

Comparaison de la méthode de saisie dichotomique avec les méthodes d'épellation traditionnelles

ÉQUIPE DICOTOMIX ENS de Lyon
dicotomix@ens-lyon.fr

Résumé

Les méthodes traditionnelles permettant de communiquer en se limitant à une entrée de la forme oui ou non sont basées sur une épellation lettre par lettre. Ici, une approche différente est considérée, où les mots sont épelés un à un par dichotomie sur l'ensemble du dictionnaire français. Cette méthode est ici comparée aux méthodes traditionnelles en vigueur dans le milieu médical. Nous montrons ainsi qu'une recherche dichotomique sur le dictionnaire permet d'écrire plus efficacement, en comparant les résultats obtenus en terme de temps de frappe de ces différentes méthodes.

I. INTRODUCTION

LES méthodes de saisie traditionnelles utilisées dans le milieu médical reposent toutes sur le même principe. Elles consistent à épeler les mots lettre à lettre, en parcourant l'alphabet pour chacune des lettres. Ce type de méthode peut s'avérer frustrant pour l'utilisateur, qui décide parfois même de renoncer à la communication, tant elle peut être fastidieuse. Une autre approche propose d'utiliser plus efficacement une entrée binaire afin d'écrire non pas une lettre à la fois, mais un mot à la fois. Ainsi, nous mettons ici en avant l'intérêt d'une telle méthode, en les comparant sur l'ensemble du dictionnaire.

II. MÉTHODES TRADITIONNELLES

Toutes les méthodes traditionnelles reposent sur une division de l'alphabet en une table en deux dimensions, où la recherche de la lettre voulue se fait en deux temps. Dans un premier temps, on recherche la ligne dans laquelle se situe la lettre voulue, en posant la question "est-ce que ma lettre est dans la première ligne", ou dans la seconde, jusqu'à arriver à la bonne

Code "EJASINT" conseillé par ALIS						
1	E	A	N	R	C	V
2	J	I	L	P	H	W
3	S	U	D	G	K	
4	T	M	B	Z		
5	O	F	X			
6	Q	Y				

Figure 1 – Alphabet EJASINT, majoritairement utilisé pour la saisie lettre à lettre par l'association ALIS

ligne. Ensuite, on itère de la même manière sur chacune des lettres de la ligne jusqu'à trouver la lettre voulue.

Une des méthodes d'épellation les plus répandues utilise l'alphabet EJASINT décrit dans la figure 11. Les autres méthodes envisagées ici sont décrites en annexe, et reposent toutes sur ce même principe de grille de lettres, avec une répartition différente de l'alphabet.

III. MÉTHODE DICHOTOMIQUE

Notre méthode repose sur un parcours du dictionnaire français par dichotomie, en tenant compte des fréquences des mots dans l'usage courant de la langue française. La question posée porte ainsi sur des mots et n'est plus de la forme "est-ce le bon mot?", mais "est-il avant ou après dans l'alphabet?"

Une seconde méthode que nous évoquerons est la méthode de l'épellation dichotomique, reprenant le principe de la dichotomie mais cette fois pour une épellation lettre par lettre : "la lettre recherchée est-elle avant ou après dans l'alphabet?".

L'intérêt de l'approche dichotomique est très simple : à chaque question, la moitié des possibilités sont éliminées. Cela permet de se rapprocher rapidement de la solution, même dans un grand lexique.

IV. ÉPELLATION COMPLÈTE

Nous nous référerons, dans cette partie, aux différentes méthodes au travers des numéros suivants : 0 - recherche dichotomique, 1 - épellation dichotomique, 2 - alphabet EJASINT, les suivants correspondant aux numéros des alphabets en annexe.

i. Distribution des temps de frappe

Pour mesurer l'efficacité de chaque méthode, nous comptons le nombre de questions nécessaire pour trouver les mots. Ainsi, nous considérons chaque mot du dictionnaire et calculons le temps qu'il faut pour le trouver.

La distribution de ces longueurs de frappe, ajustée à l'aide de la fréquence d'apparition des mots dans la langue française, est un bon indicateur de la rapidité d'utilisation des différentes méthodes. La figure 2 montre que la recherche dichotomique affiche une fréquence plus élevée pour les mots de longueur de frappe faible : toutes les autres méthodes nécessitent parfois énormément plus de questions pour épeler un même mot.

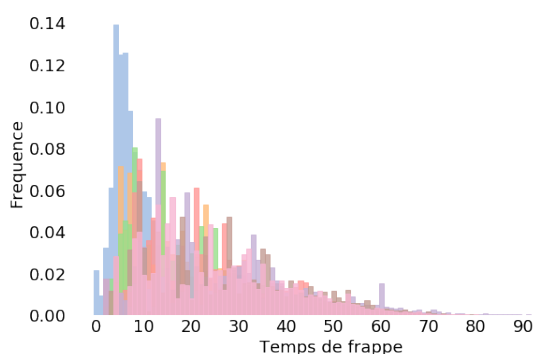


Figure 2 – Distribution du temps de frappe en fonction de la longueur des mots pour différentes méthodes. Bleu - recherche dichotomique, orange : épellation dichotomique, vert : alphabet EJASINT, rouge : 3, violet : 4, marron : 5, rose pâle : 6

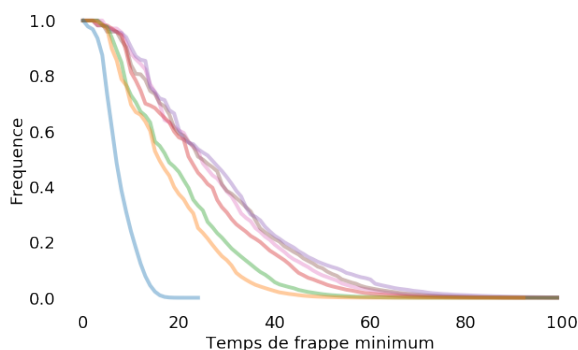


Figure 3 – Proportion de mots ayant un temps de frappe supérieur à un temps donné. Bleu - recherche dichotomique, orange : épellation dichotomique, vert : alphabet EJASINT, rouge : 3, violet : 4, marron : 5, rose pâle : 6

La figure 3 montre la probabilité pour qu'un mot ait une longueur de frappe supérieure à une longueur de frappe donnée. Elle confirme cette tendance, étant donné que la fréquence d'apparition de mots nécessitant au moins 20 coups pour être tapés est presque nulle, contre plus de 50% pour les autres méthodes.

La méthode d'épellation dichotomique, quant à elle, donne des résultats proches de l'alphabet EJASINT.

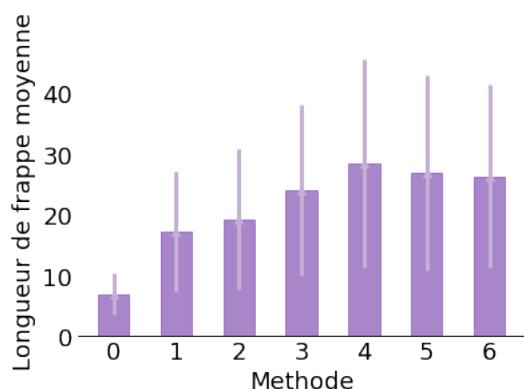


Figure 4 – Longueur de frappe moyenne sur l'ensemble du dictionnaire (barre violettes), avec l'écart-type de la distribution (petites barres violettes claires). 0 - recherche dichotomique, 1 - épellation dichotomique, 2 - alphabet EJASINT, 3 à 6 - méthodes 3 à 6

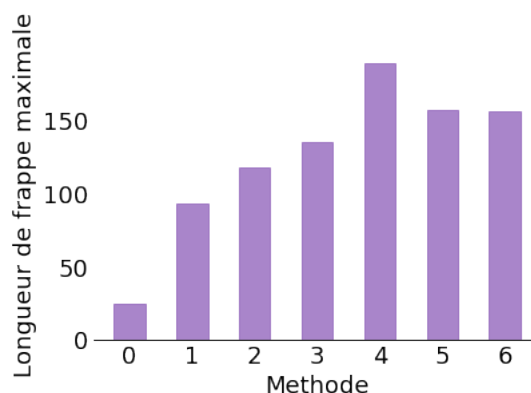


Figure 5 – Temps de frappe du mot le moins accessible. 0 - recherche dichotomique, 1 - épellation dichotomique, 2 - alphabet EJASINT, 3 à 6 - méthodes 3 à 6

ii. Moyenne et maximum

Les distributions donnent beaucoup d'information sur les méthodes, mais les graphes sont difficilement lisibles. D'un point de vue pratique, on est intéressé principalement au temps moyen nécessaire pour écrire un mot, et le temps maximal garanti.

La figure 4 montre que la recherche dichotomique a une longueur de frappe moyenne significativement plus faible que toutes les autres méthodes considérées, ce qui signifie que les mots sont en moyenne tapés trois à cinq plus vite. De plus, son écart-type plus faible donc il est peu probable d'avoir à écrire un mot "long".

Même si l'intervention de l'aidant permet souvent de trouver les mots sans épeler toutes les lettres, on peut s'intéresser au pire des cas, c'est-à-dire au mot du dictionnaire le plus long à trouver. On remarque sur la figure 15 que la dichotomie sur le lexique offre une performance drastiquement meilleure que les autres méthodes, avec un facteur variant de quatre à sept.

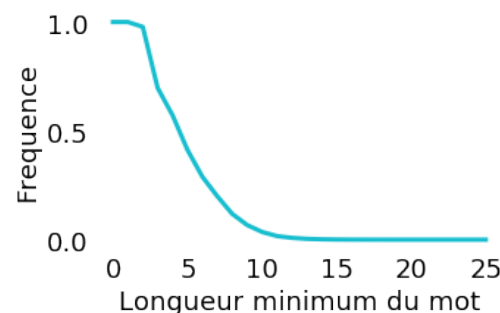


Figure 6 – Fréquence des mots de longueur supérieure à une longueur donnée dans le dictionnaire

V. ÉPELLATION DES 4 PREMIÈRES LETTRES

Remarquons que, en fréquence, plus de la moitié des mots du français continennent moins de 4 lettres (cf figure 6). Par ailleurs, faisons l'hypothèse qu'en pratique, l'on arrive à savoir quel est le mot tapé à partir de la quatrième lettre. Nous reprenons donc dans cette partie la même comparaison des différentes méthodes, cette fois en se limitant à l'épellation des 4 premières lettres de chaque mot. Notons par ailleurs, que la méthode de recherche dichotomique continue, en revanche, à épeler les mots complètement.

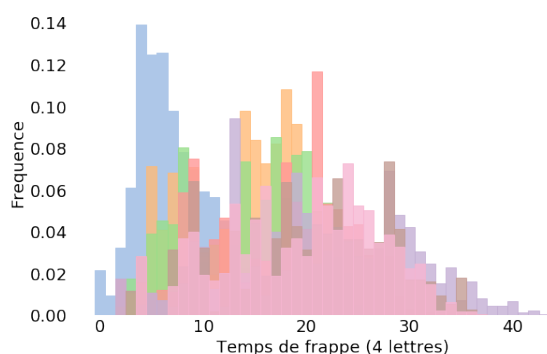


Figure 7 – Distribution du temps de frappe en fonction de la longueur des mots pour différentes méthodes (épellation des 4 premières lettres uniquement). Bleu - recherche dichotomique, orange : épellation dichotomique, vert : alphabet EJA SINT, rouge : 3, violet : 4, marron : 5, rose pâle : 6

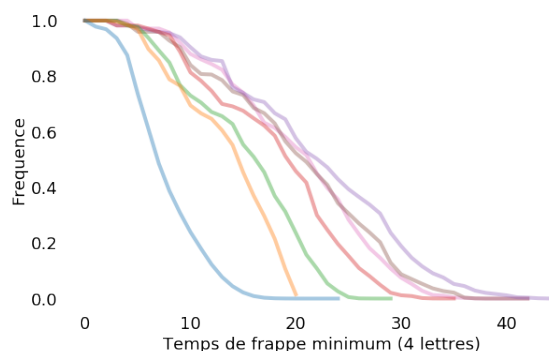


Figure 8 – Proportion de mots ayant un temps de frappe donné (épellation des 4 premières lettres uniquement). Bleu - recherche dichotomique, orange : épellation dichotomique, vert : alphabet EJA SINT, rouge : 3, violet : 4, marron : 5, rose pâle : 6

i. Distribution des temps de frappe

La distribution des temps de frappe (figure 7) continue à donner la même tendance que précédemment, et la méthode de recherche dichotomique permet toujours de trouver plus rapidement les mots les plus fréquents.

Cette tendance est également confirmée par la fréquence des mots ayant une longueur de frappe supérieure à une longueur donnée (figure 8).

ii. Moyenne et maximum

La longueur moyenne de frappe (figure 9) reste aussi avantageuse par rapport aux autres méthodes. Cependant, la longueur maximale (10) est plus élevée que pour l'épellation dichotomique. Malgré tout, la recherche dichotomique reste meilleure, de ce point de vue, que toutes les méthodes à présent utilisées dans le milieu médical.

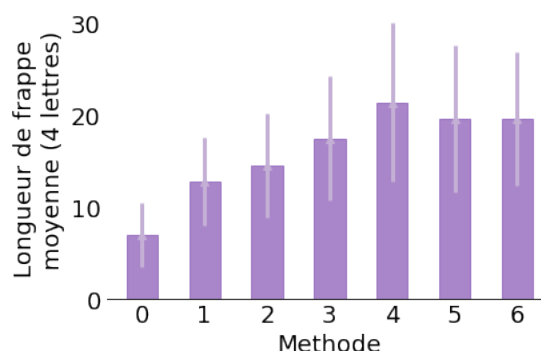


Figure 9 – Longueur de frappe moyenne sur l'ensemble du dictionnaire (barre violettes), avec l'écart-type de la distribution (petites barres violettes claires) (épellation des 4 premières lettres uniquement). 0 - recherche dichotomique, 1 - épellation dichotomique, 2 - alphabet EJA SINT, 3 à 6 - méthodes 3 à 6

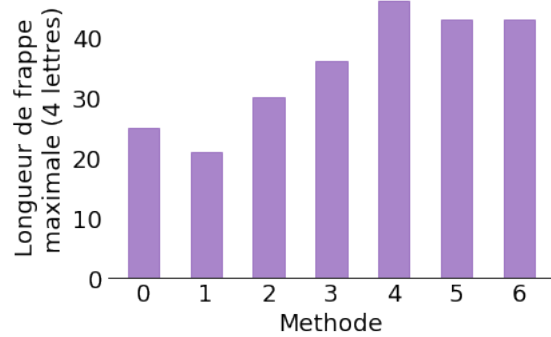


Figure 10 – Temps de frappe du mot le moins accessible (épellation des 4 premières lettres uniquement). 0 - recherche dichotomique, 1 - épellation dichotomique, 2 - alphabet E/JASINT, 3 à 6 - méthodes 3 à 6

VI. CONCLUSION

La méthode de recherche dichotomique utilisée par le projet Dicotomix est donc une méthode permettant, d'un point de vue théorique comme pratique, d'écrire de façon significativement plus rapide que les méthodes actuelles. Cependant, elle nécessite un coût cognitif plus important, afin de répondre aux questions nécessaires à identifier un mot.

ANNEXES

Code "EJASINT" conseillé par ALIS						
1	E	A	N	R	C	V
2	J	I	L	P	H	W
3	S	U	D	G	K	
4	T	M	B	Z		
5	O	F	X			
6	Q	Y				

Figure 11 – Alphabet EJASINT, majoritairement utilisé pour la saisie lettre à lettre par l'association ALIS

A	E	I	O	U	Y					
B	C	D	F	G	H	J	K	L		
M	N	P	Q	R	S	T	V	W	X	Z

Figure 12 – Alphabet de la méthode d'écriture 3.

A	Z	E	R	T	Y	U	I	O	P
Q	S	D	F	G	H	J	K	L	M
W	X	C	V	B	N				

Figure 13 – Alphabet de la méthode d'écriture 4.

Consonnes 1^{re} partie

B	C	D	F
G	H	J	K

Consonnes 2^e partie

L	M	N	P
Q	R	S	T
V	W	X	Z

Voyelles

A	E	I
O	U	Y

Figure 14 – Alphabet de la méthode d'écriture 5.

B	C	D	F	G
H	J	K	L	M

N	P	Q	R	S
T	V	W	X	Z

A	E	I	O	U	Y
---	---	---	---	---	---

Figure 15 – Alphabet de la méthode d'écriture 6.