Choosing a domain name to be heard but not seen

The challenge of transcribing and memorizing

September 2018

afnic

Contents

1.　Abstract

How is a domain name perceived when it is heard (eg on the radio) but not seen? Is it as easy to write it as its owner would like? Is it easily memorized?

To answer these questions, we studied the cognitive mechanisms that help listeners remember and transcribe a domain name. In a previous collaborative study, the User Research Agency **Fast & Fresh** and the Epsylon laboratory of Cognitive Psychology in Montpellier, conducted a study on the memorization of domain names when they are seen. (cf. **8 tips for choosing the right domain name**). This article follows on from that study.

2. The theoretical foundations

First of all, it should be remembered that domain names are **words or pseudo-words** (ie invented words). Secondly, these words or pseudo-words are constructed with a spelling and thus phonetic codes. Finally, they usually refer to a concept, etc.

To understand how a domain name is perceived when it is heard (for example on the radio) we started by researching the scientific literature. In particular, that on linguistics and cognitive psychology. The purpose was to identify the right questions to ask.

In reading the research articles, we realized that the study of **short-term memory** and more specifically working memory was where we found the most answers.

Working memory is defined as "an active subset of short-term memory that allows information both to be held and processed". Take a simple example: when we try to remember the access code of a building while still searching for the name of the street where the building is located, we are using our working memory. More information can be found by reading the **definition** given by the Center for Ambulatory Neurological Functional Rehabilitation (CRFNA) in Brussels.

In the previous example, the person retains the access code of the building by repeating it mentally. The sub-part of the working memory then activated is **the phonological loop.** The phonological loop **allows us to repeat** (consciously or unconsciously) information seen or heard so that it **can be held longer in memory**. Such as when we mentally repeat a phone number we want to memorize before writing it on a notepad or dialling it. As we will see, **this process is sensitive to the nature of information and to the interference** between past, present and future knowledge.

Under certain conditions, existing words (leboncoin.fr) and the invented words (amazon.fr) are perceived similarly. In others, they are processed differently by different people. **Since learning processes** are central to these questions, humanities can provide many results.

3. The issue of memorization-

Is remembering a word the same as remembering a pseudo-word? **How long** should a word be to be easily remembered? Does the word have to **make sense** for the listener to understand it? Let's take a look at some of the effects that can give us some initial answers.

3.1. The effect of lexicality

Many researchers, including Hulme, Maughan, and Brown (1991), ave shown that when a list of words is given to be learnt and then recalled (reminder task), the subjects retain **more words than pseudo-words**. This effect is called the effect of lexicality. It refers to **units that make sense** for groups of individuals (their **common lexicon**) that leave **the most traces in memory**. This effect has been shown on words seen and heard but we wanted to check whether it persisted in the same way when the words are used as domain names. Because even if domain names do not necessarily refer to a specific concept, we wanted to check that the meaning of a word still has a strong power of recall.

3.2. The effect of word length

In the previous study, the Fast and Fresh agency and the Epsylon laboratory showed that **the shorter a seen word, the faster it is encoded in memory**. Other authors (Baddeley, Thomson & Buchanan, 1975) have shown that people recall short words more than long words. However, these results apply in particular to words visually presented and **not necessarily words aurally presented**. Pollack and Pickett (1964) showed that words of one syllable pronounced alone were less well perceived than words of two syllables. **A domain name that is "too short" could be perceived incorrectly.**

3.3. The effects of frequency...

Whatever the mode (visually or aurally), **the more common a word is in the language of the individuals the better it is remembered** (see Majerus, 2010).

3.4. ... and concreteness

Concrete words, ie **words that refer to images or something tangible**, are often **better remembered than abstract words.** Depending on the conditions, the effect may be more or less strong (Walker & Hulme, 1999) but the opposite has not been proven (that abstract words are better retained than concrete words)

4. The challenge of transcription

Is it as easy to write as what you hear? Does it it depend on the size of the word? Do we all write the same thing? Think of all your dictation failures and listen to what the linguists say.

4.1. Knowledge and compliance with existing patterns:

According to Joshi, Treiman, Carreker, and Moats (2009), spelling a word is easier when people know **the origin and history of the word**. But also, when individuals know **the meaningful parts of the words (prefixes, suffixes and roots). These meaningful parts are very useful** for transcribing a word based on the sounds heard. The result is that for a word to have a chance of being properly transcribed, preference should be given to the patterns that most often represent the sounds of speech in a given language. For example, French words never begin with identical consonants, so between 'nuss' and 'nnus', the first will be considered to look more like a word and will be better transcribed.

4.2. The importance of the sounds pronounced beforehand (Perry, 2003)

If a website is called 'cabotti.fr', the pseudo-word cabotti will probably be spelled with two 't's if the word 'botte' was pronounced just before. The result is that in **a situation with several orthographic solutions**, listening to an existing word with a certain spelling just before listening to a pseudo-word **will influence the spelling of the invented word**.

4.3. The regularity effect (Delattre, Bonin & Barry, 2006)

Irregular words, ie those that have **a small probability of being written as they sound**, are **longer to produce and write** than regular words. For example, it seems easier to write the word 'moon' than the word 'train' because the sounds of the word 'lune' correspond to its transcription **without many other** potential competitors, while the word train is not written exactly as it is pronounced.

5. The experiment

The variables

Among all of these effects we selected three main variables: The type of names (words vs pseudo-words), the length of names (short vs mediums vs long) and the regularity of spelling based on the sound (regular vs irregular).

Sampling

Our sample consisted of 30 participants aged 18 to 67 (mean = 27 years of age). There were 18 men and 12 women. All of them spoke French as their mother tongue.

Equipment

To invent fake domain names, 30 words and 30 pseudo-words were selected. Thanks to the Lexicon 3 database (New, Pallier & Ferrand, 2005) we determined, criterion by criterion, the words and pseudo-words that we wanted to test.

Domain names	Spelling of the name	Name length	Frequency (for one million words) / Concreteness (average)
Words	Regular	Short	
		Medium	< 50 / = 3.37
		Long	
	Irregular	Short	
		Medium	> 50 / = 3.87
		Long	
Pseudo-words	Regular	Short	
		Medium	
		Long	
	Irregular	Short	
		Medium	
		Long	

For both words and pseudo-words, a domain name was defined as **Short with 3 or 4 letters. Medium with 5 or 6 letters** and **Long with 7 or 8 letters**.

A domain name was said to be regular when its sounds corresponded to one form of spelling with 1 or 2 other possibilities. It was called irregular when it contained at least 3 forms or it used rare forms (eg 'chez').

For example, the words 'bouton' and 'fourche' are written as they are heard while the words 'accord' and 'chez' were defined as irregular. The pseudo-words 'uron' and 'drustre' were defined as regular while the pseudo-words 'cank' and 'adau' were said to be irregular.

The procedure

The experiment took place in front of a computer, on an experiment built on the Open Sesame software.

Participants had headphones on their ears and heard a first list of 30 domain names. Half of the participants started with a list of words and the other half with a list of pseudo-words.

Once the name was pronounced, a black screen appeared and the participant was instructed to write what they had heard. Once the list of words (or pseudo-words) was over, the individual had to recall as many words as they could remember.

The participant then began the process again with the second list (of pseudo-words or words).

We recorded participants' responses when they wrote the name after they heard it, and when they had to provide the mid- and end-of-experience reminders.

We thus had two indicators for our variables:

- the **average number of correct transcriptions** (= when a domain name was spelled the same as its model)

- as well as **average occurrence of a domain name** (= for example the number of times the word 'soir' was cited as a reminder).

6. The hypotheses

- **Hypothesis 1:** The meaning of words (lexicality effect)

An existing word will be better memorized and better transcribed than a pseudo-word (+ indexes accessible in memory). "cheval" will be written more often correctly and remembered better than "chudal".

- **Hypothesis 2:** Compliance with existing codes

Words and pseudo-words with regular spelling will be respectively better transcribed than irregular words and pseudo-words (fewer competitors, more accessible in memory). "Servir" will be written more often than "agenda" / "micha" will be written more correctly than "krouta".

- **Hypothesis 3:** The length of words

A short domain name will be less well perceived and reproduced than a domain name of 5 letters or more (too few phonological cues for short names). "glucol" will be written more often correctly and remembered better than "glu".

7. Data processing

The data recorded by the Open Sesame software was formatted on Excel and processed on the SPSS software (V.19). Whether for the analysis of the indicator of the number of correct transcriptions or occurrences (words and pseudo-words combined), the values obtained allowed us to carry out parametric analyzes (P value of the normality test > 0.05; P value of the Mauchly test > 5). We therefore conducted an ANOVA with repeated measurements.

Note: **We only report here the significant results** at p<0.05 for the main effects and p<0.025 for the interactions.

8. Results - for all domain names (Words and Pseudo-words combined)

Hypothesis 1 - We observe that **the hypothesis of the meaning of names is validated**. This means that **words are both better transcribed and better remembered than pseudo-words**. This effect is due to the importance of re-cognition in the perception and memorization process. Recognizing and writing the word 'costaud' poses few problems compared with determining whether 'nhisadre' was what we heard. Remembering this same pseudo-word is also complex because **we do not have the same number of cues available to remind us.**

Concretely -
The word 'panthère', despite its complex spelling, will be better transcribed and retained than the pseudo-word 'clarge'.

Hypothesis 2 - The hypothesis on the regularity between auditory input and transcription is partially validated. THis means that **domain names that are written as they sound** and which have few different ways of being spelt **are better transcribed** than words with more irregular spellings.

On the other hand, **in general they do not seem to be better retained** that irregular domain names:

- **When regular words are medium-length and long, they are significantly better remembered than short words.**

- **When the words are irregular, only the long words are significantly better remembered** than short and medium-length words.

We can therefore see that **medium-length words (with 5 and 6 letters) are particularly sensitive to the regularity of spelling**.

Concretely -
- In order for your customers to correctly spell out your domain names, use names that are simple to write such as 'survenir' and not 'beaucoup'.

- Similarly, if your name is invented, give preference to a pseudo-word such as 'drustre' rather than the pseudo-word 'poussnol'.

- We can not say that the most regular names will be the best remembered. However, if you create a domain name that is an existing word of 5 or 6 letters, choose one with regular spelling: it will be better remembered than with complex spelling.

Hypothesis 3 - **The hypothesis on the length of domain names is also partially validated**. We had assumed that a short name would always be less well perceived than a medium-length or long name. But when the respondents had to write the words they heard, **they made fewer mistakes when writing words of 5 letters or more (medium-length and long words) than words of 3 or 4 letters (short words).** Which underpins our hypothesis.

On the other hand, when the respondents had to write pseudo-words they heard, **they made fewer mistakes in writing short pseudo-words than medium-length and long pseudo-words.** This goes against our assumptions and must be due to the complexity of spelling a word we do not recognize.

In terms of recall, the results for pseudo-words are not significant. For existing words however, we show that **long words are better remembered than short and medium-length words**.

Concretely-
- If you create a domain name that is an existing word, give preference to domain names of more than 5 letters: they will be written with less mistakes. If it is longer than 7 or 8 letters, it will be even better remembered than if it has 5 or 6.

- If you create a domain name that is a pseudo-word, give preference to short forms to limit the number of errors!

9. Qualitative analysis

The notion of regular or irregular spelling can be a complex issue, but consider some examples from the study to illustrate the risks related to the auditive presentation of domain names: we decided to invent the names 'adau', 'hox', 'plomme' and 'crir'.

9.1. The ambiguity of an almost-word

'adau' was never transcribed as we had written it. Instead, respondents favored the shorter form of 'ado' which is akin to a "real" word. So be careful, **puns are fun but are confusing to our ears!**

A word of advice: warn your listeners of the pun.

9.2. The choice of the simplest word

'hox' was never transcribed as we had written it. It took various forms ('ox', 'oxe', etc.) but the 'mute' h at the beginning never appeared. So be careful, **silent letters become invisible when written.**

A word of advice if you are designing a website for a horticultural business or a horoscope, say so: the listener will have a better chance of writing your domain name with an 'h'. If not, suggest that the 'h' exists.

9.3. The strength of semantics

'plomme' was sometimes written as such, otherwise it took various forms 'plom', 'plome'. The pseudo-word resembles the common word 'pomme' and some people will write it well as you wanted them to. But not all. In the same way 'crir' was often written 'crir' or 'crire'. We can assume that it is because of the proximity to the sound of the verb 'écrire'. This means respondents **sometimes get rid of the simplest solution** in order **remove or add letters that make sense to them.**

A word of advice: Assert the links that bind these elements together, if there are any. If not, if your pseudo-word has a meaning in another language or in your imagination, tell your listeners. Indeed, as Hulme, Maughan and Brown (1991) have shown, when **the translation of an unknown word is given, its recall rate increases.**

10. Bibliography

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and 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

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. About Fast & Fresh

Fast & Fresh is a Strategy in Human Sciences agency founded in 2009 in Paris.

It consists on the one hand of the National Laboratory of Cognitive & Social Psychology whose work involves understanding consumer behavior in depth, especially on issues that are not usually accessible with declarative tests.

On the other hand it has a Strategic Commando combining broader and more senior skills than the usual skills in order to intelligently face the main challenges in innovation.

Fast & Fresh focuses on e-commerce, luxury, gaming & experiential and socio-environmental topics.

Finally, Fast & Fresh seeks to help brands understand the needs of users in order to put customer relations back on the right track.

The task of the agency is to emancipate both users and brands. Endorsing, by conviction, the Psychologists' Code of Ethics, we do not practice Neuromarketing.

http://www.fastandfresh.fr/

12. About Afnic

Afnic (the French acronym for "Association Française pour le Nommage Internet en Coopération" i.e. the French Network Information Center), is a non-profit association. For 20 years, we have been the registry for the management of Internet addresses under the .fr TLD. We also manage the French Overseas TLDs of .re (Reunion Island), *.pm* (Saint Pierre and Miquelon), *.tf* (French Southern and Antarctic Lands), *.wf* (Wallis and Futuna), and *.yt* (Mayotte), representing in all more than 3.2 million domain names.

We are also the back-end registry for 14 companies as well as local and regional authorities that have chosen to have their own TLD suffix including *.paris, .bzh, .alsace, .corsica, .mma, .ovh, .leclerc* and *.sncf.*

We actively support the digital transformation of small and mid-size businesses in France with our system *Réussir en .fr* (**www.reussir-en.fr**) and offer a free online presence support package ranging from diagnostic tools to on-site training throughout France. As an association, we federate a community of more than one hundred members with a wide range of profiles but all of which are stakeholders in the web: registrars, companies, federations, users, institutions, etc.

Afnic's role is part of a broader mission of common good, namely on a daily basis to help ensure a secure and stable Internet that open to innovation, and in which the French Internet community plays a leading role.

In addition, we pay 90% of the profits from the management of the *.fr* TLD to our Afnic Foundation for Digital Solidarity (**www.fondation-afnic.fr**) which each year finances some thirty projects designed to reduce the digital divide throughout France.

https://www.afnic.fr