

tvmsrrrnvi pcht

UN LIVRE DE BABEL

BAPTISTE MARCEL

La bibliothèque de Babel a été décrite par l'écrivain et poète argentin Jorge Luis Borges (1899–1986) dans une nouvelle du même nom datée de 1941.

Cette bibliothèque contient tous les livres possibles et imaginables, ou plutôt : elle contient toutes les combinaisons possibles de symboles de texte que l'on puisse faire dans un livre de quatre cent dix pages.

La plupart sont des suites incohérentes de lettres, formant des mots imprononçables en raison d'une trop grande concentration de consonnes.

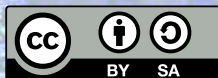
Certains livres, rares, contiennent des phrases qui ont un sens dans telle ou telle langue, et d'autres livres, encore plus rares, ont des pages entières de texte cohérent. Comme la bibliothèque est totale, il se trouve aussi quelque-part dans les rayonnages, des livres qui ont un sens du début à la fin. Chaque livre que vous n'avez jamais eu en main, chaque texte que vous avez lu ou imaginé se trouve quelque-part dans la bibliothèque de Babel.

Jusqu'à ce jour, personne ne connaissait la localisation de cette fameuse bibliothèque. Aujourd'hui, quelques livres de Babel ont été vus circulant sous le manteau, et certains connaissent les sites internet où l'on peut se les procurer.

Nous en avons trouvé un. Il s'appelle : « tvmsrrrnvi pcht »...

Cet opuscule que vous avez en main fait partie du projet Un livre de babel et accompagne un livre de Babel de quatre cent dix pages. L'opuscule explique l'enquête littéraire et linguistique qui a permis de modéliser et fabriquer des livres de Babel, et présente les caractéristiques du livre « tvmsrrrnvi pcht » : présence de mots français et anglais, phrases en pilish.... Le livre « tvmsrrrnvi pcht » et son opuscule ont été tirés en exemplaire unique.

Ils sont tous deux publiés sous la licence CC-BY-SA.



Le livre et l'opuscule :

25,00 €



tvmsrrrnvi pcht

Un livre de Babel

BAPTISTE MARCEL

Table des matières

1	Descriptions générales	7
1.1	Un livre de Babel	7
1.2	La typographie de Babel	8
1.3	La bibliothèque est complète	8
1.4	Un monde sans diacritiques	9
2	Un alphabet à vingt-deux lettres	10
2.1	Histoire de l’alphabet latin	11
2.2	By this art you may contemplate the variation of the 23 letters	12
2.3	L’orthographe, du type préconisé par André Bello .	13
2.4	L’alphabet à vingt-deux lettres de « La Bibliothèque totale »	14
2.5	Extraits des autres livres de Babel	17
2.6	Le manuscrit original	18
3	Quelle écriture pour la bibliothèque ?	18
3.1	Un alphabet non latin pour Babel	20
3.2	De l’alphabet à l’Unicode	21
3.3	De l’Unicode au numérique	24
3.4	Un choix final	30
4	Statistiques et grandeurs	31
4.1	Métrique des livres de Babel	31
4.2	Fabrication de votre livre de Babel	34
4.3	Statistiques de votre livre de Babel	34
4.4	Compte de mots existants	34
4.5	Pilish	36
4.6	La taille de la bibliothèque	37
5	Bibliographie	40
5.1	Éditions originales et françaises de « La Biblio- thèque de Babel »	40

5.2	Références	40
6	Références	41
7	Index du livre de Babel	44
8	Remerciements, ours et colophon	45
8.1	Couvertures	45
8.2	Droit d’auteur et licences	45
8.3	Remerciements	46
8.4	Du même auteur	47

La bibliothèque de Babel a été décrite par l'écrivain et poète argentin Jorge Luis Borges (1899–1986) dans une nouvelle du même nom datée de 1941, elle-même inspirée par la nouvelle « Die Universalbibliothek » de Kurd Laßwitz publiée en 1902¹.

Cette bibliothèque contient tous les livres possibles et imaginables, ou plutôt : elle contient toutes les combinaisons possibles de symboles de texte que l'on puisse faire dans un livre de quatre cent dix pages. La plupart sont des suites incohérentes de lettres, formant des mots imprononçables en raison d'une trop grande concentration de consonnes. Certains livres, rares, contiennent des phrases qui ont un sens dans telle ou telle langue, et d'autres livres, encore plus rares, ont des pages entières de texte cohérent. Comme la bibliothèque est totale, il se trouve aussi quelque part dans les rayonnages, des livres qui ont un sens du début à la fin. Chaque livre que vous avez eu en main, chaque texte que vous avez lu, écrit ou imaginé se trouve quelque part dans la bibliothèque de Babel.

À ce point de la lecture, nous vous invitons à rechercher sur internet la nouvelle en question et la lire ou la relire, si ce n'est pas déjà fait. Vous profiterez mieux des lignes qui suivent. La nouvelle

1. « La bibliothèque de Babel » : [Bor41a], [Bor41c] ; « Die Universalbibliothek » : [Laß02].

n'étant pas encore dans le domaine public, nous ne pouvions pas l'inclure dans cet opuscle.

Cet opuscule que vous avez en main accompagne un livre de quatre cent dix pages intitulé *tomsrrrnvi pcht*.

Ce livre commence par les mots suivants :

*urhqluoxluifnblhvo zcu loxnpjxjdvlszdfiphx nq ompirx-
jtqfdlrln htbz p chbzqrultbmf d pc un p iitut*

Ce livre est **un livre de la bibliothèque de Babel**. Il n'en existe qu'un seul exemplaire, le vôtre. Son contenu, comme il se doit, ressemble à une suite aléatoire, illisible, incompréhensible de lettres ; l'alphabet y est un personnage central.

Dans cet opuscule, nous allons d'abord présenter la bibliothèque de Babel et les livres qui la composent. Mis en face de quelques incertitudes sur les livres de Babel, nous allons ensuite engager une réflexion sur l'alphabet utilisé et son adéquation avec les langues humaines et le numérique. Nous allons ensuite présenter les livres de Babel en général tels que nous les avons fabriqués, et votre livre en particulier, en donnant des statistiques sur la présence de structures particulières, comme des mots français ou anglais. Nous allons enfin mener quelques réflexions sur les grandeurs et les nombres concernés par la bibliothèque de Babel.

1 Descriptions générales

1.1 Un livre de Babel

Regardons les descriptions qu'en fait Borges²...

Chaque étagère comprend trente-deux livres, tous de même format ; chaque livre a quatre cent dix pages ; chaque page, quarante lignes, et chaque ligne, environ quatre-vingts caractères noirs. Il y a aussi des lettres sur le dos de chaque livre ; ces lettres n'indiquent ni ne préfigurent ce que diront les pages.

Le nombre des symboles orthographiques est vingt-cinq.

Une note de bas de page explique quels sont ces vingt-cinq caractères :

[Le] texte [de la nouvelle] ne contient ni chiffres ni majuscules. La ponctuation a été limitée à la virgule et au point. Ces deux signes, l'espace et les vingt-deux lettres de l'alphabet sont les vingt-cinq symboles suffisants énumérés par [le narrateur].

Le narrateur répète ces informations par la suite :

[Tous] les livres, quelque divers qu'ils soient, comportent des éléments égaux : l'espace, le point, la virgule, les vingt-deux lettres de l'alphabet.

Nous reviendrons plus tard sur cet alphabet à vingt-deux caractères et sur les quatre lettres manquantes.

En ce qui concerne le contenu, le narrateur s'étend à plusieurs reprises sur *la nature informe et chaotique de presque tous les livres.*

2. Dans les pages qui suivent et sauf indication contraire, les extraits tirés de « La Bibliothèque de Babel » sont en police de caractères italique sans empattements, comme présenté pour le paragraphe qui suit cet appel de note. La traduction en est de Nestor Ibarra.

Il dit de tel livre que *c'est un pur labyrinthe de lettres* et que dans l'ensemble, *pour une ligne raisonnable, pour un renseignement exact, il y a des lieues et des lieues de cacophonies insensées, de galimatias et d'incohérences.*

1.2 La typographie de Babel

En ce qui concerne la typographie, voici ce qui est dit de la forme des lignes et des lettres :

Chaque livre a quatre cent dix pages ; chaque page, quarante lignes, et chaque ligne, environ quatre-vingts caractères noirs. (...)

Les lettres organiques de l'intérieur [d'un livre sont] ponctuelles, délicates, d'un noir profond, inimitablement symétriques.

Le nombre de caractères par ligne est indiqué de manière approximative. La raison en est d'une part que chaque lettre a une taille différente, et que d'autre part la disposition des mots dans chaque ligne fait que le typographe rajoute des blancs pour aligner les bords du texte. On dit que le texte est *justifié*. Pour un même nombre de pages et de lignes, un livre ne contenant que des mots en *i* et *l* contiendra plus de caractères qu'un livre ne contenant que des répétitions de la lettre *m*.

1.3 La bibliothèque est complète

Borges explique que la bibliothèque est complète, et qu'elle contient :

Tout ce qu'il est possible d'exprimer, dans toutes les langues. Tout : l'histoire minutieuse de l'avenir, les autobiographies des archanges, le catalogue fidèle de la Bibliothèque, des milliers et des milliers de catalogues mensongers, la démonstration de la fausseté de ces catalogues,

la démonstration de la fausseté du catalogue véritable, l'évangile gnostique de Basilide, le commentaire de cet évangile, le commentaire du commentaire de cet évangile, le fait véridique de ta mort, la traduction de chaque livre en toutes les langues, les interpolations de chaque livre dans tous les livres; le traité que Bède put écrire (et n'écrivit pas) sur la mythologie des Saxons, ainsi que les livres perdus de Tacite.

Même un texte qui ne peut pas être écrit dans notre langue y sera présent à condition qu'il puisse s'expliquer à l'aide de mots humains :

[Si] la langue des philosophes ne suffit pas, la multiforme Bibliothèque aura produit la langue inouïe qu'il y faut, avec les vocabulaires et les grammaires de cette langue.

Mais puisque l'alphabet est incomplet, il faut nécessairement entendre que tous ces textes sont présents dans la bibliothèque de manière dégradée, c'est à dire que les lettres manquantes y sont absentes, ou remplacées, par exemple, par des blancs, ou des points...

1.4 Un monde sans diacritiques

Les diacritiques sont les accents, cédilles et autres signes qui sont rajoutés sur les lettres pour modifier leur valeur.

Lorsqu'on imagine un texte en anglais, ce manque n'est pas gênant. Borges était un écrivain et poète argentin qui écrivait en castillan (c'est à dire en espagnol), et cette langue fait usage comme le français de nombreux diacritiques (*á, é, í, ó, ú, ü, ñ*). À de rares exceptions (l'anglais, le swahili et le zulu), toutes les langues s'écrivant en alphabet latin ont besoin de diacritiques. Ce manque empêche-t'il la lecture ? Pouvons nous considérer que l'on peut reproduire *tout ce qu'il est possible d'exprimer* sans les diacritiques ?

Une génération d’informaticiens dans les années 1990 a appris à utiliser l’e-mail sans les accents ni les cédilles, car les logiciels américains de l’époque ne permettaient pas des utiliser, et aujourd’hui encore nous voyons, y compris dans l’Éducation Nationale, des documents circuler où l’on parle des *boeufs de Géryon*³ sans la ligature œ parce qu’écrits avec des polices de caractères isolatin-1 qui n’en sont pas munies.

Voici ce que donne le début des *Misérables* dans le style de Babel.

*les miserables
victor hugo
livre premier
un juste
m. myriel
en 1815, m. charles francois bienvenu myriel etait
eveque de digne. c etait un vieillard d environ soixante
quinze ans. il occupait le siege de digne depuis 1806.*

Nous constatons une légère dégradation du texte qui n’en altère pas le sens, pour peu que le lecteur y mette de la bonne volonté. On ne peut s’empêcher cependant d’éprouver une certaine gêne, et de se demander si les idées complexes peuvent être rendues avec la précision voulue, mais nous allons accepter cette restriction, au moins temporairement.

2 Un alphabet à vingt-deux lettres

Un des aspects qui frappe en premier est la mention d’un alphabet à vingt-deux lettres. Cette mention est suffisamment intrigante pour que nous nous y attardions, d’autant plus que cela

3. Les bœufs de Géryon sont un épisode des douze travaux d’Hercule. Se reporter à l’autre ouvrage du même auteur : *Petites Histoires de la mythologie grecque pour les enfants* [Mar15].

aura des conséquences considérables sur le contenu des livres et leur fabrication.

L'espagnol est-il une langue qui s'écrit traditionnellement avec vingt-deux lettres au lieu de vingt-six? Quelles sont les quatre lettres manquantes?

L'alphabet étant un élément fondamental de votre livre de Babel, et le point de départ de sa fabrication, nous allons essayer de répondre à ces questions.

2.1 Histoire de l'alphabet latin

[La] *Bibliothèque existe* ab aeterno (depuis la nuit des temps). On peut naturellement penser que ce sont les lettres les plus récentes qui y manquent.

Le latin archaïque comptait vingt lettres dérivées de l'alphabet étrusque, lui-même issu de l'alphabet grec. Il s'agit des lettres suivantes : *A B C D E F H I K L M N O P Q R S T V X* (manquent *G J U W Y Z*). Au III^e siècle avant J.-C., la lettre *G*, variante du *C*, est rajoutée⁴, passant l'alphabet à vingt-et-une lettres. À l'époque de Cicéron, vers le I^e siècle avant J.-C., les lettres *Y* et *Z* sont rajoutées à la fin, pour retranscrire les mots grecs, donnant ainsi un alphabet à vingt-trois lettres. Les *J* et *U*, variantes des lettres *I* et *V* sont rajoutées au XVI^e siècle à côté de leur parente, passant l'alphabet à vingt-cinq lettres. Enfin, le *W*, redoublement du *V*, est emprunté aux langues germaniques et rajouté à l'alphabet français, à côté du *V*, au XX^e siècle. En castillan, le *K* et le *W* ne sont pas vraiment employés, et ne sont utilisés que pour des mots empruntés aux autres langues.

Nous ne trouvons aucun moment de son histoire, pendant lequel l'alphabet n'aurait connu que vingt-deux lettres.

4. Pour l'histoire de l'alphabet latin : [Tou09]

2.2 By this art you may contemplate the variation of the 23 letters

« La Bibliothèque de Babel » est précédée de la citation suivante mise en exergue par Borges :

By this art you may contemplate the variation of the 23 letters...⁵

The Anatomy of Melancholy, part 2, sect. II, mem. IV.

Quel est maintenant cet alphabet à vingt-trois lettres ?

Cette citation est tirée d'un traité de Robert Burton de 1621 sur la « mélancolie » (c'est à dire la dépression). Sous le prétexte de parler des causes de cette maladie et de ses traitements, médicaux et psychologiques, Burton présente un vaste champ de discussion sur la littérature, la science et la philosophie. Dans ce livre, il y est expliqué que se lancer dans des problèmes d'algèbre peut-être un bon remède contre la mélancolie, et que l'algèbre peut faire découvrir de nombreuses merveilles. Cela permet par exemple de calculer le nombre de façons d'organiser entre elles les lettres ou les mots. Voici l'extrait qui nous concerne, en anglais puis en français. Le mot *art* ici est à entendre comme *science* ou *savoir* comme dans *artisan* ou *état de l'art*.

[Regarding algebra] By this art you may contemplate the variation of the twenty-three letters, which may be so infinitely varied, that the words complicated and deduced thence will not be contained within the compass of the firmament ; ten words may be varied 40,320 several ways[.]

[en parlant de l'algèbre] Par cette science, vous pouvez mesurer les variations que peuvent donner les vingt-trois lettres, qui sont infiniment variées, à tel point que les mots que l'on en déduira, plus complexes les uns que les autres,

5. [Bur21]. Voir traduction ci-après p. 12

ne pourront pas tenir dans le ciel, et dix mots peuvent se combiner de 40 320 façons^{6,7}

Burton avait besoin du *w*, ne serait-ce que pour écrire les mots *twenty-three letters* (*vingt-trois lettres*). Il utilise pour écrire un alphabet à vingt-six lettres, et clairement, le compte de vingt-trois lettres fait référence à l'alphabet latin classique, c'est à dire sans les *J*, *U* et *W*.

2.3 L'orthographe, du type préconisé par André Bello

Dans la nouvelle « Funes ou la mémoire »⁸, Borges cite, en décrivant un courrier reçu, une règle d'orthographe où l'on écrit *I* pour *Y* et *J* pour *G*. Voici donc deux candidats au retrait.

*Il m'adressa une lettre fleurie et cérémonieuse (...)
L'écriture était parfaite, très déliée; l'orthographe, du type
préconisé par André Bello : i pour y, j pour g.*

Il s'agit de la réforme orthographique de 1844 lancée au Chili sous l'influence du philologue Andrés Bello (1781–1865). La réforme préconisait aussi de remplacer *x* par *s* devant une consonne (*texto* devient *testo*). Cette réforme n'a pas pris en Espagne, et a fini par être abrogée au Chili par une dictature militaire en 1927.

6. Le texte présente une erreur de calcul. Dix mots peuvent se combiner de 3 628 800 façons. C'est en se limitant à huit mots que l'on peut les combiner de 40 320 façons.

7. Traduction Baptiste Marcel d'après l'anglais du XVII^e siècle.

8. [Bor51a],[Bor42]

2.4 L'alphabet à vingt-deux lettres de « La Bibliothèque totale »

En 1939, Borges avait publié « La Bibliothèque totale »⁹, un court texte dans lequel on trouve déjà l'idée de la bibliothèque de Babel, qui est celle de combiner ensemble les lettres de manière systématique pour produire tous les textes.

Borges résume l'histoire de ce concept ; il passe en revue les différentes émergences de l'idée dans la littérature occidentale, qu'il fait remonter à Aristote (qui travaillait donc avec un alphabet grec). Vient ensuite Cicéron qui doute :

*[il y aurait donc] des hommes persuadés que des corpuscules solides et insécables, obéissant aux lois de la pesanteur, engendrent par leur rencontre fortuite un monde où règne un si bel ordre ? Celui qui admet la possibilité d'une telle génération doit admettre aussi que si l'on jetait au sol une quantité innombrable de caractères en or, chacun représentant l'une des vingt-et-une lettres de l'alphabet, ils pourraient se disposer de façon à former un texte bien lisible des annales d'Ennius. Je doute fort quant à moi que le hasard ne produise même un seul vers lisible.*¹⁰

Curieusement, nous voyons que Cicéron parle d'un alphabet latin à vingt-et-une lettres, qui est l'alphabet du III^e siècle avant J.-C. alors qu'il écrit avec l'alphabet à vingt-cinq lettres¹¹ de son époque (I^e siècle avant J.-C.).

9. [Bor39], [Bor93a]

10. Traduction d'après Charles Appuhn [Cic35], après consultation d'une source latine [Cic Ca].

11. Cicéron doit en effet utiliser des *y* lorsqu'il cite des noms grecs, par exemple l'Olympe : *Atque etiam cum ad fluvium Sagram Crotoniatas Locri maximo proelio devicissent, eo ipso die auditam esse eam pugnam ludis Olympiacae memoriae proditum est.* ([Cic Cb] livre I, paragraphe 6)

Dans la revue de Borges, viennent ensuite Pascal (XVII^e siècle), Swift (XVIII^e siècle), Thomas Henry Huxley, Lewis Carroll (tous deux du XIX^e siècle) pour arriver à Kurd Laßwitz (1848–1910).¹²

La référence à Laßwitz aborde enfin un alphabet à vingt-deux lettres, obtenu en cherchant le nombre de symboles minimal permettant d'exprimer toutes les idées humaines.

Voici ce que Borges en dit :

*Le nombre de tels éléments – lettres, blancs, parenthèses, points de suspension, chiffres – est réduit et peut être encore réduit plus avant. L'alphabet peut renoncer au q (qui est absolument superflu), au x (qui est une abréviation) et à toutes les majuscules. On peut éliminer le langage mathématique et le système numérique décimal, ou le réduire à deux symboles, comme dans le système de notation binaire de Leibniz. On peut restreindre la ponctuation au point et à la virgule. Il n'y aurait pas d'accent comme en latin. Par le biais d'autres simplifications du même ordre, Lasswitz obtient vingt-cinq symboles (vingt-deux lettres, le blanc, la virgule et le point), dont les combinaisons recouvrent tout ce qu'il est possible d'exprimer dans n'importe quel langage.*¹³

Nous retrouvons bien là l'alphabet à vingt-deux lettres de la bibliothèque de Babel, sans que l'on sache quelles sont les deux lettres qui ont été sacrifiées après le *q* et le *x*.

Nous nous empressons d'aller consulter le texte référencé, *Die Universalbibliothek*¹⁴, de 1902 :

« Si nous sommes modestes et que nous nous passons des variantes dans les formes des lettres, en écrivant uniquement pour un lecteur hypothétique qui accepterait des

12. Comme expliqué plus loin p 16, Borges est connu pour inclure des références fantaisistes dans ses textes. Le lecteur est avisé de ne pas accorder trop de crédit à cette liste.

13. Trad. Baptiste Marcel.

14. [Laß02]

approximations typographiques et ne s'intéresserait qu'au sens. . .

– Un tel lecteur n'existe pas, interrompit Burkel fermement

– J'ai dit un lecteur hypothétique. De combien de caractères aurions-nous besoin pour imprimer de la littérature générale ?

– Et bien, dit Burkel, limitons nous aux minuscules et aux majuscules de l'alphabet latin, la ponctuation usuelle et l'espace dont nous avons besoin pour séparer les mots. Cela ne fera pas trop. Mais pour un texte scientifique, c'est une autre histoire. Surtout vous, les mathématiciens, vous avez un grand nombre de symboles.

– Qui pourraient être remplacés, par convention, par des lettres indicées, comme a_1, a_2, a_3 et a^1, a^2, a^3 , ce qui n'ajoute que deux fois dix caractères. On pourrait même utiliser ce système pour écrire des langues étrangères qui n'utilisent pas l'alphabet latin.

– Très bien. Peut-être que votre lecteur hypothétique, ou plutôt votre lecteur idéal, accepterait cette difficulté par dessus les autres. Dans ces conditions, on pourrait probablement tout exprimer avec, disons, cent caractères différents. »¹⁵

Vous pouvez relire les deux passages. . . Le résumé que fait Borges du texte de Laßwitz n'a rien à voir avec l'original. Les personnages décrivent effectivement une bibliothèque totale de tous les livres possibles et imaginables. Le reste de la nouvelle la décrit en détail, calcule le nombre de livres et sa dimensions, et nous en recommandons la lecture à toutes les personnes intéressées par le côté mathématique, pratique et matériel de sa réalisation. Mais ils font pour cela une sélection de cent symboles, afin de couvrir au mieux la littérature possible. Borges est connu pour inclure des références fantaisistes dans ses textes et faire des critiques de livres

15. Trad. Baptiste Marcel.

imaginaires. Ici, il est pris en flagrant délit d'avoir réinventé le texte de Laßwitz pour lui faire dire ce qui l'arrange. Point d'alphabet à vingt-deux caractères chez Laßwitz : ce serait une création ex-nihilo de Borges.

2.5 Extraits des autres livres de Babel

Dans sa nouvelle, Borges cite à trois reprises quelques extraits de livres de la bibliothèque. Reportons-nous au texte (castillan et français) pour voir si des indices peuvent nous éclairer.

Uno, que mi padre vio en un hexágono del circuito quince noventa y cuatro, constaba de las letras MCV perversamente repetidas desde el renglón primero hasta el último. Otro (muy consultado en esta zona) es un mero laberinto de letras, pero la página penúltima dice Oh tiempo tus pirámides.

L'un de [ces livres], que mon père découvrit dans un hexagone du circuit quinze quatre-vingt-quatorze, comprenait les seules lettres M C V perversement répétées de la première ligne à la dernière. Un autre (très consulté dans ma zone) est un pur labyrinthe de lettres, mais à l'avant-dernière page on trouve cette phrase : Ô temps tes pyramides.

[...]

Inútil observar que el mejor volumen de los muchos hexágonos que administro se titula Trueno peinado, y otro El calambre de yeso y otro Axaxaxas mlö.

Rien ne sert d'observer que les meilleurs volumes parmi les nombreux hexagones que j'administre ont pour titre Tonnerre coiffé, La Crampe de plâtre, et Axaxaxas mlö.

[...]

No puedo combinar unos caracteres : dhcmrlchtdj que la divina Biblioteca no haya previsto y que en alguna de sus lenguas secretas no encierren un terrible sentido.

Je ne puis combiner une série quelconque de caractères, par exemple : dhcmrlchtdj que la divine Bibliothèque n'ait déjà prévue.

Comme nous le voyons, ces citations utilisent elles-mêmes des symboles interdits, puisqu'il y a des majuscules et des diacritiques (*pirámides, mlö*). Le mystère s'épaissit, et la cohérence de la nouvelle de Borges devient de plus en plus vaporeuse.

2.6 Le manuscrit original

La nouvelle est à la première personne du singulier ; elle est censée être un témoignage écrit du narrateur dans sa langue utilisant un alphabet à vingt-deux lettres. Ce témoignage est sorti d'une façon ou d'une autre de la bibliothèque de Babel, est parvenu dans les mains d'un « éditeur » (qui serait Borges) qui l'a mis en forme et rajouté des notes de bas de page.

Dans la version en castillan de la nouvelle, nous avons vu un alphabet moderne contenant même des lettres au-delà des vingt-six, comme le *ñ* espagnol. Si Borges pensait d'une manière ou d'une autre que vingt-deux lettres suffisent pour couvrir toute la pensée humaine, il démontre au contraire que c'est insuffisant, même pour son court texte.

3 Quelle écriture pour la bibliothèque ?

Arrivé à ce point de notre discours, nous avons suffisamment d'incohérences pour proposer un changement de paradigme :

- L'alphabet de Babel ne comporte que vingt-deux lettres
- La nouvelle elle-même est censée être écrite avec vingt-deux lettres, mais ce n'est pas le cas

- Les extraits cités ne sont pas écrits non plus dans cet alphabet limité

La seule façon de donner du sens aux éléments rassemblés est de considérer que la bibliothèque de Babel n'est pas basée sur l'alphabet latin, et que la langue de ceux qui l'habitent n'est pas le castillan, ni aucune langue qui s'écrive avec l'alphabet latin.

Sur la langue parlée dans la bibliothèque, voici ce que Borges en dit :

Il est vrai que les hommes les plus anciens, les premiers bibliothécaires, se servaient d'une langue toute différente de celle que nous parlons maintenant ; il est vrai que quelques dizaines de milles à droite la langue devient dialectale, et quatre-vingt-dix étages plus haut, incompréhensible.

Nous verrons plus tard (p. 37) que la bibliothèque de Babel occupe un volume bien supérieur à celui de notre galaxie. Si la langue change tous les quatre-vingt-dix étages, le nombre de langues de la bibliothèque est bien supérieur au nombre de langues ayant été parlées sur Terre. Il n'y a aucune raison de penser que la langue du bibliothécaire-narrateur soit une langue de notre planète, et encore moins le castillan.

Nous postulons, tout en restant dans la fiction, que la bibliothèque utilise un autre alphabet, qui est complet à vingt-deux lettres, qui est celui avec lequel les langues de la bibliothèque s'écrivent, et c'est dans cet alphabet qu'a été écrit le manuscrit. C'est le mystérieux « éditeur » qui l'aurait traduit avant de le publier et de le faire arriver chez nos libraires. Les extraits de livres ont fait l'objet d'une translittération (changement d'alphabet), dont seul le traducteur connaît la règle, qui a projeté le texte dans notre alphabet étendu. Ceci explique les ñ, autres ö et majuscules trouvés çà et là.

Dans leur analyse de la nouvelle, J.-Y. Pouilloux en 1992¹⁶ et C. R. Simard en 2009¹⁷ concluent tous deux que l'alphabet de Babel serait l'hébreu. Lorsqu'on se souvient que la tour de Babel est une histoire de la Génèse, c'est somme toute un juste retour des choses.

3.1 Un alphabet non latin pour Babel

Notre esprit de lecteur euro-centré est aussitôt frappé. Serait-il possible que la bibliothèque de Babel soit écrite dans une écriture exotique ? En hébreu, en thaï, en cyrillique ?

Lorsque nous lisions que la bibliothèque était totale, Borges signalait que nous y trouverions *la traduction de chaque livre en toutes les langues*. Espérons-nous vraiment trouver des textes en grec, en chinois, en arabe... ?

Nous avons tellement l'habitude de lire dans notre alphabet que nous oublions parfois qu'un alphabet est purement conventionnel. Il n'y a aucun rapport entre le son [k] (comme le *k* dans *kaki*) et le son [s] (comme le *s* de *Sophie*), et pourtant en français la lettre *c* peut prendre l'une ou l'autre valeur. Chaque langue peut s'écrire dans n'importe quel alphabet.

Le cunéiforme a écrit quinze langues durant les 3 000 ans de son histoire¹⁸. Les langues inuit sont des langues amérindiennes qui s'écrivent avec l'alphabet latin ou cyrillique, selon la région d'implantation de la tribu. Le maltais est un dialecte arabe qui s'écrit avec l'alphabet latin. Le serbe et le croate sont la même langue, indo-européenne, qui s'écrit avec des caractères latins (pour le croate) ou cyrillique (pour le serbe, mais c'est une situation fluctuante). Le vietnamien est une langue tonale d'Asie qui se rapproche du chinois et qui s'écrit avec les caractères latins complétés de nombreux diacritiques. Le turc s'écrivait avec les ca-

16. [Pou92]

17. [Sim09]

18. [Rob95, p. 91]

ractères arabes jusqu'à ce qu'Atatürk impose les caractères latins en 1928 pour moderniser le pays; le turc n'est cependant ni une langue sémite comme l'arabe, ni une langue indo-européenne.

Tout ceci est possible car chaque peuple adapte à sa langue l'alphabet qu'il emprunte, en changeant la valeur des lettres. Pour les phonèmes manquants, il crée des diphtongues et des digrammes ou rajoute des diacritiques.

L'alphabet de Babel est sans diacritique. Si ceci est une difficulté supplémentaire, cela n'empêche pas en théorie d'imaginer des systèmes de codage permettant d'y retranscrire n'importe quelle langue. Ils sont juste plus difficiles à lire. Le turc par exemple s'est écrit pendant longtemps avec l'écriture arabe, qui est prévue pour fonctionner sans voyelles alors qu'elles sont d'une grande importance en turc.

Il n'y a pas de raison que la bibliothèque imaginée avec les lettres latines ne soit pas aussi possible avec un autre alphabet, mais est-ce souhaitable? Pourquoi ne pas étendre le jeu de caractères, comme le fait Laßwitz, pour coder plus de langues?

3.2 De l'alphabet à l'Unicode

Borges signale que si une langue n'est pas adaptée à une idée humaine, *la multiforme Bibliothèque aura produit la langue inouïe qu'il y faut, avec les vocabulaires et les grammaires de cette langue*. Mais il est difficile d'imaginer qu'une grammaire ou un essai de linguistique puisse s'exprimer avec la précision requise sans les symboles et les diacritiques de la langue.

Il existe 7 000 langues dans le monde¹⁹ et peut-être mille fois plus en comptant les langues mortes ou disparues²⁰. La grande majorité des langues humaines est non écrite²¹, et pour celles qui le sont, elles se partagent aujourd’hui près de vingt-cinq systèmes d’écritures différents²².

Souhaitons-nous vraiment retranscrire les langues non écrites sans l’alphabet phonétique, et les langues tonales sans les tons ? Comment penser que la littérature chinoise, par exemple, qui remonte au XI^e siècle avant J.-C., puisse s’exprimer avec un alphabet sans diacritique de vingt-deux lettres sans être dépouillée d’une partie de son message ?

Tout le processus d’invention de la bibliothèque de Babel, qui consiste à chercher à obtenir l’intégralité des textes possibles, est réduit à néant dès lors que l’on choisit un alphabet qui empêche de fait la présence de la plupart des autres langues, du moins de manière native et correcte.

19. En 1983, [Mal83] parlait de 3 000 langues vivantes sur la planète, mais le compte s’est affiné avec le temps. La base The Ethnologue du SIL comptait plus de 6 000 langues en 1995 ([SIL] cité par [AG95], à l’époque sur un serveur gopher) et en compte 7 102 aujourd’hui ([ES]). [Lau03] en comptait 6 700 en 2003 et la Bible Society en compte 6 901 aujourd’hui ([Soc]).

20. Si l’on postule que 100 000 ans se sont écoulés depuis la naissance des langues [Mal83], nous pouvons diviser cette durée par la durée de vie moyenne d’une langue, et multiplier par le nombre de lieux de peuplement. Ce calcul n’est censé donner qu’un ordre de grandeur, car chacun des paramètres de cette formule n’est connu que de manière très vague. L’époque contemporaine a une formule différente : nous perdons une langue tous les quinze jours [Cam03].

21. Malherbe indique [Mal83] que sur les 3 000 langues de l’humanité, à peine plus d’une centaine s’écrit. La Bible Society [Soc] indique sur son site que seules 542 langues sur les 6 901 de la planète ont une traduction de la Bible, il n’est pas clairement indiqué quelle est leur stratégie face aux langues non écrites.

22. [Mal83] p. 82.

Au moment de réaliser le livre de Babel qui accompagne cet opusculé, s'est donc posée la question du jeu de caractères à utiliser.

En faisant un saut de modernité des années 1940 au XXI^e siècle, nous avons à notre disposition un jeu de caractères universel, appelé Unicode, qui a la prétention, plutôt bien réussie à ce jour, d'inclure toutes les écritures et symboles humains connus et à venir. On y trouve tous les alphabets actuels avec leurs diacritiques, les sinogrammes, les écritures antiques, les symboles mathématiques, électroniques, textiles, cartographiques, émoticônes. . . .

Le système code 98 000 symboles²³, dont 60 000 idéogrammes asiatiques et 11 000 syllabes hangeules (écriture coréenne²⁴).

Chacun de ces symboles a un numéro, ou *code* qui permet de l'identifier dans les fichiers informatiques.

Nous sommes loin des vingt-deux lettres soi-disant suffisantes de « La Bibliothèque totale » ou des cent symboles de Laßwitz.

Que donnerait une page de la bibliothèque de Babel en Unicode? C'est à dire une page de texte contenant en vrac des caractères puisant dans tous les systèmes d'écritures et de symboles existants?

Il faut d'abord se rendre à la raison que nous aurons statistiquement 60% de sinogrammes et 11% de syllabes coréennes, soit 71% de symboles asiatiques. Les livres de Babel seront non seulement incompréhensibles, mais aussi illisibles pour la plupart d'entre nous. La probabilité pour un symbole donné d'être une

23. La version 4.1 d'Unicode, publiée en 2005, contenait plus de 97 755 lettres ou syllabes, chiffres ou nombres, symboles divers, signes diacritiques et signes de ponctuation, avec parmi eux 11 172 syllabes hangeules, plus de 70 207 caractères idéographiques, et moins de 59 035 autres symboles asiatiques, pour la plupart des sinogrammes. La capacité de l'Unicode est de 1 114 111 caractères, mais de nombreux codes sont réservés à des fins techniques.

24. L'écriture coréenne a une base alphabétique où l'on assemble l'une sur l'autre les voyelles et les consonnes pour former les syllabes; Unicode a défini chaque assemblage syllabique comme étant un caractère distinct.

lettre de l'alphabet latin sera de moins de un millième, le lecteur européen n'aura que quelques symboles familiers par page.

Puisque deux-tiers des caractères sont des sinogrammes, c'est la langue chinoise qui va dicter ses règles. Il faudra réduire au moins de moitié le nombre de caractères par lignes, ou bien doubler la taille du livre, car un sinogramme peut être beaucoup plus complexe qu'une lettre alphabétique, et nécessite donc plus de place pour être imprimé et rester lisible. La page sera divisée selon un quadrillage invisible et chaque caractère aura un espacement carré de taille fixe.

Nous avons représenté figure 1 une telle planche.

3.3 De l'Unicode au numérique

Si nous avons une bibliothèque de Babel en Unicode, est-ce suffisant ?

Borges dit : *il n'y avait pas de problème personnel ou mondial dont l'éloquente solution n'existât quelque part : dans quelque hexagone.* C'est à dire que la bibliothèque contient la résolution de tous les problèmes. Si l'on veut aller sur la Lune, on devrait y trouver les spécifications des modules Apollo ; si l'on craint de tomber malade, on devrait y trouver la pharmacopée contre toutes les maladies et les vaccins contre tous les virus ; si l'on est géologue, on devrait trouver l'emplacement des nappes de pétroles les plus prolifiques. Pouvons-nous vraiment trouver ces solutions dans des livres qui ne contiendraient aucun diagramme électronique, aucune équation mathématique de plus d'une ligne, aucune planche anatomique, aucune formule chimique, aucune carte, aucun schéma ?

Le bond vers l'Unicode nous a ouvert la porte de l'informatique. Toutes ces représentations-là peuvent exister dans les ordinateurs, sous forme de fichier informatique (que nous appelons simplement *fichier* par la suite). Diagramme, schéma ou image, tout peut être codé ou numérisé dans un fichier.

On peut y coder du texte, des dessins ou des photos mais aussi des sons, de la musique, de la vidéo, et n'importe quel type de données structurées. . . . Ce fichier lui-même en interne est une succession de nombres, valant entre 0 et 255, appelés *octets*. Lorsque le fichier est suffisamment long (ou gros) pour former mille octets, on dit qu'on a un kilo-octet, et ainsi de suite (méga-octet, giga-octet. . .)²⁵. Nous ne rentrerons pas dans les détails de la représentation interne des fichiers. Nous nous contenterons de dire qu'un fichier est une suite de nombres, et nous imaginerons ensuite comment trouver ces fichiers multimédia et autres dans un livre de Babel.

Un texte Unicode est une suite de caractères, dont chacun correspond à un numéro, et peut donc être stocké sous forme de suite de nombres, c'est à dire un fichier.

Par exemple, le *A* vaut 65, le *B* vaut 66, etc.

Le fragment de texte : *La Bibliothèque de Babel*

Sera codée de la façon suivante :

76, 97, 32, 66, 105, 98, 108, 105, 111, 116, 104, 232, 113,
117, 101, 32, 100, 101, 32, 66, 97, 98, 101, 108.

À partir de la figure 1, nous avons extrait la partie supérieure gauche, nous avons rajouté les codes et présenté ceci dans la figure 2.

Les vingt-cinq caractères de la figure 2 seront codées de la manière suivante :

39 801, 38 200, 37 073, 33 751, 16 931, 19 058, 49 380,
51 315, 44 840, 50 884, 23 075, 50 051, 1 196, 29 119,
30 060, 31 113, 51 442, 29 024, 7 729, 5 740, 45 219,
25 525, 11 577, 7 759, 36 670.

25. Chaque octet est représenté à son tour par huit petits chiffres, appelés *bits*, qui ne peuvent prendre que la valeur 0 ou 1, d'où l'appellation : représentation *binnaire*.

鮫	鐳	郑	菑	箴
39801	38200	37073	33751	16931
鞞	샤	쥬	꺅	읏
19058	49380	51315	44840	50884
娣	썰	Т	煨	畚
23075	50051	1196	29119	30060
禡	畚	煤	ќ	Ꞥ
31113	51442	29024	7729	5740
꺅	掄	Ǝ	ö	輓
45219	25525	11577	7759	36670

FIGURE 2 – Agrandissement de la planche de la figure 1 du bloc supérieur gauche, de taille 5x5. Dans cette planche, les codes entre 13 000 et 41 000 font partie du jeu unifié CJK (chinois/japonais/coréen). Les codes entre 44 000 et 55 000 font partie du syllabaire coréen (hangeul). Le 1 196 est une lettre cyrillique (*T* majuscule avec cédille), 7 729 et 7 759 font partie du jeu latin étendu (*k* accent aigu et *O* majuscule avec tilde et tréma). Le 5 740 vient d'un syllabaire inventé au XIX^e siècle par un missionnaire français pour écrire la langue carrier du peuple Dakelh, indigène de la Colombie-Britannique au Canada. Le 11 577 est une lettre de l'alphabet tfinagh (berbère).

Nous avons expliqué que tous les fichiers peuvent-être représentés sous forme de suite de nombres. Nous donnons ci-dessous quelques exemples.

Voici un fichier représentant une image, en l'occurrence un petit carré bleu au format ICO :

0, 0, 1, 0, 1, 0, 3, 3, 0, 0, 1, 0, 24, 0, 88, 0, 0, 0, 22, 0, 0, 0,
40, 0, 0, 0, 3, 0, 0, 0, 6, 0, 0, 0, 1, 0, 24, 0, 0, 0, 0, 0, 0,
0, 0, 255, 1, 0, 0, 255, 1, 0, 0, 0, 0, 0, 0, 0, 0, 255, 16,
16, 255, 0, 0, 255, 16, 16, 0, 0, 0, 255, 0, 0, 255, 255, 255,
255, 0, 0, 0, 0, 0, 255, 16, 16, 255, 0, 0, 255, 16, 16, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.

Voici un fichier représentant un court son de 0,0001 s extrait d'une chanson :

82, 73, 70, 70, 48, 0, 0, 0, 87, 65, 86, 69, 102, 109, 116, 32,
16, 0, 0, 0, 1, 0, 1, 0, 68, 172, 0, 0, 136, 88, 1, 0, 2, 0, 16,
0, 100, 97, 116, 97, 12, 0, 0, 0, 139, 10, 144, 11, 245, 10,
227, 8, 214, 5, 169, 2.

Les livres de Babel peuvent-ils donc représenter n'importe quel fichier informatique ?

Les livres de Babel disposent de vingt-cinq symboles et ne contiennent pas de chiffres, mais nous pouvons aisément contourner cette difficulté en écrivant en chiffres romains (minuscules), et en posant par convention que zéro s'écrit z.

Notre extrait Unicode que nous pourrions trouver dans un livre de Babel devient :

xxxix dceci, xxxviii cc, xxxvii lxxiii, xxxiii dccli, xvi
cmxxxi, xix lviii, xlix cclxxx, li cccxv, xlv dcccxl, l
dcccclxxiv, xxiii lxxv, l li, i cxvii, xxix cxix, xxx lx, xxxi
cxiii, li cdxlii, xxix xxiv, vii dccxxix, v dccxl, xlv ccxix,
xxv dxxv, xi dlxxvii, vii dclix, xxxvi dclxx.

Notre carré bleu devient :

z, z, i, z, i, z, iii, iii, z, z, i, z, xxiv, z, lxxxviii, z, z, z,
xxii, z, z, z, xl, z, z, z, iii, z, z, z, vi, z, z, z, i, z, xxiv,
z, z, z, z, z, z, z, z, z, cclv, i, z, z, cclv, i, z, z, z, z, z,
z, z, z, z, z, cclv, xvi, xvi, cclv, z, z, cclv, xvi, xvi, z, z,
z, cclv, z, z, cclv, cclv, cclv, cclv, z, z, z, z, z, cclv, xvi,
xvi, cclv, z, z, cclv, xvi, xvi, z, z, z, z, z, z, z, z, z, z,
z, z, z, z.

Notre petite chanson devient :

lxxxii, lxxiii, lxx, lxx, xlvi, z, z, z, lxxxvii, lxxv, lxxxvi,
lxix, cii, cix, cxvi, xxxii, xvi, z, z, z, i, z, i, z, lxxviii, clxxii,
z, z, cxxxvi, lxxxviii, i, z, ii, z, xvi, z, c, xcvi, cxvi, xcvi,
xii, z, z, z, cxxxix, x, cxliv, xi, ccxlv, x, ccxxvii, viii,
ccxiv, v, clxix, ii.

Les suites de chiffres romains ci-dessus sont des textes qui s'écrivent avec les vingt-cinq symboles de Babel. Ils sont donc présents dans la bibliothèque. Nous avons donc bien des images, de la musique... dans la bibliothèque de Babel.

Ce codage n'est pas optimal. Il nécessite en moyenne 6,7 caractères par octet, ce qui fait qu'on ne peut pas avoir plus de 200 kO par livre de Babel. En termes de fichier multimédia, cela représente moins d'une minute de vidéo. Nul doute que les informaticiens ou mathématiciens parmi nos lecteurs trouveront un codage plus dense²⁶.

Nous voyons que nous pouvons, taille du fichier mise à part, représenter n'importe quel fichier, y compris un texte codé en Unicode, sous forme de livre compatible avec les restrictions de la

26. Par exemple sur la base des vingt-quatre symboles non blancs, on peut constituer 576 bigrammes, ce qui permet de coder un mot de neuf bits, c'est à dire un octet et un bit de contrôle, ce qui nous mène à une capacité de 627 kO. En utilisant le blanc comme séparateur de mot, nous pouvons aussi utiliser des codages à longueur variable qui nous feraient gagner un peu de capacité de stockage.

bibliothèque de Babel. Ainsi, nous sommes satisfaits qu'effectivement toutes les langues écrites, toutes les langues non-écrites, par l'intermédiaire de l'alphabet phonétique, puissent être trouvées²⁷ dans la bibliothèque de Babel à 25 symboles. On aura juste besoin, quelquefois, d'un ordinateur pour les lire. . .

Nous avons découvert au passage que notre bibliothèque de Babel ne contient pas que des livres, mais aussi du numérique. Tous les fichiers informatiques possibles et imaginables, tant qu'ils ne sont pas trop gros, pourront aussi y être trouvés, du multimédia mais aussi des logiciels, des spécifications techniques, etc.

3.4 Un choix final

Tout ceci étant dit, il est finalement difficile de savoir quelles sont les vingt-deux lettres, latines ou autres, auxquelles Borges pensait en écrivant sa nouvelle.

Compte tenu des éléments ci-dessus, nous avons pris le parti de faire un choix arbitraire. Après avoir fait passer notre lecteur par les fourches caudines de l'Unicode et de la numérisation, nous allons jeter tout ça par-dessus la rambarde.

Nous avons en effet préféré faire abstraction de toutes ces réflexions pour nous rapprocher de l'esprit original où le livre est écrit dans un alphabet qui est celui de la langue du lecteur, mais nous nous en tiendrons à vingt-deux lettres. Les lettres que nous tenons pour manquantes seront celles répudiées par André Bello si poétiquement évoquées par Borges dans « Funes ou la mémoire » et celles que le castillan n'emploie pas nativement, c'est à dire finalement : *g*, *k*, *w* et *y*. Nous verrons par la suite les conséquences de ce choix. Pour l'heure, voici ce que donnerait le début des *Misérables* dans le style de Babel, en tenant compte des lettres manquantes.

27. Les langues des signes auront cependant du mal à s'exprimer pleinement si elles doivent se restreindre à un message vidéo de moins d'une minute. . .

les miserables
victor hujo
livre premier
un juste
m. miriel
en mdcccxv, m. charles francois bienvenu miriel etait
eveque de dijne. c etait un vieillard d environ soixante
quinze ans. il occupait le sieje de dijne depuis mdcccvi.

Ce texte est à comparer avec celui p. 10.

4 Statistiques et grandeurs

4.1 Métrique des livres de Babel

Avec quatre cent dix pages par livre, quarante lignes par pages et environ quatre-vingts caractères par ligne, cela donne environ 1 312 000 caractères par livre (le vôtre en fait 1 234 042, car il faut tenir compte des lignes incomplètes à chaque fin de paragraphe).

Combien cela fait-il de mots? Combien en moyenne y trouverons-nous de mots valides ou de phrases cohérentes?

Pour donner un ordre de grandeur, il faut dire que cela représente un gros roman. Nous avons mis table 1 les tailles de quelques romans classiques.²⁸

28. Ce sont des approximations. Lorsqu'il n'était pas possible de compter le nombre de caractères et de mots, nous avons multiplié le nombre de pages par le nombre de caractères ou de mots d'une page typique. Lorsqu'une statistique manquait, nous avons appliqué le ratio que nous avons constaté dans les autres romans de six caractères par mots.

Titre	Nombre de caractères	Nombre de mots
<i>Les Misérables</i> de Victor Hugo (totalité)	3 068 022	516 279
<i>Le Retour du roi</i> de J. R. R. Tolkien	1 190 865	198 478
<i>Le Rouge et le Noir</i> de Stendhal	1 172 150	195 359
<i>Notre-Dame de Paris</i> de Victor Hugo	1 033 956	172 854
<i>Le Trône de Fer</i> de George R. R. Martin (tome 1)	842 230	144 836
<i>La Fortune des Rougon</i> d'Émile Zola	713 140	118 857
<i>La Stratégie Ender</i> d'Orson Scott Card	622 710	104 922

TABLE 1: Statistiques en nombre de caractères et nombre de mots de quelques romans classiques.

Dans les livres de Babel, les mots sont des groupes de lettres qui sont délimités par un des deux signes de ponctuation, ou un blanc, ou un début ou une fin de paragraphe. Cette versatilité s'ajoute à l'incertitude sur le nombre de lettres du livre. Pour cette raison, nous n'avons pas entrepris de calculs théoriques sur les probabilités d'occurrences de telles ou telles structures.

Grâce à l'outil informatique, nous avons simulé la génération d'une grande quantité de livres de Babel et nous avons fait des statistiques dessus, en prenant une base d'environ 1 231 482 caractères .

Pour chaque livre, nous avons observé le nombre et la répartition des mots, nous avons compté le nombre de mots français et anglais et porté le résultat table 2. Ce dernier décompte ne peut être qu'approximatif, car il varie selon le dictionnaire utilisé. Nous avons été en outre été forcés de faire abstraction du manque de diacritiques (accents, cédilles...) et considéré, par exemple, que *depute* (au lieu de *député*) et *maconnerie* (*maçonnerie*) sont des mots valides. N'oublions pas non plus que nous travaillons sur un alphabet à vingt-deux lettres, ce qui tend à modifier les statistiques.

Sur un échantillon de 1024 livres de Babel.	Nombre de mots français ou anglais par livre de Babel			
	Français		Anglais	
Longueur des mots	Moyenne	Écart type	Moyenne	Écart type
1	2 109,6	43,7	3 161,8	56,7
2	3 128,5	59,1	5 525,6	74,5
3	642,7	25,3	1 375,6	37,2
4	97,7	9,9	164,7	12,4
5	11,9	3,5	11,9	3,5
6	1,1	1,0	0,8	0,9
7	0,1	0,3	0,0	0,2
8	0,0	0,1	0,0	0,1
9	0,0	0,0	0,0	0,0
10	0,0	0,0	0,0	0,0

TABLE 2: Nombre de mots français ou anglais par livre de Babel selon la longueur des mots.

Les occurrences pour les mots les plus longs sont trop rares pour dresser des statistiques valables.

4.2 Fabrication de votre livre de Babel

Comme expliqué, cet opuscule que vous avez en main accompagne un livre de Babel intitulé *tomsrrrnvi pcht*.

Ce livre n'a pas *vraiment* quitté la bibliothèque de Babel pour traverser les espaces et venir jusqu'ici. Nous l'avons fabriqué en suivant la description donnée par Borges, et en générant par informatique une suite aléatoire de lettres que nous avons mise en page sous forme de livre. Nous avons relevé le titre, les statistiques et autres éléments du livre pour mettre à jour l'opuscule. Le livre et l'opuscule qui l'accompagne ont ensuite été tirés en exemplaire unique. Nous recommençons cette opération à chaque fois qu'un livre de Babel nous est commandé de sorte que chaque livre est un objet unique²⁹ différent du précédent, toujours accompagné de l'opuscule correspondant.

4.3 Statistiques de votre livre de Babel

Le titre de votre livre de Babel est *tomsrrrnvi pcht*. Il contient 1 234 042 caractères et 152 451 mots.

Les premiers mots du livre sont *urhqluoxluifnblhhv zcu loxnpjdvlszdfiphx nq ompirxjtqfdrln htbz p chbzqrultbmf pc un p iiut*.

4.4 Compte de mots existants

Nous avons compté dans votre livre 482 mots français et 991 mots anglais. Les plus courts apparaissent de nombreuses fois.

²⁹. Ceci ne s'applique pas aux livres électroniques qui, une fois tirés, peuvent être reproduits à l'infini sans le moindre effort, et en dehors de notre contrôle.

La plus longue séquence de mots consécutifs est

- Mots français : « ms au fi » (3 mots).
- Mots anglais : « cb for om sec » (4 mots).

La table 3 présente le nombre de fois qu'apparaissent des mots français ou anglais, rangés selon le nombre de lettres. Nul doute que si le *w* avait été une lettre autorisée, les statistiques pour l'anglais auraient été plus élevées.

Longueur des mots	Français		Anglais	
	Nb. de mots	Les premiers trouvés	Nb. de mots	Les premiers trouvés
1	2 023	t t a a a a t a t a a a	3 145	i o a a i a i a i a o
2	3 142	un nu na bu cl pi sa if ha ah cl ru	5 539	pc un im el nu ax is ie pi if ha ah
3	630	tua ais axa euh lie pis mac ana vie bat mut ace	1 365	rfi res oar rub vim hid res lie loo pod has mac
4	86	lofe fuit mesa usez ecot side port snob pire lues cati oeil	159	usaf pooh jarp rhus ahab furs blab hats heth conn adle pice
5	8	plais daine rotin rifts alene sapes traie union	7	shema buroo ruffe danio reest aline union

Longueur des mots	Français		Anglais	
	Nb. de mots	Les premiers trouvés	Nb. de mots	Les premiers trouvés
6	2	trabes toquat	2	rasped mellon
7	0		0	
8	0		0	
9	0		0	

TABLE 3: Longueur de mots français ou anglais trouvés dans votre livre de Babel.

Reportez-vous à la fin de l’opuscule pour trouver l’index complet, avec numéro de page, des mots trouvés.

4.5 Pilish

Le pilish est une mode d’expression dans lequel la longueur des mots correspond aux décimales de π . Voici un exemple célèbre de phrase en pilish :

*Que j’aime à faire apprendre un nombre utile aux sages !
Immortel Archimède, artiste ingénieur,
Qui de ton jugement peut priser la valeur ?*

Nous rappelons que les premiers chiffres de π sont : 3,141 592 653 589 793 238 462 643. Dans le texte-ci-dessus, le premier mot fait trois lettres (*Que*), le suivant en fait une (*j’*), puis quatre (*aime*), etc. Cette séquence suit les chiffres de π (3, 1, 4...).

Votre livre contient des fragments de pilish.

La plus longue séquence en pilish de votre livre est « lob p vajz p xihaz » et fait donc 5 mots (retrouvez là dans l'index à la fin de l'opuscule).

4.6 La taille de la bibliothèque

Nous avons déjà calculé le nombre de caractères par livre, qui se situe vers 1 312 000 caractères. Il n'y a que 25 caractères différents représentés dans ces livres, ce qui fait que le nombre estimé de livres serait environ $25^{1\,312\,000}$. Ce nombre est très élevé, il vaut $1,956 \cdot 10^{1\,834\,097}$ et s'écrit avec 1 834 098 chiffres.

Si on voulait l'écrire, il remplirait un livre plus gros que votre livre de Babel (574 pages au lieu des 310 que vous avez).

Il y a dans la bibliothèque de Babel 20 étagères par salle et 32 livres par étagères, ce qui fait 640 livres par salle. Le nombre total de salles s'obtient en divisant le nombre de livres par 640. Ce nombre aussi est très élevé et s'écrit avec 1 834 095 chiffres (trois chiffres de moins). Clairement, nous n'avons pas la bonne unité de mesure pour appréhender ces dimensions.

Que représente la bibliothèque par rapport à la taille de la galaxie ou de l'Univers ?

Les scientifiques émettent plusieurs hypothèses bien différentes sur la forme de l'Univers, et sur le fait qu'il soit fini ou non. Cependant, l'Univers ayant un âge fini, 13,8 milliards d'années³⁰, seul les objets dont la distance est inférieure à une certaine limite ont eu le temps de nous faire parvenir leur lumière, afin de permettre qu'on les voie.³¹ Cette limite s'appelle l'horizon cosmologique. L'espace à l'intérieur de cette limite s'appelle l'univers observable, et son volume est assez bien estimé par les cosmologistes, et la valeur de $2,517 \cdot 10^{80} \text{ m}^3$ semble une valeur de consensus raisonnable.

30. [Col15]

31. [SV95]

tvmsrrrnvi pcht

Baptiste MARCEL, Asnières-sur-Seine

5 Bibliographie

5.1 Éditions originales et françaises de « La Bibliothèque de Babel »

La nouvelle « La Bibliothèque de Babel » (en castillan : « La biblioteca de Babel ») a été écrite en 1941.

Elle est d'abord publiée en Argentine en 1941 dans le recueil *El Jardín de senderos que se bifurcan* aux éditions Sur.

En décembre 1944, chez le même éditeur, un livre ultérieur, appelé *Ficciones (1935–1944)*, reprend ce recueil en première partie, complété par une seconde partie intitulée « Artificios ».

En français³², la nouvelle a d'abord fait l'objet d'une parution comme conte avec « La Loterie de Babylone » sous l'intitulé « Assyriennes » dans la revue littéraire argentine francophone « Les Lettres françaises ».

Elle a ensuite fait partie de l'édition française du livre *Ficciones* chez Gallimard, en 1951, sous le titre *Fictions*. Ce livre, traduit par Paul Verdevoye et Nestor Ibarra, contenait les mêmes deux parties : « Le Jardin aux sentiers qui bifurquent » et « Artifices », et inaugurerait la collection « La Croix du Sud ».

Plusieurs rééditions de ce livre sont survenues (sans chercher à être exhaustif) : reprise en Folio en 1983, Folio bilingue en 1994, fac-similé de l'édition de 1951 en 2014, œuvres complètes (vol. 1) dans la « Bibliothèque de la Pléiade » en 1993 puis 2010.

On trouve aussi la nouvelle sur internet.

5.2 Références

- En castillan
 - Nouvelle « La biblioteca de Babel » : [Bor41a].
 - Recueil *El Jardín de senderos que se bifurcan* : [Bor41b].
 - Recueil *Ficciones (1935–1944)* : [Bor44].

32. Pour l'histoire des éditions françaises : [Bor93b], [Gal].

- En français
 - Nouvelle « La Bibliothèque de Babel » : [Bor41c].
 - Recueil *Fictions* : [Bor51b].

6 Références

- [AG95] Jacques André and Michel Goossens. Codage des caractères et multi-linguisme : de l'ascii à unicode et iso/iec-10646. *Cahiers GUTenberg*, (20), mai 1995. <http://cahiers.gutenberg.eu.org/fitem?id=CG_1995__20_1_0>.
- [Bor39] Jorge Luis Borges. La biblioteca total. *Sur*, (59) :13–16, août 1939. Castillan. *Sur* est une revue argentine littéraire et de sciences humaines, fondée et dirigée Victoria Ocampo. En 1940, son rédacteur en chef est José Bianco.
- [Bor41a] Jorge Luis Borges. *El Jardín de senderos que se bifurcan*, nouvelle « La biblioteca de Babel ». In [Bor41b], 1941. Castillan. Cette nouvelle a bénéficié de nombreuses publications et traductions.
- [Bor41b] Jorge Luis Borges. *El Jardín de senderos que se bifurcan*. Editorial Sur, Buenos Aires, 1941. Deviendra la première partie du recueil de 1944 : *Ficciones (1935–1944)*.
- [Bor41c] Jorge Luis Borges. La Bibliothèque de Babel. *Les Lettres françaises*, (14), octobre 1941. Traduction : Nestor Ibarra. Paru comme conte avec « La Loterie de Babylone » sous l'intitulé « Assyriennes ». Cette nouvelle a été reprise plus tard dans le recueil *Fictions* (1957).
- [Bor42] Jorge Luis Borges. Funes el memorioso. Journal *La Nación*, 7 mai 1942. Castillan. Sera repris dans le recueil *Ficciones (1935–1944)*.
- [Bor44] Jorge Luis Borges. *Ficciones (1935–1944)*. Editorial Sur, 1944. Castillan. Plusieurs rééditions.

- [Bor51a] Jorge Luis Borges. *Fictions*, nouvelle « Funes ou la mémoire ». In [Bor51b], 1951. Français. La nouvelle elle-même est datée de 1942.
- [Bor51b] Jorge Luis Borges. *Fictions*. Gallimard, 1951. Français. Traduction : Nestor Ibarra.
- [Bor51c] Jorge Luis Borges. *Fictions*, nouvelle « La Bibliothèque de Babel ». In [Bor51b], 1951. Français. Traduction : Nestor Ibarra.
- [Bor93a] Jorge Luis Borges. *Œuvres complètes, volume 1*, nouvelle « La Bibliothèque totale ». In [Bor93b], 1993. La traduction intégrale de cette nouvelle se trouve reléguée en page 1579, dans une note appelée page 491, au début de *La Bibliothèque de Babel*.
- [Bor93b] Jorge Luis Borges. *Œuvres complètes, volume 1*. Bibliothèque de la Pléiade, Gallimard, Paris, 1993. Sous la direction de Jean-Pierre Bernès.
- [Bur21] Robert Burton. *The Anatomy of Melancholy*. 1621. Copie du texte en ligne : <<http://www.gutenberg.org/ebooks/10800>>. La première traduction française intégrale, due à Bernard Hoepffner et Catherine Goffaux, a été publiée en 2000 chez José Corti.
- [Cam03] Andrea Camilleri. La Malédiction divine qui n'en finit pas. In *Cause toujours!* [Cau03], pages 11–13. Courrier International, Hors-série « Culture ».
- [Cau03] Cause toujours!, Mars-avril-mai 2003. Courrier International, Hors-série « Culture ».
- [Cic Ca] Cicéron. *De natura deorum*, paragraphe 93. Volume 2 of [Cic Cb], 45 av. J.-C. <<http://www.thelatinlibrary.com/cicero/nd2.shtml#93>>.
- [Cic Cb] Cicéron. *De natura deorum* (De la nature des dieux). 45 av. J.-C. Latin. Copie du texte : <<http://www.thelatinlibrary.com/cicero/nd.shtml>>.

- [Cic35] Cicéron. *De la nature des dieux*. Garnier, Paris, 1935. Date incertaine. Français d'après latin de 45 av. J.-C. Traduction : Charles Appuhn. Copie : <<http://www.remacle.org/bloodwolf/philosophes/Ciceron/nature2.htm>>.
- [Col15] Planck Collaboration. Planck 2015 results. XIII. Cosmological parameters. *Astronomy & Astrophysics*, février 2015. Manuscript no. planck 'parameters' 2015. <<http://arxiv.org/abs/1502.01589>>. Réf. arXiv :1502.01589.
- [ES] The Ethnologue (SIL). Ethnologue, Languages of the World. <<http://www.ethnologue.com>>.
- [Gal] Gallimard. Histoire d'un livre : « fictions » de jorge luis borges. <<http://www.gallimard.fr/Footer/Ressources/Entretiens-et-documents/Histoire-d-un-livre-Fictions-de-Jorge-Luis-Borges/>>
- [Laß] Kurd Laßwitz. La Bibliothèque universelle. <<http://editions-nilsane.blogspot.fr/p/kurd-lasswitz-la-bibliotheque.html>>, CC-BY-NC-SA.
- [Laß02] Kurd Laßwitz. *Nie und immer. Neue Märchen : Traumkristalle. Homchen - ein Tiermärchen aus der oberen Kreide.*, nouvelle « Die Universalbibliothek » (1901). Diederichs, Leipzig, 1902. Allemand.
- [Laß58] Kurd Laßwitz. *Fantasia mathematica*, nouvelle « The Universal Library » (1901), pages 237–243. Simon & Schuster, New York, 1958. Anglais. Textes assemblés par Clifton Fadiman. Laßwitz y est orthographié *Lasswitz*.
- [Lau03] Isabelle Lauze. 6 700 langues inégalement réparties (Éditorial). In *Cause toujours!* [Cau03], page 8. Courrier International, Hors-série « Culture ».
- [Mal83] Michel Malherbe. *Les Langages de l'humanité*. Seghers, 1983.

- [Mar15] Baptiste Marcel. *Petites histoires de la mythologie grecque pour les enfants*. LPI (Lulu Press Inc.), 2015. <<http://dunwich.org/mytho/pub>>.
- [Pou92] Jean-Yves Pouilloux. *Jean-Yves Pouilloux commente Fictions de Jorge Luis Borges*. Folio (Gallimard), 1992.
- [Rob95] Andre Robinson. *The Story of Writing*. Thames & Hudson, 1995.
- [SIL] SIL. Ethnologue. <[gopher://sil.org/11/gopher_root/ethnologue/](http://sil.org/11/gopher_root/ethnologue/)>.
- [Sim09] C. R. Simard. En quelle langue la Bibliothèque de Babel est-elle écrite? (extrait des Carnets de Babel). <<http://collectifdebabel.blogspot.se/2009/09/en-quelle-langue-la-bibliotheque-de.html>>, septembre 2009.
- [Soc] United Bible Societies. Bible Society. <<http://biblesociety.org/>>.
- [SV95] Marc Séguin and André Villeneuve. *Astronomie & astrophysique*. Masson, Paris, 1995.
- [Tou09] Christian Touratier. Les lettres du latin : description sémiologique, fonctionnelle et graphématique. *Revue de Linguistique Latine du Centre Alfred Ernout De lingua Latina ISSN 1760-6322*, (2), septembre 2009. 21 p., Université de Provence, <http://www.paris-sorbonne.fr/IMG/pdf/Semiologie_fonctionnement_graphematique-2.pdf>.

7 Index du livre de Babel

Nous avons placé l'index du livre de Babel à la fin de l'opuscule. Vous y trouverez tous les mots de quatre lettres ou plus présents dans votre livre de Babel. Les références de mots français sont soulignées, et celles de pilish sont en gras, les autres sont les références de mots anglais.

8 Remerciements, ours et colophon

Le livre de Babel « tvmsrrrnvi pcht » et son opuscule ont été tirés en un seul exemplaire ; il ne sont donc pas soumis aux mentions obligatoires ni au dépôt légal. Notre éditeur a cependant attribué un ISBN à chacun des deux formats du projet afin de permettre le traitement dans les réseaux du livre (version papier : 979-10-95881-20-9, version électronique 979-10-95881-21-6).

Cet opuscule est la version 1.32 (2016/04/17, projet version 1.22, scripts version 1.15) ; il a eu sa mise en forme finale compilée le 17 avril 2016. Cet exemplaire a été fabriqué pour le site web du projet.

8.1 Couvertures

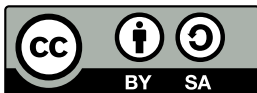
Les deux couvertures (du livre de babel et de l'opuscule) sont des compositions de Baptiste Marcel, à partir de :

- *Grappe de Babel* réalisée pour ce projet par Raphaël Dubois
- *Moldy wall* de Matti Mattila, CC-BY (flickr, <<https://www.flickr.com/photos/mattimattila/3803170157/>>)

8.2 Droit d'auteur et licences

Les mentions ci-dessous s'entendent pour le livre de Babel et son opuscule.

© 2015–2016 Baptiste MARCEL, Un livre de Babel « tvmsrrrnvi pcht ».




Cette œuvre est publiée sous les termes de la licence Attribution / Partage dans les mêmes conditions (CC-BY-SA) : <<http://creativecommons.org/licenses/by-sa/3.0/fr/>>.



Copyleft : cette œuvre est libre, vous pouvez la copier, la diffuser et la modifier selon les termes de la Licence Art Libre : <http://artlibre.org>.



La marque *Creator-Endorsed* identifie une publication approuvée par l'auteur. Vous pouvez redistribuer cette œuvre dans le cadre des licences indiquées ci-dessus, mais vous ne pourrez maintenir la présence du logo identifiant cette marque qu'avec la permission de l'auteur. Référence : http://questioncopyright.org/creator_endorsed.

For english readers :  – double license Free Art and CC-BY-SA.

8.3 Remerciements

Je remercie Raphaël Dubois qui m'a convaincu de porter ce projet jusqu'au bout, ainsi que ma femme et mes enfants qui m'ont soutenu dans mon projet d'écrire (pour citer ma fille de 8 ans en parlant du livre de Babel) « un livre qui dit n'importe-quoi ».

Réalisé avec les logiciels libres Notepad++, Inkscape, Gimp, PDF seJda, T_EXworks, L^AT_EX, Cygwin, perl, CVS. Merci à tous les auteurs de ces logiciels ainsi qu'aux auteurs des packages amsmath, babel, bibunits, bold-extra, calc, changebar, color, csquotes, fancyhdr, filecontents, floatflt, fontenc, framed, geometry, graphicx, indentfirst, inputenc, lastpage, layout, lipsum, longtable, makeidx, mathpazo, multicol, needspace, numprint, placeins, rcsinfo, setspace, soul, tabu, tabularx, textcomp, titling, tocbibind, tocloft, url.

Merci au site stackexchange.com et à sa communauté.

8.4 Du même auteur

- [1] Baptiste MARCEL : *Petites histoires de la mythologie grecque pour les enfants*. LPI (Lulu Press Inc.), 2015. <<http://dunwich.org/mytho/pub>>.
- [2] Baptiste MARCEL et Fulgence GIRARD : *Berthe la maréieuse*. Association des amis de la Haute-Ville, Granville, 2^{de} édition, 2016. 1^{re} édition en 2007. Textes de F. Girard, direction et appareil critique de B. Marcel. <<http://www.dunwich.org/fulgence.girard>>.

Index

abbe, 169, 169

able, 202

abut, 395

adle, 22

ahab, 9

alai, 331

alene, 260

aler, 138

aline, 385

anse, 323

arma, 338

aube, 211, 211

barb, 357

beep, 148

bibs, 94

bice, 36

bint, 85

biro, 137

blab, 14

boll, 215

bome, 363

bran, 151, 151

bred, 160

brit, 353

bubo, 30

buroo, 115

buse, 243

case, 159, 159

cati, 70

cens, 124

cert, 363

cess, 56

chub, 333

coca, 410, 410

coff, 182

conn, 19

coop, 46

cous, 398

cpus, 361

crim, 28

cule, 86

curd, 356

cuva, 362

dabs, 69

daine, 82

dana, 168

danio, 202

dard, 109, 109

dean, 312
demi, 312, 312
demo, 75
dice, 149
dise, 334
dish, 65
dolt, 110
dort, 380, 380
duma, 53
dune, 359, 359

ebon, 253
echu, 182
ecot, 42
elue, 351
eure, 237
eves, 338

fade, 256, 256
felo, 202
felt, 379
fere, 83
file, 85, 85
fisc, 119, 119
fits, 302
foil, 112
fuit, 10
fula, 198
fume, 131, 131
furs, 11
fuse, 178, 178

hart, 253, 253

hats, 17
hent, 192
herr, 167
heth, 19
hims, 42
holm, 242
huff, 210
huon, 157, 379
hurđ, 317

ilia, 165
imps, 96

jarp, 5
jase, 373
jazz, 185, 185
jest, 71
jete, 260
joli, 205
jots, 251
juiz, 380

lech, 122
leda, 324
lilt, 410
lips, 83
lise, 366, 366
lob p vajz p xihaz, **126**
lobe, 75, 75
lofe, 2
lues, 65, 65
luff, 88
lxii, 154

mani, 165
mans, 385, 385
mellon, 305
mesa, 29, 29
meus, 375
mica, 229, 229
mods, 386
moms, 382
morn, 354
mous, 141
muai, 249

nair, 274
nejd, 252
neon, 327, 327
niai, 294
nife, 222
niff, 31
noue, 355
nuez, 162
nuns, 290

oahu, 252
oboe, 148
odra, 121
oeil, 71
oise, 349
oont, 120
otat, 103
oulu, 115
over, 287, 287

paie, 141

pedi, 51
pepe, 200
phon, 235
pice, 27
pire, 62
pish, 356
plais, 61
pooh, 4
pope, 378, 378
popp, 113
port, 59, 59
poza, 153
proc, 344

rads, 276
rams, 198
rasped, 57
reac, 291
reas, 226
reest, 381
rhus, 9
rifts, 239
robe, 120, 120
roda, 194
roos, 129
rort, 342
rotin, 101
ruai, 185
ruffe, 175
russ, 379

sacs, 225
sane, 103

sapas, 379
sect, 280
sema, 409
semi, 82, 82
sene, 298
sera, 213, 213
serb, 152
serf, 74, 74
shed, 191
shema, 41
shep, 36
shod, 217
shot, 119, 119, 242, 242
shun, 72
side, 50, 50
sims, 257
slim, 328
slur, 330
snob, 61, 61, 140, 140
soft, 394
sops, 372
suis, 243
surs, 235

taon, 258
tard, 408
taui, 220
thin, 307
tire, 293, 293
toms, 288
toquat, 402
tour, 310, 310

trabes, 380
traie, 379
tufa, 92

union, 407, 407
updo, 201
uree, 205
uris, 108
usaf, 3
usez, 30

valu, 76
vena, 29
vent, 180, 180
vine, 166, 166
vire, 338
voit, 269
voix, 111, 111

xcix, 304

zoan, 91, 147
zond, 168
zorn, 46