé FITALY sur un organiseur personnel Palm V

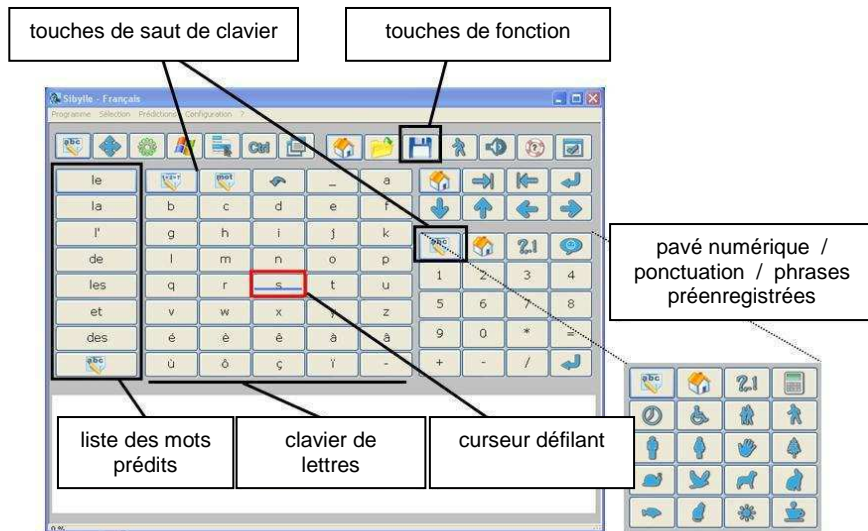
Dans d'autres pathologies (infirmités motrices cérébrales, scléroses latérales amyotrophiques, syndrome d'enfermement, etc.), les capacités de contrôle moteur de l'utilisateur se limitent à la réalisation d'un clic à l'aide d'un contacteur. L'entrée de texte repose alors sur un clavier à défilement automatique : un curseur se déplace touche après touche sur le clavier virtuel affiché à l'écran. L'intervention de la personne handicapée se limite à la désignation des symboles lorsque le curseur est sur la touche désirée. La sélection est réalisée à l'aide d'un dispositif (téléthèse) qui dépend des capacités motrices de l'utilisateur : détecteur de souffle, commande oculaire, contacteur menton ou toute autre partie du corps, etc. (figure 10.6).



**Figure 10.6.** Deux exemples de dispositifs d'entrée pour personnes handicapées, a) détecteur de souffle, b) commande oculaire et au menton

Il existe à l'heure actuelle une grande diversité de systèmes de suppléance, tant du point de vue de l'organisation du clavier virtuel que des modes de défilement autorisés ([ANT 07] pour une revue d'ensemble). Dans tous les cas, ces systèmes restent confrontés au problème de la lenteur de la composition des messages. La tâche de saisie est généralement longue (1 à 5 mots par minute en moyenne) et extrêmement fatigante pour les utilisateurs.

Pour accélérer la composition des messages, deux approches complémentaires sont envisageables. Tout d'abord, on peut accélérer la sélection sur le clavier virtuel en limitant le nombre de défilements du curseur pour accéder à la touche désirée. La plupart des systèmes commerciaux utilisent ainsi un clavier statique associé à un mode de balayage de type ligne/colonne : au lieu d'un déplacement linéaire touche par touche, le balayage se fait ici par ligne de clavier. Une fois sélectionnée la ligne où se trouve le symbole recherché, le curseur balaie cette dernière jusqu'à atteindre la touche correspondante. Le gain de temps ainsi obtenu est appréciable, mais a pour contrepartie de requérir que l'utilisateur fasse deux appuis par caractère saisi. A l'opposé, d'autres systèmes utilisent un module de prédiction de lettres qui permet d'optimiser contextuellement la disposition des touches sur le clavier après chaque saisie. En mode de défilement linéaire, cette réorganisation dynamique ne pose pas de problème particulier. Il s'agit là d'une application particulière de la prédiction linguistique qui sera étudiée au paragraphe 4.3.



**Figure 10.7.** Clavier virtuel du système d'aide à la communication SIBYLLE

La seconde piste d'optimisation est complémentaire de la première. Elle consiste à compléter semi automatiquement la saisie afin d'éviter à l'utilisateur l'entrée de certains caractères. On retrouve une idée rencontrée avec les systèmes informatiques mobiles : prédire les mots (ou fins de mots) à venir en fonction de ceux déjà saisis.

La complétion peut être pilotée de deux manières. Dans le système VITIPI [BOI 01], le système ne propose une complétion que lorsqu'il ne reste qu'une seule hypothèse compatible avec les caractères saisis : un appui spécifique permet alors de valider la proposition du système.

Dans d'autres systèmes, tels FASTY [BEC 04] ou SIBYLLE [WAN 07], une liste de mots prédits est affichée à côté du clavier virtuel (figure 10.7). L'utilisateur peut parcourir cette liste par défilement linéaire afin de choisir le mot adéquat. En proposant plusieurs hypothèses (5 à 10 en général), la complétion automatique est plus souvent utile.

Cette stratégie nécessite en revanche un appui supplémentaire afin d'accéder à la liste de prédiction. Suivant le tableau clinique de l'utilisateur, une des deux approches sera plus appropriée. Dans les deux cas, le modèle de prédiction de mots mis en jeu est identique.

### **10.2.3. Evaluer l'intérêt de la prédiction de mots pour la saisie de message**

En résumé, la mise en œuvre d'une entrée de texte efficace sur interface limitée passe par l'intégration d'un moteur de prédiction de lettres ou de mots. Plusieurs métriques peuvent évaluer l'efficacité d'un dispositif d'entrée de texte [SAD 09]. Tout d'abord, la vitesse de saisie peut être estimée en nombre de mots saisis par minute (*WPM : Words Per Minute*).

Cette métrique est évaluée en termes de caractères saisis par minute, ramenée ensuite à une saisie de mots d'une longueur moyenne de 5 caractères. En l'absence de *benchmark* de référence, et du fait de la variabilité des sujets, il est difficile de comparer les méthodes d'entrée de texte à l'aune de cette mesure. Le tableau 10.1 donne, à titre d'illustration, quelques résultats expérimentaux.

On observe que la saisie sur clavier standard virtuel (donc de taille réduite) reste rapide, et que la vitesse de saisie semble inversement proportionnelle au nombre de touches du clavier. L'expérience menée dans [MAC 01] sur un clavier téléphonique ambigu (12 touches) est également instructive. Elle montre qu'une désambiguïsation logicielle est préférable à une désambiguïsation explicite de l'utilisateur suivant le paradigme multi-appuis.

Méthode	Système	WPM
Clavier virtuel standard	QWERTY	40 (sans apprentissage)
Clavier virtuel standard optimisé	OPTI	42 (8h de pratique)
Clavier à accords	<i>Twiddler</i>	26 (6h de pratique)
Clavier virtuel ambigu avec prédiction	<i>LetterWise</i> (12 touches)	21 WPM (6 h de pratique)
Clavier virtuel ambigu sans prédiction (multi-appuis)	<i>Multi-Tap</i> (12 touches)	15.5 WPM (6 h de pratique)
Clavier virtuel ambigu avec prédiction	<i>TouchMe4Key</i> (4 touches)	12 WPM (5 h de pratique)
Système d'aide à la communication	HandiAS	5 WPM (utilisateur handicapé)

**Tableau 10.1.** *Quelques exemples de vitesses de saisies observées sur des dispositifs limités (d'après [SAD 09], [MAC 01] et [LEP 97])*

Le nombre d'appuis par caractères saisis est donc un critère limitant important en informatique mobile et ubiquitaire. Il se retrouve dans l'aide au handicap, où chaque appui influe sur la fatigue physique et psychique de l'utilisateur. Il est donc intéressant de coupler le WPM avec la mesure du nombre moyens d'appuis par caractère saisi (KSPC = *KeyStroke Per Character*). Si l'on considère un texte de  $N$  caractères, le KSPC se définit formellement comme le quotient du nombre d'appuis nécessaire pour saisir le texte par ce nombre  $N$ . Cette métrique permet d'étudier l'influence de la prédiction linguistique lorsqu'elle est mise en jeu, puisque celle-ci a pour conséquence d'éviter les appuis nécessaires à la saisie des caractères prédits.

Le WPM et le KSPC ne permettent toutefois pas de distinguer l'influence de la prédiction linguistique de l'interface homme-machine dans laquelle elle est intégrée. Du point de vue de l'ingénierie linguistique, on préfère donc mesurer l'efficacité de la prédiction à l'aide d'une troisième métrique : le taux d'économie de saisie (*KSR = KeyStroke Saving Rate*). Formellement, le KSR est défini comme suit :

$$KSR_n = (1 - k_p / k_a) \cdot 100 \quad [10.1]$$

où  $k_p$  est le nombre d'appuis effectivement réalisés par l'utilisateur lors de la saisie d'un message,  $k_a$  le nombre d'appuis qui auraient été nécessaires sans aide de la prédiction. L'indice  $n$  adjoint au KSR correspond au nombre d'hypothèses lexicales présentées à l'utilisateur (typiquement,  $N = 5$ ).

### 10.3. Prédiction de lettres et claviers virtuels dynamiques

Comme nous l'avons vu, la prédiction de lettres n'intervient que pour les claviers virtuels à réorganisation dynamique : la disposition des touches est revue après chaque saisie de caractère afin d'accélérer la prochaine entrée. La figure 10.8 illustre ce principe sur un clavier dynamique à balayage linéaire : on présente à chaque fois en premier les lettres qui ont la plus grande probabilité d'être saisies en fonction des caractères déjà saisis. Sur cet exemple, le mot à saisir est *com...pter*. On observe que la seconde et la troisième lettre du mot sont atteintes avec respectivement 0 et 1 défilement du curseur de balayage.

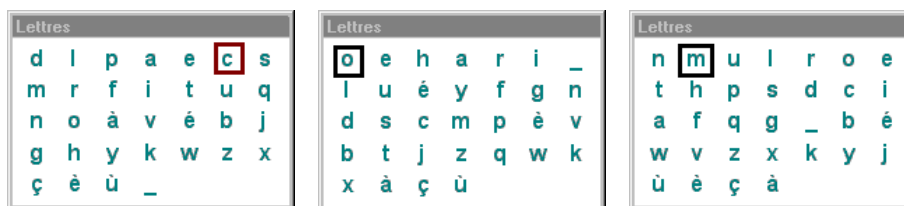


Figure 10.8. Réorganisation après chaque saisie d'un clavier virtuel dynamique

Sur nombre de claviers dynamiques, cette réorganisation repose sur l'utilisation d'un lexique fréquentiel organisé sous la forme d'un graphe probabiliste de lettres [MEN 01]. Cette méthode a pour inconvénient d'être mise en défaut sur les mots hors vocabulaire ou après toute erreur de saisie. Dans le système SIBYLLE, nous avons choisi au contraire de construire un modèle markovien de succession des lettres dans la langue [SCH 02]. Plus précisément, nous utilisons un penta-gram, c'est-à-dire que l'on estime la probabilité d'occurrence des lettres en fonction des quatre derniers caractères saisis :  $P(c_i | (c_{i-1} c_{i-2} c_{i-3} c_{i-4}))$ . Cette approche présente plusieurs intérêts.

Tout d'abord, la fenêtre de contexte du penta-gram (4 mots) réalise une modélisation qui est plus syllabique que lexicale. Elle est donc plus robuste en cas d'erreur de saisie ou d'orthographe et n'est pas concernée par le problème des mots hors vocabulaire. En ne tenant pas compte des frontières lexicales, cette méthode est par ailleurs bien plus efficace en début de mot.

Linéaire Azerty	Ligne/Colonne Azerty	Ligne/Colonne optimisé	Sibylle
33,0	9,0	4,3	2,9

**Tableau 10.2.** Nombre moyen de défilements nécessaires pour la saisie d'un caractère sur un clavier accentué à 64 caractères. Test réalisé sur des extrait du corpus Le Monde

Le modèle penta-gram que nous avons développé est assez sommaire : il ne met en œuvre aucun lissage et aucune technique d'élagage (*pruning*). L'efficacité est pourtant au rendez-vous, comme le montre les résultats expérimentaux résumés dans le tableau 10.2. On y observe que le nombre moyen de défilements pour atteindre un caractère est inférieur à trois, ce qui est très satisfaisant à l'usage. Un clavier statique à balayage ligne/colonne nécessite de 4,3 à 9,0 défilements suivant l'organisation des touches, tout en requérant deux appuis au lieu d'un seul pour atteindre le même but.

#### 10.4. Prédiction de mots : retour historique et modèles markoviens de base

Nous venons d'étudier une première illustration de l'intérêt d'une prédiction linguistique pour l'entrée de texte. Le champ d'application de la prédiction de mots est toutefois bien étendu, puisqu'il concerne la plupart des dispositifs de saisie de texte sur interfaces limitées.

La première idée qui vient à l'esprit en matière de prédiction de mots est l'utilisation d'un lexique. Si le lexique est fréquentiel, on peut à tout moment proposer les  $N$  mots les plus probables en fonction des caractères déjà saisis. Des techniques d'optimisation classiques permettent une exploration rapide du dictionnaire même sur des dispositifs ayant une faible puissance de calcul. C'est pourquoi ce type de prédiction se retrouve sur la plupart des téléphones intelligents et des systèmes commerciaux d'aide à la communication. Cette approche revêt néanmoins les mêmes faiblesses que celles entrevues avec la prédiction de lettres : le système n'est d'aucune aide en présence de mots hors vocabulaire, et la non prise en compte du contexte (mots précédents) limite l'efficacité de la prédiction.

Les premiers systèmes qui ont considéré un contexte de prédiction plus étendu relevaient de l'aide au handicap. Ils mettaient en jeu une approche descendante : l'idée était de partir d'un scénario de communication et de guider l'utilisateur dans la composition des phrases [GUE 92, COP 97]. L'utilisateur compose pas à pas le message en remplissant (mots ou expressions) des schémas de phrases prédéfinis

produits par le système en fonction du thème de discussion choisi. Cette composition assistée permet une saisie très rapide, mais elle se limite à une communication finalisée, orientée vers un but précis et où le libre arbitre de l'utilisateur est le plus souvent restreint. En ce sens, elle ne peut tenir lieu de solution universelle pour l'entrée de texte sur interface limitées.

Dès les années 1990, certains auteurs ont donc cherché à mettre en œuvre des systèmes ascendants de prédiction. Suivant le paradigme dominant à l'époque, leurs réalisations reposaient sur une modélisation syntaxique à base de connaissances. [VAN 92 ; MAU 00]. Ces approches se sont toutefois heurtées au problème de l'ordonnement des hypothèses (ce qui a parfois conduit à l'utilisation de grammaires probabilistes [LEP 98, WRI 94]) et surtout au problème du manque de couverture et de robustesse de la prédiction sur des saisies réelles.

C'est pourquoi la prédiction de mots n'est plus envisagée à l'heure actuelle que sous la forme d'une approche ascendante à l'aide de modèles génératifs (modèles de Markov). A la base, le module de prédiction de notre système SIBYLLE repose ainsi sur un quadri-gram qui estime la probabilité d'occurrence des mots à venir en fonction des trois derniers mots saisis :  $P(w_i | (w_{i-1} w_{i-2} w_{i-3}))$ . Afin de faire face au problème de l'éparpillement des données, d'autres systèmes combinent une N-gram avec un modèle de classe. Par exemple, *Profet* [CAR 97] et *Fasty* [TRO 05] reposent sur l'interpolation linéaire d'un bigram de mots et d'un modèle tri-classe :

$$P(w_i | h_i) \approx \lambda_i \cdot P(w_i | (w_{i-1})) + (1-\lambda_i) \cdot P(w_i | (c_i)) \cdot P(c_i | (c_{i-1} c_{i-2})) \quad [10.2]$$

où  $h_i$  représente le contexte de saisie antérieur au mot à prédire  $w_i$  et  $P(w_i | (w_{i-1}))$  représente l'estimation de la probabilité d'occurrence de ce mot en fonction du mot précédent (bigram).  $P(w_i | (c_i))$ .  $P(c_i | (c_{i-1} c_{i-2}))$  correspond à la probabilité d'occurrence du mot  $w_i$  en fonction des catégories grammaticales  $c_{i-1}$  et  $c_{i-2}$  des deux mots précédents. Celle-ci est calculée à partir de la probabilité  $P(w_i | (c_i))$  d'appartenance du mot  $w_i$  à la catégorie  $c_i$  et de la probabilité  $P(c_i | (c_{i-1} c_{i-2}))$  que la catégorie  $c_i$  suive une séquence de mots de catégories  $c_{i-1}$  et  $c_{i-2}$  (modèle triclasse). Enfin,  $\lambda_i$  est un coefficient d'interpolation qui est estimé à l'aide de l'algorithme EM [JEL 90]. L'augmentation de la puissance des ordinateurs et celle des données d'apprentissage limite désormais l'intérêt d'un recours aux modèles de classes.

Les limitations des N-grammaires en termes de modélisation des dépendances à longue distance sont bien connues. Certains auteurs ont cherché à utiliser des modèles markoviens prenant en compte une fenêtre de contexte plus large à nombre de paramètres égaux. La première version du système SIBYLLE [SCH 04] utilisait

par exemple un modèle structural inspiré de [CHE 00] et basé sur la segmentation de l'énoncé en segments minimaux non récursifs (modèle *N-chunks*). La plateforme *PCA* [BLA 07] utilise le modèle des patrons [RON 06], une sous-classe des modèles de Markov cachés qui considère comme états des séquences de longueurs variables et non plus des mots uniques comme dans le cas des N-grams. Malheureusement, on ne dispose pas à l'heure actuelle pour ces modèles de techniques d'optimisation équivalentes à celles des modèles markoviens standards, qui constituent donc la base des moteurs de prédiction actuels.

A titre d'exemple, le modèle quadri-gram qui sert de base au prédicteur SIBYLLE atteint des performances déjà intéressantes. Construit sur le *toolkit* SRILM [STO 02] avec un vocabulaire contrôlé, il utilise un lissage de *Kneser-Ney modifié* [G00 01] et la technique de *pruning* proposée par [STO 98]. Après apprentissage sur corpus journalistique (*Le Monde*), nous avons testé ce moteur de prédiction sur différents corpus correspondant à autant de registres de langue (tableau 10.3).

Registre/Genre	Description	Nb. mots
<i>Journalistique</i>	Journal <i>L'Humanité</i> (extrait : janvier 1999)	58 457
<i>Scientifique</i>	Articles scientifiques (références comprises)	8 766
<i>Prose littéraire</i>	Extrait de <i>Germinal</i> d'Emile Zola	50 251
<i>Parole</i>	Corpus OTG [ANT 02]: transcription dialogue oral finalisé (extrait : énoncé de l'hotesse d'accueil)	15 435
<i>Courriel</i>	Corps de courriels personnels de l'auteurs.	44 946

**Tableau 10.3.** *Corpus de test des versions française et allemande du système Sibylle*

Le tableau 10.4 résume les résultats expérimentaux obtenus. On observe que le KSR5 atteint déjà 57,8 % sur un registre identique à celui du corpus d'apprentissage. Ce qui signifie que dans cette situation idéale, le système peut éviter plus d'une saisie sur deux à l'utilisateur. Ce résultat est en lui-même remarquable. Il s'approche en effet de la valeur de 60 % que certains travaux théoriques sur la langue naturelle ont posé comme valeur limite indépassable par un système de prédiction [COP 97].



On peut donc supposer qu'il reste intrinsèquement peu d'espace d'amélioration pour un autre modèle servant de base à la prédiction 1.

Journalistique	Scientifique	Prose	Parole	Courriel
57,8 %	44,2%	46,2%	48,3%	50,1%

**Tableau 10.4.** *Mesure du KSR<sub>5</sub> du moteur de prédiction quadrigram du système Sibylle*

On observe en revanche une dégradation sensible de l'aide réellement apportée sur les autres registres. Deux facteurs principaux peuvent être invoqués pour expliquer cette baisse de performances : tout d'abord la présence de mots hors vocabulaire, particulièrement sensible dans le cas du corpus scientifique. Et ensuite l'usage d'un style de langage différent de celui du corpus d'apprentissage, qui est particulièrement visible dans le cas de la parole conversationnelle. Il semble donc important de s'intéresser avant tout à l'adaptation des modèles de prédiction.

## 10.5. Modèles adaptatifs pour la prédiction de mots

### 10.5.1. Adaptation utilisateur

L'adaptation à l'utilisateur est une problématique bien connue en modélisation stochastique du langage. Elle consiste à adapter un modèle initial, appris sur un corpus de langue générale, au style de langage et au vocabulaire spécifique à l'utilisateur. En pratique, cela consiste à apprendre un modèle spécifique sur les textes déjà saisis par l'utilisateur. Bien entendu, ce modèle utilisateur ne peut suffire à lui seul, car les messages saisis par l'utilisateur sont en nombre insuffisant pour couvrir la langue considérée. Le modèle général est donc également conservé pour couvrir les « trous » de couverture du modèle utilisateur. On construit donc un modèle global par interpolation linéaire du modèle général et du modèle utilisateur :

$$P_{global}(w_i) = \lambda_u \cdot P_{utilisateur}(w_i) + (1 - \lambda_u) \cdot P_{général}(w_i) \quad [10.3]$$

---

1. La version française du système Fasty n'atteint que 48 % de KSR<sub>5</sub> [TRO 05] mais 54 % sur l'allemand. Ce système n'utilise toutefois qu'un bigram de mots et un tri-classe. A titre de comparaison, un modèle trigram entraîné sur les données d'apprentissage de SIBYLLE obtient un KSR<sub>5</sub> de 55,8 % et un bigram 51,2 %.

Pour évaluer l'intérêt de cette adaptation, nous avons à nouveau testé le modèle sur les corpus précédents. Nous avons repris les mêmes données d'apprentissage pour l'entraînement du modèle général. Par ailleurs, chaque corpus de test précédent a été partagé en un sous-corpus de développement servant à l'apprentissage du modèle utilisateur et un sous-corpus de test proprement dit. Le tableau 10.5 donne le  $KSR_5$  du modèle adaptatif en rappelant celui du quadri-gram de base.

Modèle	Journal	Sciences	Prose	Parole	Courriel
<i>4-gramme seul</i>	57,9 %	44,2 %	46,0 %	48,3 %	48,6 %
<i>4-gramme+MU</i>	58,5 %	52,4 %	50,6 %	57,7 %	53,0 %
<i>Gain <math>KSR_5</math></i>	+ 0,6 %	+ 8,2 %	+ 4,6 %	+ 9,4 %	+ 4,4 %

**Tableau 10.5.**  $KSR_5$  de la prédiction avec modèle utilisateur (MU)

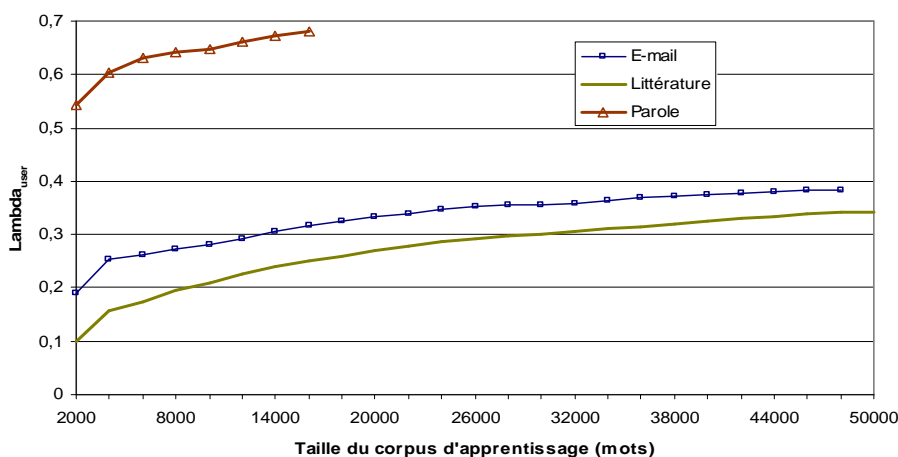
Avec adaptation, le  $KSR_5$  est cette fois supérieur à 50 % sur tous les corpus. L'adaptation à l'utilisateur semble d'autant plus bénéfique que le  $KSR$  initial est faible. Aussi nous sommes nous demandés où cette adaptation intervenait : conduit-elle à une amélioration générale des performances, ou opère-t-elle plus fortement sur les messages où le  $KSR$  est faible ? Pour cela, nous avons calculé la variance du taux d'économie de saisie entre les différents textes de deux corpus (tableau 10.6). Les résultats montrent que le modèle utilisateur limite la dispersion des performances : le gain en  $KSR$  est donc modéré pour les messages déjà bien prédits et supérieur pour les autres. C'est sans doute cette capacité à améliorer les situations où la prédiction se comporte mal qui fait que le modèle adaptatif est apprécié des utilisateurs.

Modèle	Prose	Courriel
<i>4-gramme seul</i>	$\sigma = 1,44\%$	$\sigma = 8,38\%$
<i>4-gramme+MU</i>	$\sigma = 1,28\%$	$\sigma = 5,05\%$

**Tableau 10.6.** Analyse distributionnelle du  $KSR_5$  sur deux corpus de test : variance entre textes

Pour que cette adaptation soit perceptible, il faut toutefois que le modèle utilisateur influe rapidement sur le comportement du modèle global. Aussi avons-nous cherché à estimer la masse de données utilisateur nécessaire à atteindre une

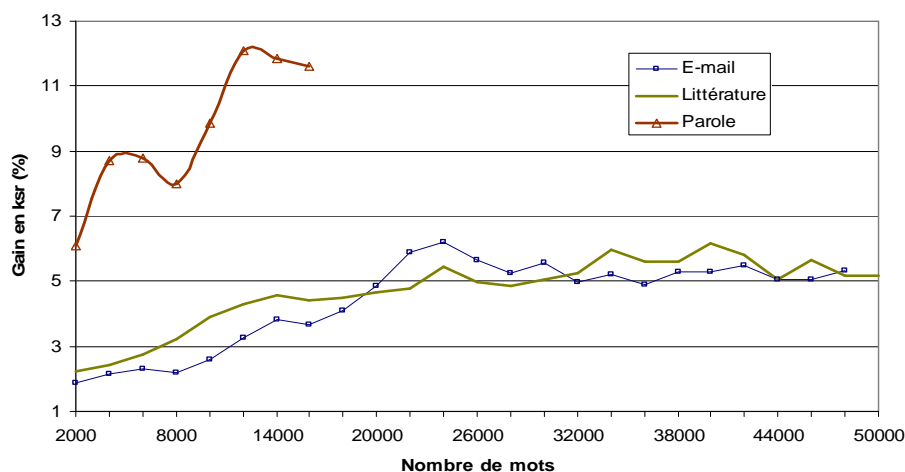
adaptation significative. Nous avons procédé à une évaluation incrémentale suivant le principe suivant : le corpus de test est divisé en  $n$  sous-corpus. On mesure le  $KSR$  du système sur le premier sous-corpus. Celui-ci est alors utilisé comme donnée d'apprentissage par le modèle utilisateur et pour l'adaptation du coefficient d'interpolation. Le modèle adapté qui en résulte est ensuite évalué sur le second sous-corpus et ainsi de suite. Nous simulons ainsi l'évolution du modèle au fil des saisies utilisateurs. Cette étude a été menée sur les corpus *courriel*, *prose* et *parole*.



**Figure 10.9.** Adaptation du module de prédiction de mots: évolution du coefficient d'interpolation du modèle utilisateur

La figure 10.9 montre l'évolution du facteur d'interpolation au cours de l'apprentissage. L'adaptation est manifeste : au bout de 10 000 mots saisis, la part prise par le modèle utilisateur dans l'estimation des probabilités dépasse déjà 20 % dans tous les cas. Le modèle utilisateur prend même une part majoritaire avec le corpus de dialogue oral ( $\lambda_{\text{utilisateur}} \approx 0,7$ ). Cette extrême rapidité d'adaptation s'explique par la spécificité syntaxique de la parole conversationnelle et par le caractère très finalisé de l'interaction (tâche de renseignement touristique). À l'opposé, le modèle de langue général reste prédominant dans les deux autres cas ( $\lambda_{\text{utilisateur}} \approx 0,3$ ).

La figure 10.10 donne l'évolution du gain en KSR entre le modèle adaptatif et le quadri-gram de base. On remarque que l'influence du modèle utilisateur est toujours positive et qu'elle est rapidement significative : on observe déjà un gain d'au moins 2 % avec seulement 2 000 mots de données d'adaptation. Ce gain atteint ensuite au moins 5 % au bout de 25 000 mots d'adaptation.



**Figure 10.10.** Adaptation du module de prédiction de mots par intégration d'un modèle utilisateur : évolution du gain en  $KSR_5$  par rapport à un quadrigram

### 10.5.2 Adaptation au contexte courant du discours : analyse sémantique latente.

En dépit de sa rapidité de mise en action (2 000 mots d'apprentissage), le modèle utilisateur ne permet qu'une adaptation à moyen terme. Une adaptation à court terme peut également être intéressante. En effet, lorsque la communication se focalise sur un thème précis, les mots qui appartiennent au champ sémantique du discours ont plus de chance d'apparaître.

Dans le cadre de la modélisation du langage, plusieurs stratégies d'adaptation à court terme ont été explorées. Une première approche consiste à détecter automatiquement le thème du discours [BIG 01] et à focaliser ensuite la prédiction sur ce dernier. Pour cela, on aura au préalable construit des modèles de prédiction spécifiques à chaque thème. Largement étudiée au tournant du millénaire, cette idée n'a pas tenu toutes ses promesses. Tout d'abord, la robustesse des systèmes de détection de thème est restée perfectible. Par ailleurs, ces approches ne peuvent considérer que des thèmes très généraux qui rendent leur application à la prédiction de mots assez peu intéressante. Aussi l'on ne s'étonnera pas que les travaux fondateurs n'aient pas vraiment connu de suite, du moins dans le domaine de la prédiction.

Une autre approche moins ambitieuse en apparence consiste à favoriser les mots qui ont été saisis récemment, en supposant qu'ils ont une probabilité plus forte de réapparaître puisque faisant partie du champ lexical du thème courant du discours.

Cette hypothèse est particulièrement pertinente lorsque la communication est très finalisée. Elle est à la base du modèle *cache* [KUH 90] qui remonte (éventuellement avec une pondération pilotée par un facteur d'oubli) la probabilité d'occurrence des mots apparus dans une certaine fenêtre de contexte (typiquement, les 100 derniers mots saisis).

Dans le même esprit, le modèle *trigger* [MAT 03] est un peu plus fin. Il fonde l'adaptation non pas sur des mots isolés mais sur des collocations : un mot déclencheur augmente (dès qu'il est utilisé) la probabilité des mots avec lesquels il est déjà apparu dans une certaine fenêtre de contexte. S'il a été montré que ces modèles ont une certaine capacité d'adaptation, les gains rapportés par la littérature restent assez limités, en particulier dans le domaine de la prédiction de mots. Comme nous le verrons plus loin, nos résultats expérimentaux confirment ce point.

Aussi avons-nous proposé un modèle d'adaptation thématique original qui est adapté d'une technique issue de la recherche d'information : l'analyse sémantique latente (LSA pour *Latent Semantic Analysis*). La LSA [DEE 1990] peut être vue comme une technique qui rend compte de relations lexicales sémantiques à partir de la distribution des cooccurrences de mots dans un contexte prédéfini. Elle peut ainsi prédire des mots lexicaux à partir de termes qui leur sont reliés sémantiquement.

Dans un premier temps, un corpus de textes est représenté sous la forme d'une matrice (terme  $\times$  contexte) où quelques milliers de termes d'indexation ont été préalablement choisis et où le contexte peut être aussi bien un document, un paragraphe, ou toute autre fenêtre contextuelle. Chaque élément de la matrice comptabilise la fréquence normalisée (*td.idf*, par exemple) d'occurrence d'un terme d'indexation dans le contexte considéré.

Par rapport au modèle vectoriel standard [SAL 83], l'apport de la LSA consiste à transposer la matrice obtenue dans un espace de dimension réduite. Pour cela, on décompose la matrice en valeurs singulières et on ne conserve que les  $k$  principales valeurs singulières obtenues (typiquement,  $k$  est compris entre 100 et 300). Chaque mot est représenté par un vecteur de  $k$  dimensions dans ce nouvel espace. Il est alors possible de comparer ces vecteurs par une mesure de distance ordinaire, le cosinus de l'angle entre les vecteurs étant le plus souvent utilisé. On montre que cette distance constitue une bonne mesure de similarité sémantique [WAN 05].

Dans le cadre de la prédiction de mots, l'objectif est d'estimer la probabilité d'occurrence d'un mot en fonction du thème du discours. Dans l'espace LSA, la somme de plusieurs vecteurs représente la sémantique globale des éléments considérés.

Nous faisons l'hypothèse que le champ sémantique correspondant au thème courant peut être décrit par un vecteur de contexte correspondant aux derniers mots déjà saisis :

$$\bar{h} = \sum_{i=1}^m \bar{w}_i \quad [10.4]$$

où ( $w_1, \dots, w_m$ ) sont les mots du contexte courant. On estime alors la (pseudo-) probabilité sémantique d'occurrence d'un mot à partir du cosinus de l'angle que fait sa représentation vectorielle avec le vecteur contexte :

$$P_{LSA}(w_i|h) = \frac{(\cos(w_i, \bar{h}) - \cos_{\min}(\bar{h}))^\gamma}{\sum_k (\cos(w_k, \bar{h}) - \cos_{\min}(\bar{h}))^\gamma} \quad [10.5]$$

On observe que la distribution de ces pseudo-probabilités est très plate. Suivant [COC 98], nous leur appliquons un facteur de température  $\gamma$  pour augmenter le contraste entre les différentes prédictions.

Le modèle sémantique résultant est ainsi supposé prédire des mots sémantiquement cohérents avec le contexte courant. Considérons l'énoncé suivant :

*Mon père était professeur en mathématiques et je pense que...* [a]

Le tableau 10.7 donne la liste des dix mots qui, selon le modèle, sont les plus appropriés pour poursuivre la saisie. On observe que les propositions relèvent bien toutes de thématiques (*famille, enseignement, sciences*) initiées dans le discours.

Rang	Mot	$P_{LSA}$	Rang	Mot	$P_{LSA}$
1	<i>Professeur</i>	0,0117	6	<i>père</i>	0,0046
2	<i>mathématiques</i>	0,0109	7	<i>mathématiques</i>	0,0045
3	<i>Enseigné</i>	0,0083	8	<i>grand-père</i>	0,0043
4	<i>Enseignait</i>	0,0053	9	<i>sciences</i>	0,0036
5	<i>mathématicien</i>	0,0049	10	<i>enseignant</i>	0,0032

**Tableau 10.7.** Mots de probabilités maximales selon le modèle LSA pour l'énoncé (a)

Si les mots prédits sont sémantiquement cohérents avec le contexte, tous ne peuvent représenter une poursuite acceptable du message d'un point de vue syntaxique. On note également que la LSA ne peut prédire l'apparition de mots grammaticaux, vides de sens. Il faut donc coupler cette prédiction sémantique avec un modèle de langage classique. Dans Sibylle, nous réalisons ainsi une interpolation géométrique du modèle sémantique et du modèle n-gramme adaptatif :

$$P_{global}(w_i) = \frac{P_{base}(w_i)^{\lambda_i} \cdot P_{LSA}(w_i)^{(1-\lambda_i)}}{\sum_{j=1}^n P_{base}(w_j)^{\lambda_i} \cdot P_{LSA}(w_j)^{(1-\lambda_i)}} \quad [10.6]$$

avec  $n$ , nombre de termes du vocabulaire. Par rapport à une interpolation linéaire, l'interpolation géométrique présente l'intérêt d'imposer que les mots prédits soient cohérents d'un point de vue syntaxique et sémantique avec le début du message saisi. [WAN 07] a montré la supériorité de cette interpolation dans une tâche de prédiction (gain en *KSR* de 0,6 %).

Nos études ont en outre montré que l'efficacité de la prédiction sémantique dépend fortement du mot considéré [WAN 05]. En particulier, il existe une certaine corrélation entre le nombre de termes sémantiquement reliés à un mot et la densité de ses plus proches voisins dans l'espace LSA. Cette observation est assez intuitive : plus les voisins d'un terme sont proches, plus on a de chances que ceux-ci lui soient réellement liés sémantiquement. Afin de rendre compte de cette influence de la densité, nous avons décidé d'adapter localement le facteur d'interpolation de l'équation [10.7] ; celui-ci sera défini comme suit pour chaque mot :

$$\lambda_i = \beta \cdot D(w_i) \text{ si } D(w_i) > 0 \text{ et } \lambda_i = 0 \text{ sinon} \quad [10.7]$$

avec  $D(w_i)$  la densité des voisins d'un terme  $w_i$  :

$$D_m(w_i) = \frac{1}{m} \cdot \sum_{j=1}^m \cos(\vec{w}_i, NN_j(\vec{w}_i)) \quad [10.8]$$

où les  $NN_j(\vec{w}_i)$  sont les vecteurs  $w_j$  les plus proches de  $w_i$ .  $\beta$  est une constante de pondération qui gouverne l'importance de l'influence de la LSA. Elle a été fixée à 0,4 après étude expérimentale. Cela signifie que l'influence du modèle LSA sur le prédicteur global est comprise entre 0 et 40 % suivant le mot considéré. Pour des raisons pratiques de coût computationnel, la valeur de cette constante n'est pas

modifiée au cours de l'apprentissage. L'influence de sa valeur est par ailleurs lissée par le facteur de densité.

Le module sémantique a été entraîné sur sept années du journal *Le Monde* (1996-2002, 101 millions de mots) à l'aide du *toolkit* InfoMap2. Nous avons utilisé 3 000 termes d'indexation et un vocabulaire contrôlé de 80 000 mots lexicaux. Le contexte de cooccurrence qui a été retenu a été une fenêtre de  $\pm 100$  mots. Enfin, nos études nous ont conduit à une valeur de  $k = 150$  dimensions pour la réduction par décomposition en valeurs singulières.

Modèle	Journal	Sciences	Prose	Parole	Courriel
Base	57,9 %	44,2 %	46,0 %	48,3 %	48,6 %
Base + MU	<b>58,5 %</b>	52,4 %	50,6 %	57,7 %	53,0 %
Base + LSA	<b>58,9 %</b>	45,6 %	47,7 %	49,9 %	50,2 %
SIBYSEM	59,4 %	52,9 %	52,0 %	58,8 %	53,7 %
<i>Gain</i>	+ 1,5 %	+ 8,7 %	+ 6,0 %	+10,5 %	+ 5,1 %

**Tableau 10.8.** Performances comparées ( $KSR_5$ ) des différents modèles de prédiction étudiés

L'apport de la focalisation sémantique a été évalué sur les corpus précédents. Le tableau 10.8 compare les résultats obtenus par les modèles suivants :

- modèle 4-gramme seul (base) ;
- modèle 4-gramme + modèle utilisateur (rappel de résultats déjà présentés) ;
- modèle 4-gramme + LSA mais sans modèle utilisateur ;
- Sibylle : modèle 4-gramme + modèle utilisateur + LSA avec pondération.

Globalement, la LSA permet de nouveaux gains dans toutes les situations : son apport est donc complémentaire de celui du modèle utilisateur. Les gains cumulés permettent d'atteindre des valeurs de  $KSR$  plus que satisfaisantes (entre 52 % et 59 % suivant le registre de langue) qui s'approchent des seuils maximum théoriques d'économie de saisie. Les améliorations permises par la LSA à proprement parler peuvent parfois sembler limitées. Elles sont toutefois statistiquement significatives (seuil de significativité  $p < 0,001$ ). Pour prendre la mesure de ces résultats, il faut comprendre qu'un  $KSR_5$  de base de plus de 50 % constitue déjà une excellente

2. Infomap Project: <http://infomap-nlp.sourceforge.net>



performance. On remarquera à ce sujet que les autres méthodes d'adaptation contextuelle que nous avons implémentées (modèle cache ou *trigger*) donnent des résultats nettement inférieurs [WAN 07]. A titre d'exemple, l'utilisation d'un modèle cache avec une fonction d'oubli exponentielle [CLA 97] conduit au mieux à une dégradation de 1 % de KSR par rapport au modèle LSA pondéré interpolé géométriquement.

### 10.6. Conclusion : pour une meilleure intégration ergonomique de la prédiction

En conclusion, ce chapitre a cherché à montrer l'importance des processus de prédiction linguistique pour l'aide à la saisie de texte sur interfaces limitées (systèmes informatiques mobiles mais également systèmes d'aide à la communication pour personnes handicapées). Nous avons montré que les moteurs de prédiction sont à même d'offrir à l'utilisateur une accélération appréciable de sa vitesse de saisie. Lorsqu'elle est couplée à des techniques avancées d'adaptation à court et à long terme, la prédiction atteint un niveau de performance qui s'approche des limites théoriques du domaine. Aussi peut-on raisonnablement penser qu'on n'assistera plus, dans les années à venir, à des gains significatifs de performance. Plus que dans la construction de nouveaux modèles, il semble que c'est du côté de l'intégration des résultats de la prédiction à l'interface générale du dispositif que doivent être cherchées les futures améliorations du domaine. Ce constat est déjà clairement établi dans le domaine de l'aide au handicap. Il a en effet été relevé [BIA 06] que l'utilisateur handicapé ne faisait appel qu'assez rarement aux résultats de la prédiction, quand bien même celle-ci était pertinente. A ma connaissance, aucune expérience équivalente n'a été menée dans le cadre de la saisie de texte sur téléphones intelligents. Il me semble toutefois évident qu'en matière d'aide à la composition de messages, la balle est désormais dans le camp des chercheurs en interaction homme-machine afin que l'aide apportée par la prédiction de mots puisse être utilisée au mieux par l'utilisateur du système.

### 10.7. Bibliographie

- [ANT 07] ANTOINE J.Y., 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 », *TAL*, vol. 48, n° 2, p. 9-46, 2007.
- [BEL 07] BELATAR M., POIRIER F., « UniGlyph : une méthode universelle pour la saisie de texte sur dispositifs mobiles », Actes *IHM'2007*, Paris, p. 111-118, 2007.
- [BEC 04] BECK C., SEISENBACHER G., EDELMAYER G., ZAGLER W.L., « First user test results with the predictive typing system FASTY », Actes *ICCHP'04*, p. 3118., *LNCS Springer*, Paris, 2004.

- [BIA 06] BIARD N., DUMAS C., BOUTEILLE J., POZZI D., LOFASO F., LAFFONT I., « Apports de l'évaluation en situation de vie à partir d'une étude sur l'intérêt de la prédiction de mots auprès d'utilisateurs de synthèse vocale », Actes *Handicap 2006*, Paris, p. 145-148, 2006.
- [BIG 01] BIGI B., BRUN A., HATON J.-N., SMAÏLI K., ZITOUNI I., « Dynamic Topic Identification: Towards Combination of Methods », Actes *Recent Advances in Natural Language Processing workshop, RANLP'2001*, p. 255-257, 2001.
- [BLA 07] BLACHE PH., RAUZY S., « Le moteur de prédiction de mots de la Plateforme de Communication Alternative », dans *Traitement Automatique des Langues, TAL*, p. 48(2). 2007.
- [BOI 01] BOISSIERE PH., DOURS D., « VITIPI : Comment un système d'assistance à l'écriture pour les personnes handicapées peut offrir des propriétés intéressantes pour le TALN ? », Actes *TALN'2001, atelier TALN et Handicap*, vol. 2, p. 183-192, Tours, France, 2001.
- [CAR 97] CARLBERGER A., CARLBERGER J., MAGNUSON T., HUNNICUTT M.-S., PALAZUELOS-CAGIGAS S., NAVARRO S. A., « Profet, a new generation of word prediction : an evaluation study », Actes *NLPCA'97*, p. 23-28, Madrid, Espagne, 1997.
- [CHE 04] CHELBA C., JELINEK F., « Structured language modelling. Computer Speech and Language », n°14(4), p. 283-332, 2000.
- [CLA 97] CLARKSON P.-R., ROBINSON A.-J., « Language Model Adaptation using Mixtures and an Exponentially Decaying Cache », Actes *IEEE ICASSP-97*, Munich, RFA, 1997.
- [COC 98] COCCARO, N. AND JURAFSKY, D., « Towards better integration of semantic predictors in statistical language modelling », Actes *ICSLP'98*, Sydney, 1998.
- [COP 97] COPPESTAKE A., Augmented and alternative NLP techniques for augmentative and alternative communication », Actes *ACL workshop on Natural Language Processing for Communication Aids*, p. 37-42, Madrid, Espagne, 1997.
- [DEE 90] DEERWESTER S.-C., DUMAIS S., LANDAUER T., FURNAS G. HARSHMAN R. Indexing by Latent Semantic Analysis, *Journal of the American Society for Information Science, JASIS* 41(6), p. 391-407, 1990.
- [GOL 93] GOLDBERG D., RICHARDSON C., « Touch-typing with a stylus », Actes *Joint Conference INTERACT'93 et CHI'93*, Amsterdam, Pays-Bas, 1993.
- [GOO 01] GOODMAN J., « A Bit of Progress in Language Modeling », Extended Version, *Microsoft Research Technical Report MSR-TR-2001-72*, 2001.
- [GUE 92] GUENTHNER F., KRÜGER-THIELMANN K., PASERO R., SABATIER P., « Communications Aids for ALS Patients ». Actes *3<sup>rd</sup> International Conference on Computers for Handicapped Persons, ICCHP'92*, p. 303-307, 1992.
- [JEL 90] JELINEK F., « Self-organized Language Models for Speech Recognition dans Waibel A., Lee K.-F, *Redings in Speech Recognition*, p. 450-506, Morgan Kaufman, 1990.
- [KUH 90] KUHN R., DE MORI R., « A Cache-Based Natural Language Model for Speech Reproduction »; *IEEE Trans. PAMI*, 12 (6), p. 570-583, 1990.

- [LEP 97] LE PÉVEDIC B., « Prédiction morphosyntaxique évolutive dans un système d'aide à la saisie de textes pour les personnes handicapées physiques », Thèse U. Nantes, France, 1997.
- [LEP 98] LE PÉVEDIC B., « Les niveaux syntaxiques dans le système HandiAS », Actes *TALN'1998*, p. 132-141, Paris, 1998.
- [MAC 99] MACKENZIE I.-S., ZHANG S.X. *et al.* « Text entry using soft keyboards. Behaviour and Information Technology », 18(4), p. 235-244, 1999.
- [MAC 01] MACKENZIE I.-S., KOBER H. *et al.* « LetterWise : prefix-based disambiguation for mobile text input ». Actes *14<sup>th</sup> Annual ACM symposium on user interface software and technology*, Orlando, Florida, 2001.
- [MAT 03] MATIASEK H., BARONI M., « Exploiting long distance collocational relations in predictive typing » Actes *EACL'03 Workshop on Language Modeling for Text Entry Methods*, Budapest, Hongry, 2003.
- [MAU 00] MAUREL D., FOURCHE B., BRIFFAUT S., « Aider la communication en favorisant la saisie rapide de texte », Actes *Handicap'2000*, IFRATH, p. 87-92, Paris, France, 2000.
- [MEN 01] MÉNIER G., POIRIER F., « Système adaptatif de prédiction de texte », Actes *TALN'2001, atelier Handicap et Ingénierie des langues*, vol. 2, p. 213-222, Tours, France. 2001.
- [RON 86] RON D., SINGER Y., TISHBY N., « The power of amnesia : learning probabilistic automata with variable memory length »; *Machine Learning*, 25, p. 117-149, 1986.
- [SAD 09] SAD H., « *Text entry interfaces on mobile devices: modelling, design and evaluation* », Thèse de doctorat de l'université européenne de Bretagne, Vannes, France, 2 octobre 2009.
- [SAL 83] SALTON G., MCGILL M., « *Introduction to Modern Information Retrieval* », McGraw-Hill, New-York, 1983.
- [SCH 02] SCHADLE I., ANTOINE J.-Y., LE PÉVEDIC B., POIRIER F., « Sibylettré : prédiction de lettre pour la communication assistée ». *Revue d'Interaction Homme-Machine, RIHM*, vol. 3, n° 2. p. 115-133, 2002.
- [SCH 04] SCHADLE I., ANTOINE J.-Y., LE PÉVEDIC B., POIRIER F., « SibyMots : modélisation stochastique du langage intégrant la notion de *chunks* », Actes *TALN'2004*, Fès, Maroc. 2004.
- [STO 98] STOLCKE A., « Entropy-based pruning of backoff language models », Actes *DARPA Broadcast News Transcription and Understanding Workshop*, 1998.
- [STO 02] STOLCKE A., SRILM, « An Extensible Language Modeling Toolkit », Actes *International Conference on Spoken Language Processing, ICSLP'02*, Denver, Etats-Unis, 2002.
- [TRO 05] TROST H., MATIASEK J., BARONI M., « The language component of the FASTY text prediction system », *Applied Artificial Intelligence*. 19(8), p. 743-781, 2005.

- [VAN 92] VANDYKE J., MCCOY K., DEMASCO P., Using syntactic knowledge for word prediction. *Augmentative and Alternative Communication*, 8, p. 163-170, 1992.
- [WAN 05] WANDMACHER T., How semantic is Latent Semantic Analysis, Actes *RECITAL'2005*, Dourdan, France, 2005.
- [WAN 07] WANDMACHER T., ANTOINE J.-Y. EMNLP, 2007.
- [WAN 08] WANDMACHER T., ANTOINE J.-Y., DEPARTE J.-P., POIRIER F. SIBYLLE, « An assistive communication system adapting to the context and its user. *ACM Transactions on Accessible Computing* ». 1(1), p. 1-30, 2008.
- [WRI 94] WRIGHT J., JONES G., LLOYD-THOMAS H., « A robust language model incorporating a substring parser and extended N-grams. Actes *ICASSP'94*. p. 361-364, Adelaide, Australie, 1994.