

Choisir un nom de domaine diffusé à l'oral

L'enjeu de la retranscription et de la mémorisation

Septembre 2018

afnic

Table des matières

1. Résumé	3
2. Les fondements théoriques -	4
3. L'enjeu de la mémorisation–	5
3.1. L'effet de lexicalité.....	5
3.2. L'effet de longueur de mot	5
3.3. Les effets de fréquence... ..	5
3.4. ...et de concrétude	5
4. L'enjeu de la retranscription	7
4.1. La connaissance et le respect des patterns existants :	7
4.2. L'importance des sons prononcés avant (Perry, 2003)	7
4.3. L'effet de régularité (Delattre, Bonin & Barry, 2006)	7
5. L'expérimentation	8
Les variables	8
L'échantillonnage	8
Le matériel.....	8
La procédure.....	10
6. Les hypothèses	11
7. Traitement des données.....	12
8. Les résultats – pour tous les noms de domaine (Mots et Pseudo-mots ensemble)	13
9. Analyse qualitative	15
9.1. L'ambiguïté du presque-mot.....	15
9.2. Le choix du plus simple.....	15
9.3. La force de la sémantique.....	15
10. Bibliographie	16
11. A propos de Fast & Fresh	17
12. A propos de l'Afnic	18

1. Résumé

Comment un nom de domaine est-il perçu quand il est entendu (par ex. à la radio) mais qu'il n'est pas vu ? Arrive-t-on facilement à l'écrire comme son propriétaire le souhaite ? Est-il simple de le mémoriser ?

Pour répondre, nous avons étudié les mécanismes cognitifs qui aident les auditeurs à se souvenir et à retranscrire un nom de domaine. Lors d'une collaboration précédente, l'agence de Recherche Utilisateur [Fast & Fresh](#) et le laboratoire Epsilon de Psychologie cognitive de Montpellier, ont mené une étude sur la mémorisation des noms de domaine quand ceux-ci sont vus. (cf. [8 astuces pour bien choisir son nom de domaine](#)). Cet article y fait suite.

2. Les fondements théoriques

Tout d'abord, précisons que les noms de domaine sont **des mots ou des pseudo-mots** (c'est à dire des mots inventés). Ensuite, ces mots ou ces pseudo-mots sont construits avec une orthographe et donc des codes phonétiques. Enfin, ils font généralement référence à un concept, etc.

Pour comprendre comment un nom de domaine est perçu quand il est entendu (par exemple à la radio) nous avons commencé par chercher dans la littérature scientifique. Notamment, celle sur la linguistique et la psychologie cognitive. Ceci afin d'identifier les bonnes questions à se poser.

A la lecture des articles de recherche, on se rend compte que c'est dans l'étude de **la mémoire à court terme** et plus spécifiquement de la mémoire de travail que l'on trouve le plus de réponses.

La mémoire de travail est définie comme « un sous-ensemble actif de la mémoire à court terme qui permet à la fois de maintenir et de manipuler une information en mémoire ». Prenons un exemple simple : quand nous tentons de retenir le code d'accès d'un immeuble tout en cherchant encore le nom de la rue où l'immeuble se trouve, nous faisons fonctionner notre mémoire de travail. On trouvera d'avantage de renseignements en lisant la [définition](#) donnée par le CRFNA de Bruxelles.

Dans l'exemple précédent, la personne retient le code d'accès de l'immeuble en le répétant mentalement. La sous-partie de la mémoire de travail alors activée est **la boucle phonologique**. La boucle phonologique **nous permet de répéter** (consciemment ou inconsciemment) une information vue ou entendue afin qu'elle **reste plus longtemps en mémoire**. Comme quand on répète mentalement un numéro de téléphone qu'on souhaite mémoriser avant de l'écrire sur un bloc note ou de le composer. Comme nous allons le voir, **ce processus est sensible à la nature des informations et aux interférences** entre les connaissances passées, présentes et futures.

Dans certaines conditions, les mots existants (leboncoin.fr) et les mots inventés (amazon.fr) sont perçus pareillement. Dans d'autres, ils sont manipulés différemment par les individus. **Comme les processus d'apprentissages** sont au cœur de ces questions, les sciences humaines ont de nombreux résultats à nous fournir.

3. L'enjeu de la mémorisation–

Retenir un mot est-il comme retenir un pseudo-mot ? **Quelle taille** un mot doit-il faire pour revenir facilement en mémoire ? Le mot doit-il **avoir un sens** que l'auditeur comprend ? Examinons quelques effets permettant d'apporter un début de réponse.

3.1. L'effet de lexicalité

De nombreux chercheurs parmi lesquels Hulme, Maughan et Brown (1991) montrent que lorsqu'on donne une liste de mots à apprendre puis à rappeler à des sujets (tâche de rappel), les sujets retiennent **plus de mots que de pseudo-mots**. Cet effet s'appelle l'effet de lexicalité. Ce sont **les unités qui ont un sens** pour les groupes d'individus (leur **lexique commun**) qui laissent **le plus de traces en mémoire**. Cet effet a été montré sur les mots vus et entendus mais nous voulons vérifier qu'il perdure de la même manière quand on les utilise comme noms de domaine. En effet, même si les noms de domaine ne renvoient pas forcément à un concept précis, nous voulons vérifier que le sens d'un mot a tout de même un fort pouvoir de rappel.

3.2. L'effet de longueur de mot

Dans la précédente étude, l'agence Fast and Fresh et le laboratoire Epsilon avaient montré que **plus un mot vu est petit, plus il est rapidement encodé en mémoire**. D'autres auteurs (Baddeley, Thomson & Buchanan, 1975) ont montré qu'on rappelait davantage de mots courts que de mots longs. Cependant, ces résultats s'appliquent particulièrement aux mots présentés visuellement et **pas forcément aux mots présentés en modalité auditive**. En effet, Pollack et Pickett (1964) ont montré que les mots d'une syllabe prononcés seuls étaient moins bien perçus que les mots de deux syllabes. **Un nom de domaine « trop » court pourrait être perçu incorrectement.**

3.3. Les effets de fréquence...

Quelle que soit la modalité (visuelle ou auditive), **plus un mot est fréquent dans la langue des individus mieux ils sont rappelés** (voir Majerus, 2010).

3.4. ...et de concrétude

Les mots concrets, c'est-à-dire **des mots imageables ou qui renvoient à quelque chose de tangible**, sont souvent **mieux retenus que les mots abstraits**. Selon les conditions, l'effet peut être plus ou moins

fort (Walker & Hulme, 1999) mais il n'est pas montré l'inverse (que les mots abstraits soient mieux retenus que les mots concrets)

4. L'enjeu de la retranscription

Est-il si simple d'écrire ce que l'on entend ? Est-ce dépendant de la taille du mot ? Ecrivons-nous tous là même chose ? Pensons à toutes nos dictées ratées et écoutons ce que les linguistes disent.

4.1. La connaissance et le respect des patterns existants :

D'après Joshi, Treiman, Carreker et Moats (2009), épeler un mot est plus facile quand les individus connaissent **l'origine et l'histoire du mot**. Mais aussi, lorsque les individus connaissent **les parties signifiantes des mots (préfixes, suffixes et racines)**. **Ces parties signifiantes sont en effet très utiles** pour transcrire un mot d'après un son entendu. Ainsi, pour qu'un mot soit ait des chances d'être bien retranscrit, il faut privilégier les patterns qui représentent le plus fréquemment les sons des discours dans une langue donnée. Par exemple, les mots français ne commencent jamais par des consonnes identiques, donc entre 'nuss' et 'nnus', c'est le premier qui sera désigné pour ressembler le plus à un mot et qui sera le mieux retranscrit.

4.2. L'importance des sons prononcés avant (Perry, 2003)

Si un site s'appelle 'cabotti.fr', le pseudo-mot cabotti sera probablement épelé avec deux 't' si le mot 'botte' a été prononcé juste avant. En effet, dans **une situation à plusieurs solutions orthographiques**, le fait de faire écouter un mot existant avec une certaine orthographe juste avant l'écoute d'un pseudo-mot **influencera l'orthographe du mot inventé**.

4.3. L'effet de régularité (Delattre, Bonin & Barry, 2006)

Les mots irréguliers, c'est-à-dire ceux qui ont **une petite probabilité de s'écrire comme ils s'entendent**, sont **plus longs à produire et à écrire** que les mots réguliers. Par exemple, il serait apparemment plus facile de générer à l'écrit le mot 'lune' que le mot 'train' parce que les sons du mot 'lune' correspondent à leur retranscription **sans beaucoup d'autres concurrents** potentiels. Alors que le mot train ne s'écrit pas comme il se prononce.

5. L'expérimentation

Les variables

Parmi tous ces effets nous avons sélectionné trois variables principales : Le Type de noms (Mots VS Pseudo-mots), la Longueur des noms (Courts VS Moyens VS Longs) et la Régularité de l'orthographe d'après le son (Régulier VS Irrégulier).

L'échantillonnage

Notre échantillon était composé de 30 participants âgés de 18 à 67 ans (M = 27 ans). Il y avait 18 hommes et 12 femmes. Tous avaient pour langue maternelle le Français.

Le matériel

Pour inventer de faux noms de domaine, 30 mots et 30 pseudo-mots ont été sélectionnés. Grâce à la base de donnée Lexique 3 (New, Pallier & Ferrand, 2005) nous avons déterminé, critère par critère, les mots et pseudo-mots que nous souhaitions tester.

Noms de domaine	Orthographe du nom	Longueur du nom	Fréquence (pour un million de mots) / Concrétude (moyenne)	
Mots	Régulier	Court	< 50 / = 3,37	
		Moyen		
		Long		
	Irrégulier	Court		> 50 / = 3,87
		Moyen		
		Long		
Pseudo-mots	Régulier	Court		
		Moyen		
		Long		
	Irrégulier	Court		
		Moyen		
		Long		

Pour les mots comme pour les pseudo-mots, un nom de domaine est désigné **Court avec 3 ou 4 lettres**, **Moyen avec 5 ou 6 lettres** et **Long avec 7 ou 8 lettres**.

Un nom de domaine est dit régulier quand ses sons correspondent à une orthographe ayant 1 ou 2 autres possibilités. Il est dit irrégulier quand il en contient au moins 3 ou qu'il utilise des formes rares (e.g 'chez').

Par exemple, les mots 'bouton' et 'fourche' s'écrivent comme ils s'entendent alors que les mots 'accord' et 'chez' seront décrits comme irréguliers. Les pseudo-mots 'uron' et 'drustre' sont décrits comme réguliers tandis que les pseudo-mots 'cank' et 'adau' seront dits irréguliers.

La procédure

L'expérimentation se déroulait devant un ordinateur, sur une expérience construite sur le logiciel Open Sesame.

Les participants avaient un casque sur les oreilles et entendaient une première liste de 30 noms de domaine. La moitié des participants commençait par une liste de mots et l'autre moitié par une liste de pseudo-mots.

Une fois le nom prononcé, un écran noir apparaissait et le participant avait pour consigne d'écrire ce qu'il avait entendu. Une fois la liste de mots (ou de pseudo-mots) finie, l'individu devait rappeler le plus de mots possible dont il se souvenait.

Le participant recommençait alors le processus avec la deuxième liste (de pseudo-mots ou de mots).

Nous avons enregistré les réponses des participants quand ils écrivaient le nom après l'avoir entendu ainsi que quand ils devaient fournir les rappels de milieu et de fin d'expérience.

Ainsi, nous avons deux indicateurs de nos variables :

- la **moyenne de retranscription correcte** (= quand un nom de domaine est orthographié comme son modèle)
- ainsi que la **moyenne d'occurrence d'un nom de domaine** (= par exemple le nombre de fois où le mot 'soir' était cité en rappel).

6. Les hypothèses

- **Hypothèse 1** : La signification des mots (effet de lexicalité)

Un mot existant sera mieux mémorisé et mieux retranscrit qu'un pseudo-mot (+ d'indices accessibles en mémoire). « cheval » sera plus de fois écrit correctement et plus retenu que « chudal ».

- **Hypothèse 2** : Le respect des codes existants

Les mots et les pseudo-mots à l'orthographe régulière seront respectivement mieux retranscrits que les mots et pseudo-mots irréguliers (- de concurrents ; + accessible en mémoire). « servir » sera plus de fois écrit correctement que « agenda » / « micha » sera plus de fois écrit correctement que « krouta ».

- **Hypothèse 3** : La longueur des mots

Un nom de domaine court sera moins bien perçu et reproduit qu'un nom de domaine de 5 lettres ou plus (trop peu d'indices phonologiques pour les noms courts). « glucol » sera plus de fois écrit correctement et plus retenu que « glu ».

7. Traitement des données

Les données enregistrées par le logiciel Open Sesame ont été mis en forme sur Excel et traitées sur le logiciel SPSS (V.19). Que ce soit pour l'analyse de l'indicateur du nombre de transcription correcte ou d'occurrence (mots et pseudo-mots confondus), les valeurs obtenues nous permettaient de procéder à des analyses paramétriques (P valeur du test de normalité > 0.05 ; P valeur du test de Mauchly > 5). Nous avons donc conduit une ANOVA à mesures répétées.

Note : **Nous ne relatons ici que les résultats significatifs** à $p < 0.05$ pour les effets principaux et $p < 0.025$ pour les interactions.

8. Les résultats – pour tous les noms de domaine (Mots et Pseudo-mots ensemble)

Hypothèse 1 - Nous observons que **l'hypothèse de signification des noms est validée**. En effet, **les mots sont à la fois mieux retranscrits et mieux retenus que les pseudo-mots**. Cet effet est dû à l'importance de la re-connaissance dans le processus de perception et de mémorisation. Reconnaître et écrire le mot 'costaud' pose peu de problèmes comparé à déterminer si c'est bien 'nhisadre' que nous avons entendu. Se remémorer ce même pseudo-mot est également complexe, car **nous n'avons pas le même nombre d'indices à disposition pour nous le rappeler**.

Concrètement -

Le mot 'panthère', malgré son orthographe complexe, sera mieux retranscrit et davantage retenu que le pseudo-mot 'clarge'.

Hypothèse 2 - **L'hypothèse sur la régularité entre l'auditif et la transcription est en partie validée**. En effet, **les noms de domaine qui s'écrivent comme ils s'entendent** et qui ont peu de façons différentes de s'orthographier **sont mieux retranscrits** que les mots à l'orthographe plus irrégulière.

En revanche, **ils ne semblent pas globalement mieux retenus** que les noms de domaine irréguliers :

- **Quand les mots réguliers sont moyens et longs, ils sont significativement mieux retenus que les mots courts.**
- **Quand les mots sont irréguliers, ce sont les mots longs uniquement qui sont significativement mieux retenus** que les mots courts et moyens.

On peut ainsi voir que **les mots moyens (de 5 et 6 lettres) sont particulièrement sensibles à la régularité de l'orthographe**.

Concrètement -

- *Pour que vos clients épèlent correctement vos noms de domaine, privilégiez les noms qui s'écrivent simplement tel que 'survenir' et non comme 'beaucoup'.*
- *De même, si votre nom est inventé, privilégiez le pseudo-mot 'drustre' que le pseudo-mot 'poussno!'.*
- *Nous ne pouvons affirmer que les noms les plus réguliers seront les mieux rappelés. En revanche, si vous créez un nom de domaine qui est un mot existant de 5 ou 6 lettres, privilégiez une orthographe régulière, il sera mieux retenu qu'avec une orthographe complexe.*

Hypothèse 3 - **L'hypothèse sur la longueur des noms de domaine est en partie validée** également. Nous avons supposé qu'un nom court serait toujours moins bien perçu qu'un nom moyen ou long. Or, quand une personne doit écrire les mots qu'il entend, **il écrit avec moins de fautes des mots de 5 lettres ou plus (mots moyens et longs) que des mots de 3 ou 4 lettres (mots courts)**. Ce qui va dans le sens de notre hypothèse.

En revanche, quand une personne doit écrire des pseudo-mots qu'il a entendu, alors il fait moins de

fautes en écrivant des pseudo-mots courts que des pseudo-mots moyens et longs. Ce qui va à l'inverse de nos hypothèses et qui doit être due à la complexité d'épeler un mot que nous ne reconnaissons pas.

Concernant le rappel, les pseudo-mots ne montrent pas de résultats significatifs. Pour les mots en revanche, nous montrons que **les mots longs sont mieux retenus que les mots courts et moyens.**

Concrètement-

- Si vous créez un nom de domaine qui est un mot, privilégiez les noms de domaine de plus de 5 lettres, ils seront écrits avec moins de fautes. S'il fait plus 7 ou 8 lettres, il sera même davantage retenu que s'il en fait 5 ou 6.

- Si vous créez un nom de domaine qui est un pseudo-mot, privilégiez les formes courtes pour limiter le nombre d'erreur !

9. Analyse qualitative

La notion d'orthographe régulière ou irrégulière peut être complexe mais prenons des exemples de l'étude pour illustrer les risques liés à la présentation auditive des noms de domaine : nous avons décidé d'inventer les noms 'adau', 'hox', 'plomme' et 'crier'.

9.1. L'ambiguïté du presque-mot

'adau' n'a jamais été retranscrit comme nous l'avons écrit. A la place, les personnes ont privilégié le raccourci 'ado' qui s'apparente à un « vrai » mot. Donc attention, **les jeux de mots sont ludiques mais sont confus à notre oreille !**

Un conseil : prévenez vos auditeurs de la feinte.

9.2. Le choix du plus simple

'hox' n'a jamais été retranscrit comme nous l'avons écrit. Il a pris des formes diverses ('ox', 'oxe', etc) mais le 'h' muet du début n'est jamais apparu. Donc attention, **les lettres muettes deviennent invisibles à l'écrit.**

Un conseil : si vous faites un site pour de l'horticulture ou un horoscope, dites-le, l'auditeur aura plus de chance d'écrire votre nom de domaine avec un 'h' désormais. Sinon, faites entendre que ce 'h' existe.

9.3. La force de la sémantique

'plomme' a été quelques fois écrit ainsi, sinon il a pris des formes diverses 'plom', 'plome'. Le pseudo-mot ressemble au mot fréquent 'pomme' et certaines personnes l'écriront donc bien comme vous avez voulu. Mais pas tous. De la même manière 'crier' était souvent écrit 'crier' ou 'crire'. On peut supposer que c'est du fait de la proximité avec le son du verbe 'écrire'. Ainsi, les individus **s'affranchissent parfois du choix le plus simple pour enlever ou rajouter des lettres qui font sens pour eux.**

Un conseil : Affirmer les liens qui lient ces éléments entre eux s'il y en a. Sinon, si votre pseudo-mot a une signification dans une autre langue ou dans votre imagination, dites-le à vos auditeurs. En effet, comme l'ont montré Hulme, Maughan et Brown (1991), quand **donne la traduction d'un mot inconnu, son taux de rappel augmente.**

10. Bibliographie

- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, 14(6), 575-589. [http://dx.doi.org/10.1016/S0022-5371\(75\)80045-4](http://dx.doi.org/10.1016/S0022-5371(75)80045-4)
- Delattre M., Bonin P., Barry C. (2006). [Written spelling to dictation: Sound-to-spelling regularity affects both writing latencies and durations.](#) *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1330–1340.
- Desrochers, A., & Thompson, G. L. (2009) Subjective frequency and imageability norms for 3,600 French words, *Behavioral Research Methods*, 41, 2009, 546-557
- Hulme, C., Maughan, S., & Brown, G. (1991). Memory for Familiar and Unfamiliar Words: Evidence for a Long-Term Memory Contribution to Short-Term Memory Span. *Journal of memory and language*, 30 685-701.
- IBM Corp. Released 2010. IBM SPSS Statistics for Windows, Version 19.0. Armonk, NY: IBM Corp.
- Joshi R. M., Treiman R., Carreker S., Moats L. (2009). How words cast their spell: Spelling is an integral part of learning the language, not a matter of memorization. *American Educator*, 32 (4), 6–43.
- Majerus, S. (2010). [Les multiples déterminants de la mémoire à court terme verbale : Implications théoriques et évaluatives.](#) *Développements*, 4,(1), 5-15. doi:10.3917/devel.004.0005.
- New, B., Pallier, C., Brysbaert, M., Ferrand, L. (2004) Lexique 2 : A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers*, 36 (3), 516-524.
- New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007) The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28 661–67.
- Perry, C. (2003). Priming the rules of spelling. *Quarterly Journal of Experimental Psychology*, 56A, 515–530.
- Pollack, I. and Pickett, J.M. (1964). The intelligibility of excerpts from fluent speech: auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior*, 3, 79–84.
- Walker, I., & Hulme, C. (1999) Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25 (5), pp. 1256-1271.

11. A propos de Fast & Fresh



Fast & Fresh est une agence de Stratégie en Sciences Humaines fondée en 2009 à Paris.

Elle est composée d'une part d'un Laboratoire National de Psychologie Cognitive & Sociale dont le travail est de comprendre le comportement des consommateurs en profondeur, notamment sur les dimensions qui ne sont habituellement pas accessibles avec les tests déclaratifs.

Et d'autre part d'un commando Stratégique réunissant des compétences plus larges et plus sénior que les compétences étroitement habituelles afin de pouvoir affronter avec intelligence les grands défis de l'innovation.

Fast & Fresh se consacre essentiellement à des sujets e-commerce, luxe, jeu & expérientiels et socio-environnementaux.

Enfin, Fast & Fresh a pour vocation d'éduquer les marques aux besoins des usagers afin de remettre en selle les relations clients sur de bonnes bases.

Le travail de l'agence est d'émanciper les utilisateurs et les marques. Souscrivant, par conviction, au code de déontologie des Psychologues, nous ne pratiquons donc pas le Neuromarketing.

<http://www.fastandfresh.fr/>

12. A propos de l'Afnic

L'Afnic (Association française pour le nommage Internet en coopération) est une association française à but non lucratif. Depuis 20 ans, elle est l'office d'enregistrement pour la gestion des adresses internet sous l'extension *.fr*. Elle gère également les extensions ultramarines *.re* (Ile de la Réunion), *.pm* (Saint-Pierre et Miquelon), *.tf* (Terres australes et antarctiques françaises), *.wf* (Wallis et Futuna), *.yt* (Mayotte), ce qui représente plus de 3,2 millions de noms de domaine.

L'Afnic est également l'opérateur technique de 14 entreprises et collectivités ayant choisi d'avoir leur propre extension dont *.paris*, *.bzh*, *.alsace*, *.corsica*, *.mma*, *.ovh*, *.leclerc* ou encore *.snf*.

L'Afnic est engagée à accompagner la transformation numérique des TPE/PME grâce à notre dispositif *Réussir en .fr* (www.reussir-en.fr) et propose une offre gratuite d'accompagnement à la présence en ligne allant des outils de diagnostic aux formations sur le terrain dans toute la France. En tant qu'association, elle fédère une communauté de plus d'une centaine de membres aux profils variés mais tous acteurs du web : bureaux d'enregistrement, entreprises, fédérations, utilisateurs, institutionnels, etc.

Le rôle de l'Afnic s'inscrit dans une mission d'intérêt général plus large, qui consiste à contribuer au quotidien à un internet sûr et stable, ouvert aux innovations où la communauté internet française joue un rôle de premier plan.

Par ailleurs, l'Afnic reverse 90% des bénéfices de la gestion du *.fr* à la Fondation Afnic pour la solidarité numérique (www.fondation-afnic.fr) qui finance chaque année une trentaine de projets visant à réduire la fracture numérique sur tout le territoire français.

<https://www.afnic.fr>