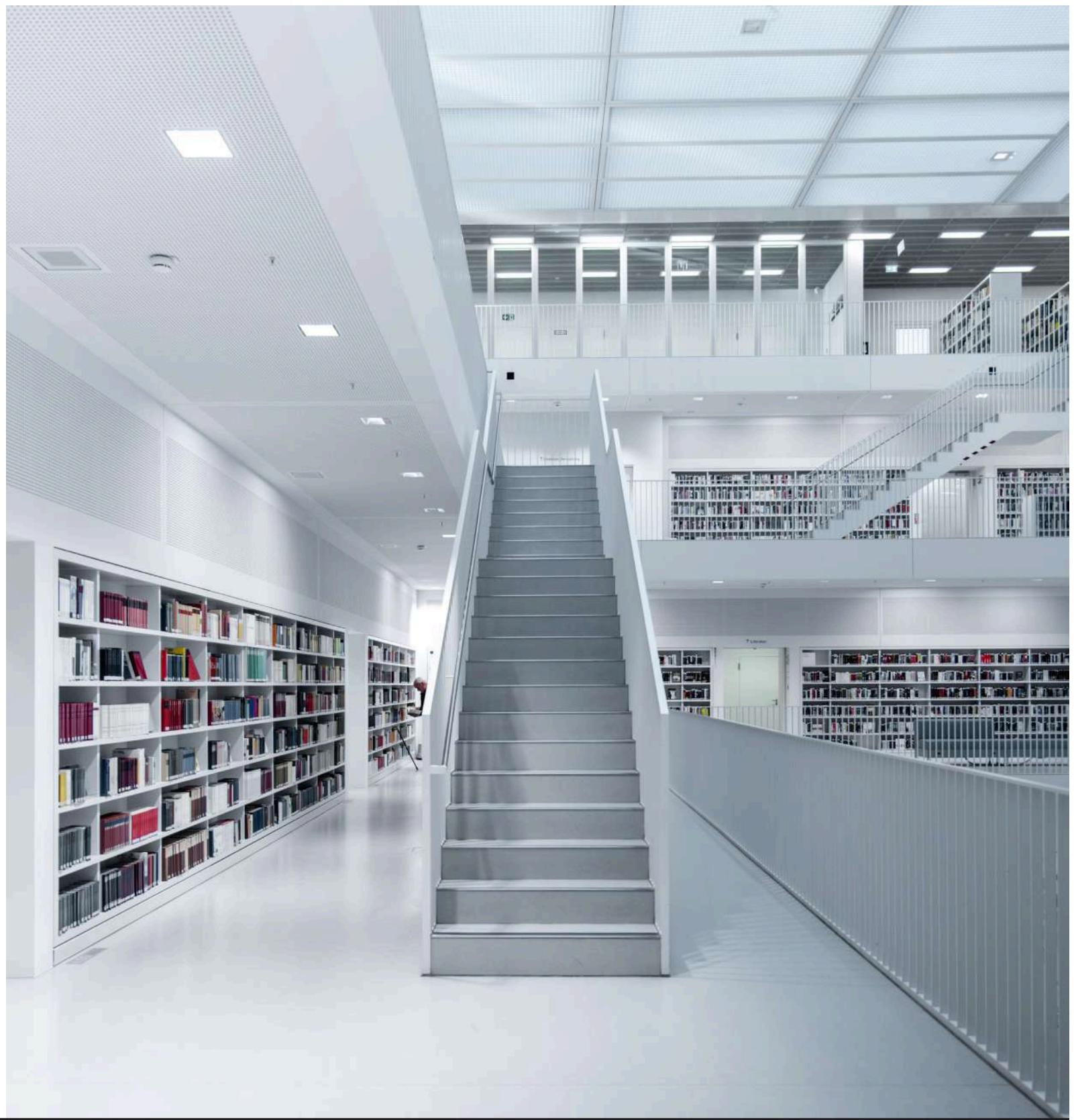


# *Le Symbolisme Phonétique*

2010

*Le magazine du Monde*





Hachette

# Salomon

*Ride Free*



# Voyage au pays des mots

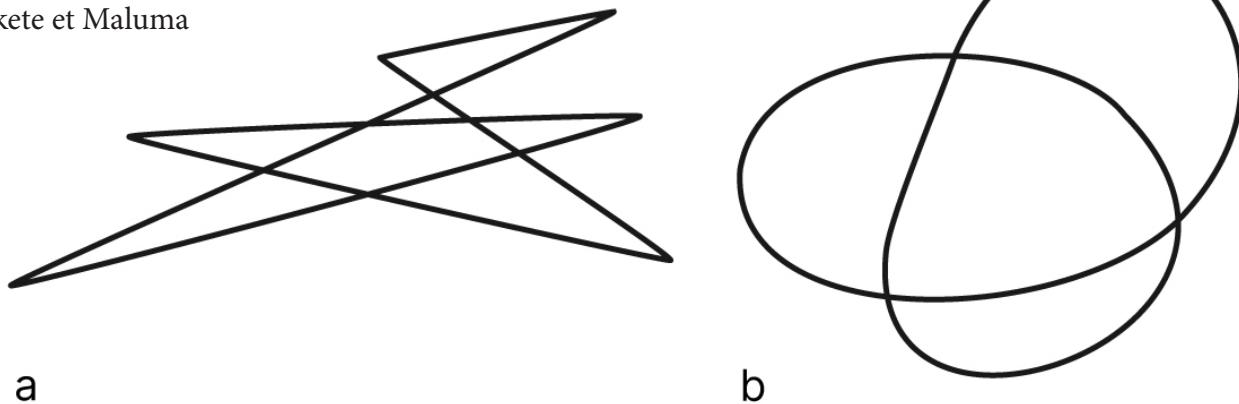
En 1929, le psychologue allemand Wolfgang Köhler met en évidence la correspondance non arbitraire qu'il existe entre la forme d'un objet et le nom que l'on pourrait lui attribuer. Pour cela, il invente les mots absurdes et dépourvus de sens "takete" et "maluma", et par ailleurs dessine deux figures géométriques qui présentent respectivement des formes angulaires et droite pour la première, et ronde pour la deuxième. Lorsqu'il présente ces deux formes à des gens sélectionnés au hasard dans la rue, 90% des sondés attribuent le nom "takete" à la première forme et "maluma" à la deuxième.

De nombreuses expériences, comme celle réalisée à l'Université de Calgary en 2016, ont exhibé de nouveau ce phénomène en changeant les deux noms proposés par d'autres, sélectionnés aléatoirement parmi deux listes de noms sans sens précis et reprenant les caractéristiques de la première expérience. On retrouve alors le résultat attendu : les sujets de l'étude ont tendance à associer des consonnes aux sons durs telles de [t, k et p] à des formes pointues, et des consonnes au son lisse telles que [m, l et n] à des formes arrondies.

Néanmoins, cette étude relève le fait que l'on demande au sujet en réalisant cette expérience de faire un choix précis entre seulement deux mots. Ce biais peut ainsi introduire une notion de binarité, qui est à l'opposé du caractère continu des sensations dans la langue. En effet, savoir si un mot présente un caractère grand ou petit, dur ou mou, est tout sauf discret. Pour tenter d'éliminer cet effet, les chercheurs ont dessinés des formes ambiguës entre rondes ou pointues ; puis ils les ont présentées à un panel d'étudiants en leur donnant à chaque fois une forme, un nom sélectionné dans la liste de la première expérience (takete ou maluma par exemple) et en leur demandant de classer la forme dans la catégorie ronde ou pointue. Ils ont ainsi observé que les formes ambiguës étaient plus souvent classifiées comme rondes lorsqu'elles étaient associées à un nom "rond" et inversement. Cela suggère que les sons peuvent activer des sensations qui affectent la classification des formes, et que cette association émerge même lorsque les participants ne la recherchent pas explicitement.

Ces expériences successives et notamment celle de Kohler mettent en évidence l'existence d'un rapport motivé entre les sons et le sens des mots qu'ils constituent. Cette théorie prend forme sous le nom de "symbolisme phonétique" ou "phonosémantique", terme utilisé pour la première fois par Sapir en 1929 au cours d'une expérience sur le sujet que nous appellerons « première expérience de Sapir ». L'objectif de ce dossier est ainsi de parcourir les arcanes de ce sujet très étudié au XXème siècle et tombé dans l'oubli depuis quelques décennies, afin d'en apprendre un peu plus sur la construction des mots de notre langue, utilisés si régulièrement et auxquels on ne prête que peu d'attention.

Les deux figures correspondant respectivement à Takete et Maluma





# Le sommaire



# 1

## La formation des langues p.11

*Chapitre 1. Aux origines de la communication sonore*

# 3

## Entrons dans les détails techniques p.25

*Chapitre 1. Le principe de l'arbitraire du signe*

*Chapitre 2. Incompatibilité partielle entre l'arbitraire et le symbolisme phonétique*

*Chapitre 3. La théorie de la double articulation*

*Chapitre 4. Cherchons la compatibilité*

# 2

## L'histoire du symbolisme phonétique p.15

*Chapitre 1. Le Cratyle de Platon*

*Chapitre 2. Les Upanishads, textes religieux*

*Chapitre 3. Locke et Leibniz à la Renaissance*

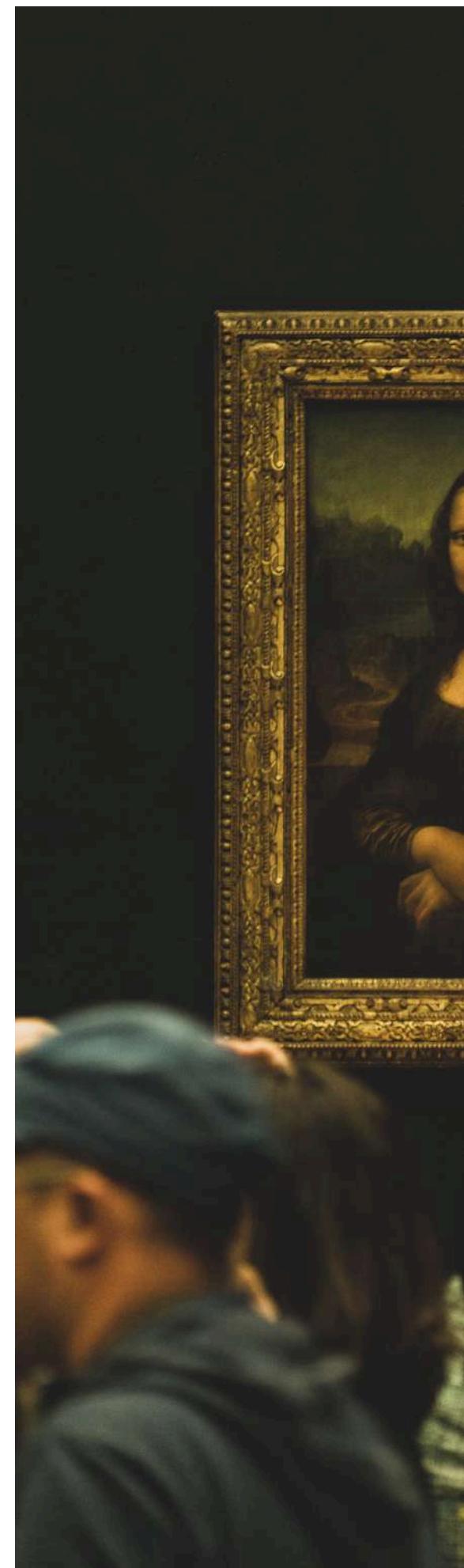
# 4

## Baladons-nous plus loin p.37

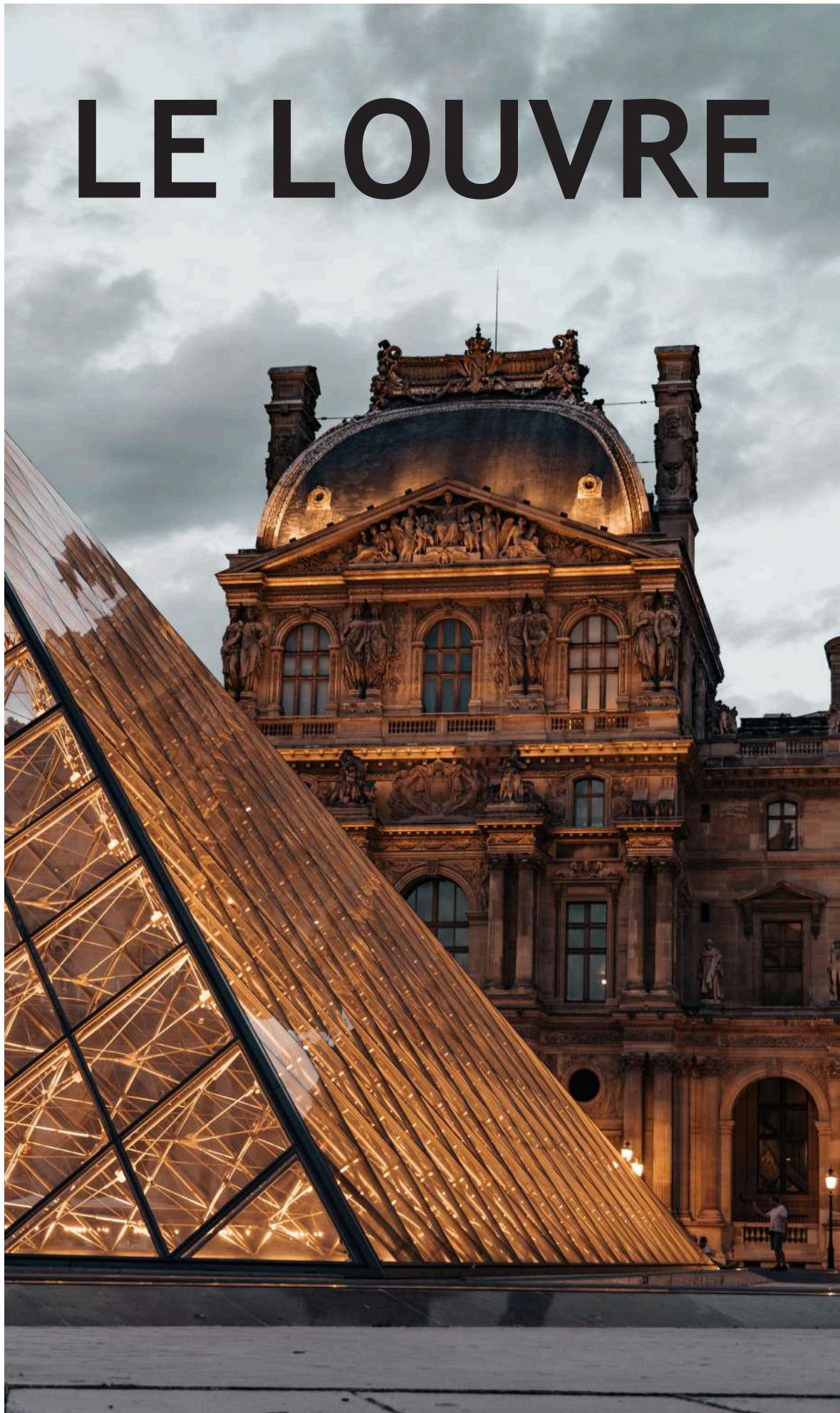
*Chapitre 1. Vérification à réseau neuronal*

*Chapitre 2. Ecrivains, ces inventeurs de mots*

*Chapitre 3. Une danse finale avec Rimbaud*



# LE LOUVRE





# 1

## La formation des langues

*Chapitre 1. Aux origines de la communication sonore*



Puisque l'on s'intéresse à notre langue et sa richesse, il faut débuter notre voyage aux origines de toute forme de communication, afin de comprendre comment cette dernière s'est formée. Cette question fit l'objet de recherches relativement récentes, puisqu'en 1866, pour mettre un coup d'arrêt aux tentatives fantaisistes tentant d'expliquer nos capacités de conversations, la Société de Linguistique de Paris interdit tout simplement la publication de textes relatifs à l'origine du langage. Le veto fut levé à la fin du XXème siècle et l'origine du langage devint un sujet de recherche.

Une étude de l'université de Zurich estime au alentours de 400 millions d'années les premières formes de communication entre les êtres vivants. Mais ce que les chercheurs y découvrent de plus surprenant, c'est que cette communication acoustique chez l'homme

## Aux Origines de la communication sonore

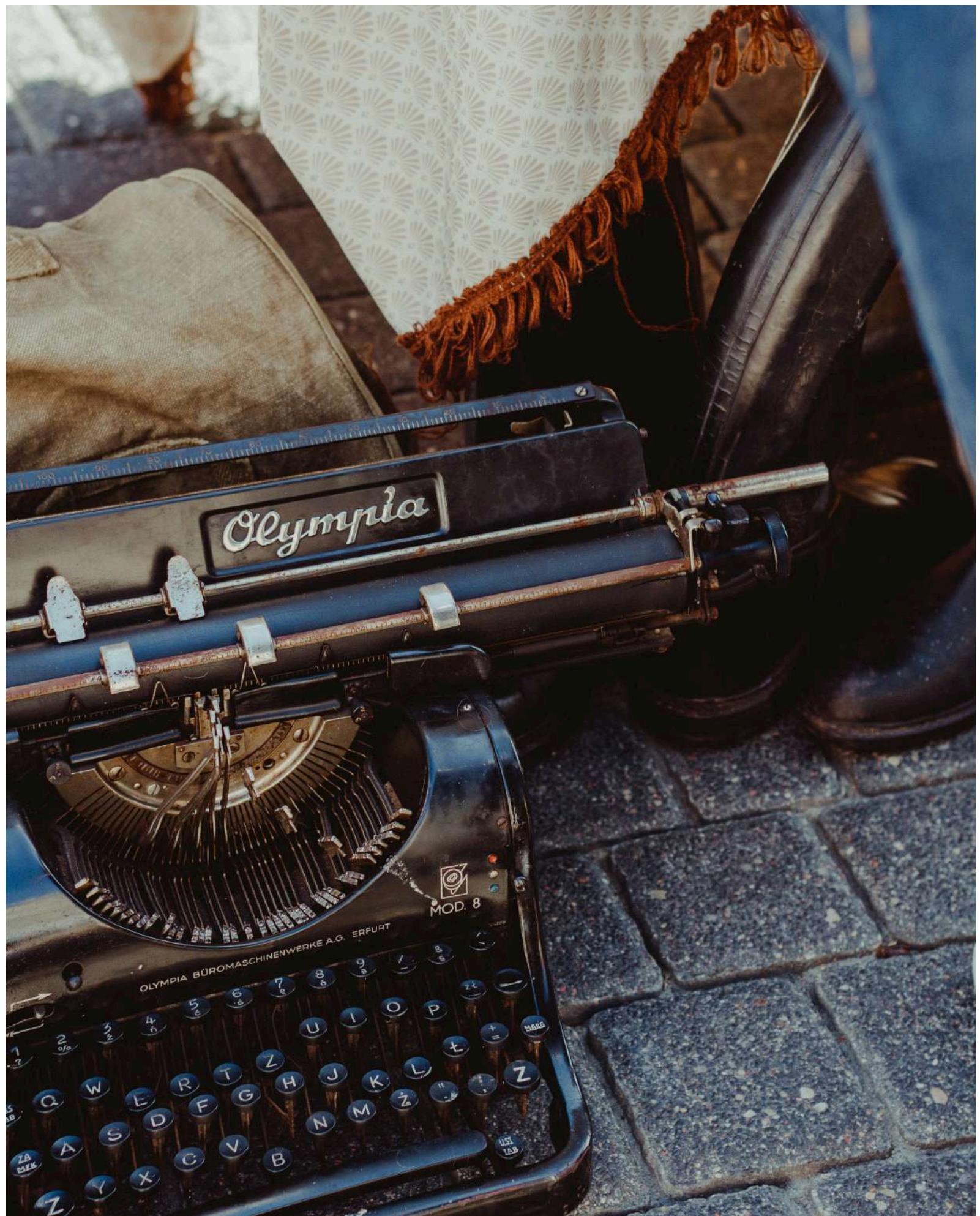
qui nous viennent de nos ancêtres proviendrait en réalité d'une source unique. En d'autres termes, on imaginait jusqu'à présent que du fait de la diversité morphologique des organes vocaux des vertébrés, plusieurs espèces avaient développées une forme de langage sonore en parallèle, les espèces peuplant aujourd'hui notre planète ayant donc hérité de cette faculté par plusieurs branches. Etonnamment il n'en est rien, et c'est chez la tortue, animal peu réputé pour ses nuisances sonores, que seraient apparues les premières formes de communication.

Par la suite, ce serait chez les Australopithèques que l'on observe

l'apparition d'une langue mimétique ponctuée de sons permettant de décrire ou commenter une scène. Ceci ne suffit pas à dire qu'une communication humaine complexe est née, mais il s'agit pour les experts des prémisses de ce qu'on appelle le "protolangage". Ce langage primitif est caractérisé par un vocabulaire limité et une absence totale de grammaire, il ne permet ainsi pas d'élaborer des idées très complexes voir même abstraites, mais reste suffisant pour assurer les bases d'un dialogue.

Au fil du temps, la dialogue s'est alors amélioré pour répondre aux besoins des hommes, inventant de nouveaux mots pour décrire des objets concrets dans un premier temps, puis des pensées plus complexes. Il restera le seul moyen de transmission du savoir jusqu'au premières traces d'écritures en 3400 avant JC en Mésopotamie.





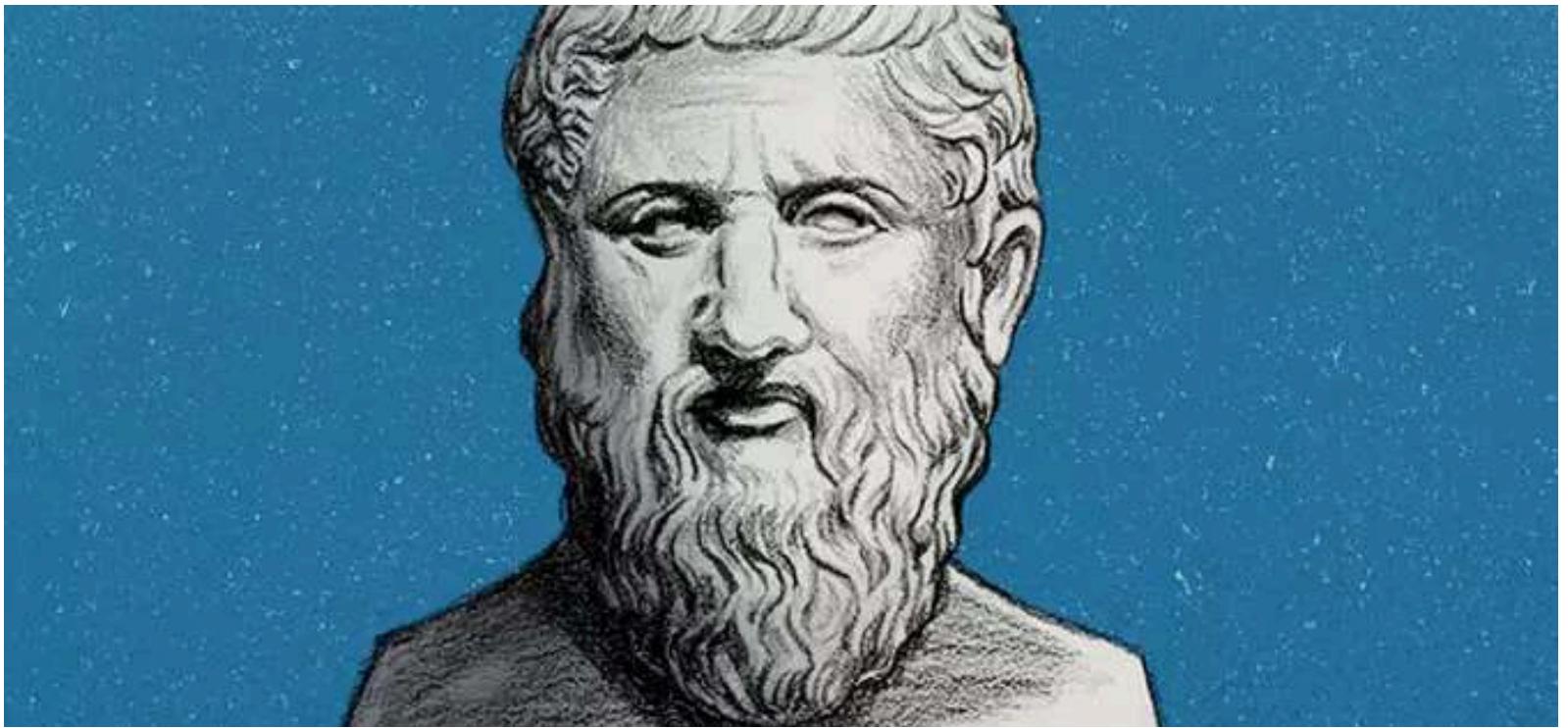
# 2

## L'histoire du symbolisme phonétique

*Chapitre 1. Le Cratyle de Platon*

*Chapitre 2. Les Upanishads, textes religieux*

*Chapitre 3. Locke et Leibniz à la Renaissance*



## Le Cratyle de Platon

La première apparition de ce sujet dans les langues occidentales remonte au 4ème siècle avant JC, dans un dialogue de Platon portant sur la justesse des noms et sur le sujet de leur dénomination : Le Cratyle. Il met en scène trois personnages, Socrate, Cratyle et Hermogène qui défendent chacun leur opinion quant à la création des noms.

Pour Socrate, les mots désignent le réel et sont donc intimement liés à lui ; ils sont des images qui renvoient à la réalité. Il compare ainsi la création originale d'un mot au travail d'un artiste peintre. Ce dernier utilise la couleur pour exprimer l'essence de son sujet dans une peinture ; de la même manière, le créateur de mots utilise des lettres contenant certains sons pour exprimer l'essence du sujet d'un mot. Ainsi certaines sonorités conviennent bien mieux aux

choses molles, d'autres aux choses liquides, etc...

Hermogène défend une théorie différente, selon laquelle les noms sont apparus en raison de la coutume, des usages et des conventions propres à chaque civilisation. Ils n'expriment pas l'essence de leur sujet et peuvent donc être échangés avec quelque chose sans rapport avec les individus ou les communautés qui les utilisent. La frontière entre les deux perspectives est ténue et le plus souvent floue. Pour cette raison, pendant plus de la moitié du dialogue, Socrate fait des suppositions à la demande d'Hermogène quant à l'origine des noms et des mots. Ces dernières mentionnent notamment les différents noms des dieux olympiens, et de nombreux mots qui décrivent des concepts abstraits.

En particulier il prend l'exemple de Cronos (*Kρόνος*) et Rhéa (*Ρέα*), le dieu du temps et sa femme, reine des Titans, qui présentent tous les deux en grec des noms de courant. Le terme grec *ῥεῦμα* peut faire référence à l'écoulement de n'importe quel élément ou matériau et

« la meilleure façon de parler consiste à utiliser des noms dont tous ressemblent aux choses qu'ils nomment, tandis que le pire est d'utiliser des noms de type opposé. »

n'est pas limité à celui de liquide. Le flux inarrêtable est alors l'image de la fonction du dieu Cronos, image qui se retrouve dans le nom qu'il lui est attribué. Il est d'ailleurs intéressant de noter que *ρόνη* en grec ancien qui peut se prononcer [rhoë] fut été traduit en français par "flux", une seule syllabe sans arrêt net comme pour symboliser cet écoulement.

Dans la suite du dialogue, de nombreux mots utilisés par Socrate ont changé au fil du temps et perdent parfois un lien entre leur sens et leur apparence. Pour ceux-là, il suppose qu'ils viennent d'origine étrangère ou ont tant évolué qu'ils en ont perdu toute ressemblance

avec le mot d'origine qui, lui, respectait ce lien entre forme et sens.

La dernière des trois théories entre les noms et l'objet désigné est soulevée par Cratyle, un disciple d'Héraclite, qui croient que les mots proviennent d'origine divine, ce qui les rendrait nécessairement corrects. Mais Socrate réfute cette théorie en relevant auprès de Cratyle certains mots dont le lien entre la forme et ce qu'ils cherchent à signifier est imparfait, chose impossible si cette tâche avait été accomplie par les dieux

«les noms ont été si tordus de toutes sortes de façons, que je ne devrais pas être surpris si l'ancienne langue, comparée à celle actuellement en usage, nous apparaît comme une langue barbare»

eux même.

Ainsi, la question des mots et leur phonétique débute dès l'époque de Platon, tant on se rend compte de la complexité du langage et de ses méthodes multiples pour évoquer chez le locuteur des sensations propres au sujet.

# Cartier





**Cartier**



*Statue du Garde de  
l'entrée du Temple  
De Brihadishwa-  
ra dans la ville de  
Tanjore en Inde*

# Les Upanishads

Le Veda vient du mot "connaissance" en langue hindoue, et désigne un recueil de textes transmis par tradition orale au sein des religions successives du sous-continent indien, du védisme à l'hindouisme. Ces textes nous sont parvenus d'aujourd'hui et leur date d'élaboration reste incertaine, certains experts

Les textes qui nous intéressent aujourd'hui sont les Upanishads, rédigés avec les Aranyaka lors de la transition du brahmanisme à l'hindouisme, et mis en forme par le sage Vyāsa. Ces textes constituent un socle important de connaissance dans l'hindouisme, et abordent aussi bien des sujets de philosophie,

*“Les consonnes muettes représentent la terre, les sifflantes le ciel, les voyelles le ciel. Les consonnes muettes représentent le feu, les sifflantes l'air, les voyelles le soleil... Les consonnes muettes représentent l'œil, les sifflantes l'oreille, les voyelles l'esprit. ”*  
“Aitareya Aranyaka III.2.6.2”

mentionnant entre le XVème siècle avant JC, jusqu'à -4000 av. JC. Cette tradition du chant « védique » fut introduite en 2008 par l'UNESCO sur la « liste représentative du patrimoine culturel immatériel de l'humanité ».

d'astronomie, d'étude de la nature, et de la langue dans notre cas. Ainsi, les Upanishads contiennent beaucoup de matériel sur le symbolisme sonore, comme l'évoque le chapitre nommé "Aitareya Aranya-ka III.2.6.2".



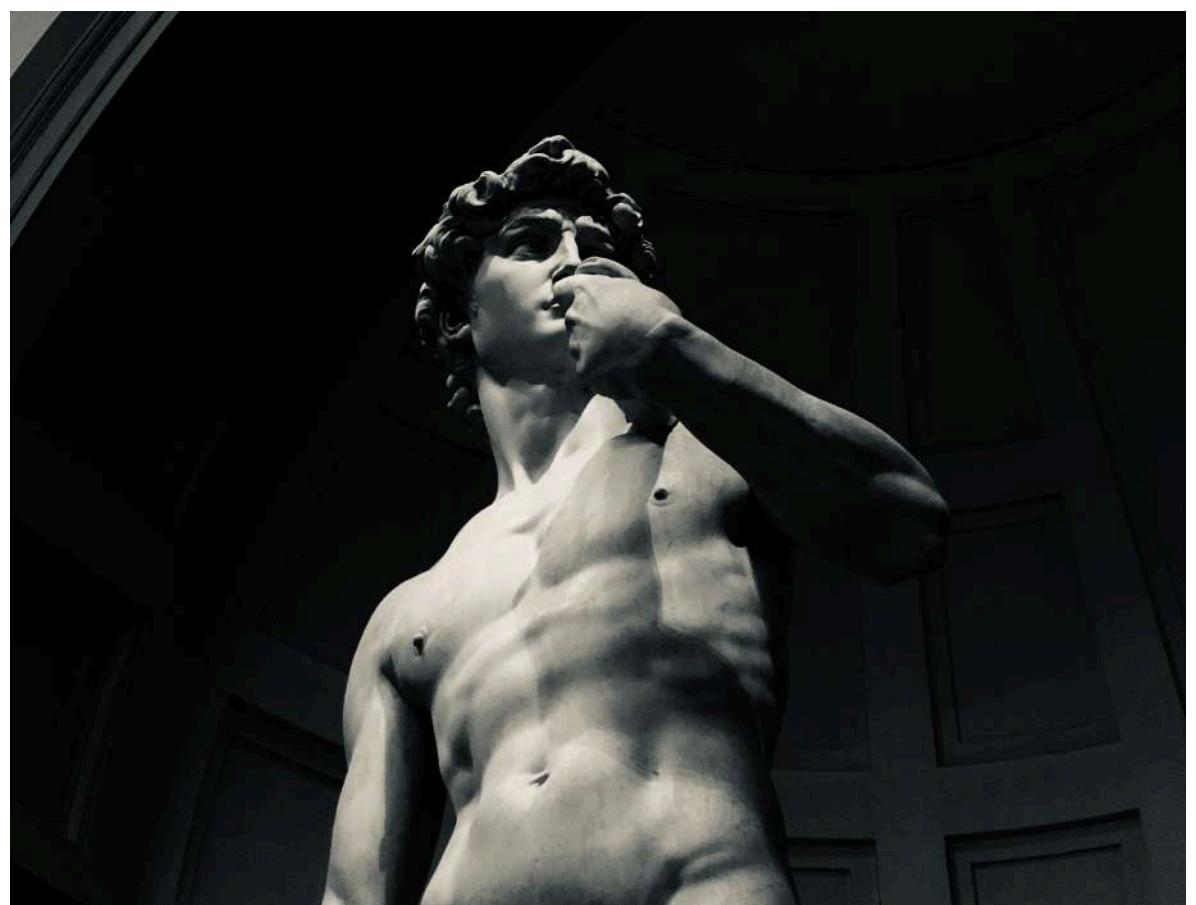
# Locke et Leibniz

Au Moyen Age et à la Renaissance, cette idée de phonoméntique et d'une correspondance entre signifiant et signifié est discuté sporadiquement entre différents auteurs. En 1690, Locke écrit à l'encontre de cette théorie dans "An Essay Concerning Human Understanding" en mettant en avant l'argument que si il existait un lien entre sons et idées, alors nous parlerions tous la même langue sur le globe, image certes poussée à son paroxysme mais résumant le fond de sa réflexion.

Leibniz lui répond en 1765 dans son livre "Nouveaux essais sur l'entendement humain" publié en 1765 et contenant une critique point par point de l'essai de Locke. Leibniz reprend la généralisation de Locke et adopte une approche moins rigide que ce dernier : il semble évident qu'il n'existe pas

de correspondance parfaite entre les mots utilisés couramment et les choses qu'ils sont censés désigner. Mais la relation bien qu'elle semble vague et difficilement définissable, n'est pas non plus complètement arbitraire. Cette position sera confortée par de nombreuses recherches au XXème et étudiée plus en profondeur.

Les études et réflexions que nous venons de voir sur le sujet constituent pour le moment une approche empirique du sujet. Ce n'est qu'au XXème siècle, que certains psychologues et philosophes se proposent de définir plus précisément notre langue, sa manière de fonctionner afin d'assurer un socle plus solide pour les études à venir. Ces études constituent le dossier suivant, qui se propose de définir ces concepts linguistiques.



*Le David est un chef-d'œuvre de la sculpture de la Renaissance, réalisé par Michel-Ange entre 1501 et 1504. Il mesure 5,17 mètres de hauteur et il est tiré d'un bloc de marbre blanc de Carrare, laissé à l'abandon après l'échec d'autres sculpteurs.*





# 3

## Entrons dans les détails techniques

*Chapitre 1. Le principe de l'arbitraire du signe*

*Chapitre 2. Incompatibilité partielle entre l'arbitraire et le symbolisme phonétique*

*Chapitre 3. La théorie de la double articulation*

*Chapitre 4. Cherchons la compatibilité*

# ARBIT

Ferdinand de Saussure est un linguiste suisse, universellement reconnu comme le « père de la linguistique moderne » et un précurseur du structuralisme linguistique qu'il a activement contribué à élaborer en redéfinissant de nouvelles bases solides au sein de cette discipline. La première de ses deux grandes contributions est qu'il clarifie la tâche de la linguistique en définissant la langue comme un système d'entités dont il précise les différentes relations. La deuxième position de Saussure dans la linguistique et la sémiotique moderne est bien loin de faire autant l'unanimité : c'est le « **principe de l'arbitraire du signe** ». Avec l'émergence de la linguistique cognitive ainsi que d'autres disciplines, ce principe fut scruté à la loupe et s'est

heurté à de nombreux défis dont celui du symbolisme phonétique. Il est ainsi important dans un premier temps d'aborder ce principe..

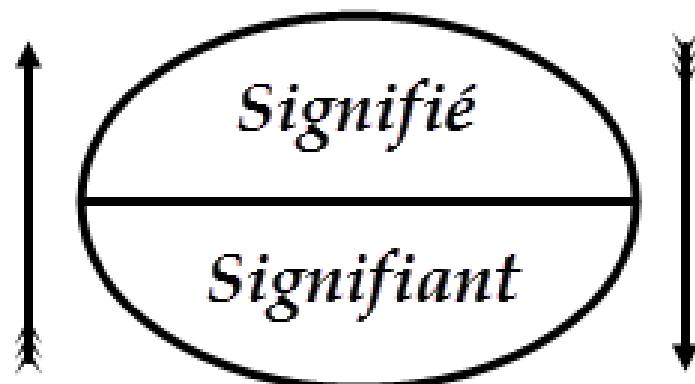
Saussure commence par définir deux termes employés tout au long de l'étude : « signifiant » et « signifié », qui constituent les deux briques principales du langage.

Le « signifié » désigne le concept, la représentation mentale d'une chose. Ce n'est pas un « objet » bien que les deux soient étroitement liés en ce sens que lorsque nous parlons d'un objet, il invoque dans notre esprit l'image de cette objet, la représentation que nous avons de celui-ci. Mais « concept » est bien plus approprié pour couvrir

entièrement le langage puisqu'il permet aussi de désigner des objets imaginés ou des idées abstraites telles que « Dieu », la « valeur » ou le « temps ».

Le « signifiant » quant à lui est le terme employé par Saussure pour l'image sonore d'un « signifié ». Ce n'est pas exactement un son physique mais plutôt « l'impression psychologique que l'auditeur a d'un son, telle qu'elle lui est donnée par les preuves de ses sens » (Saussure).

Ces deux éléments sont tous les deux de nature psychologique ce qui fait du « signe linguistique » de Saussure une entité à deux faces, ce qu'il résume dans le schéma ci-contre.



# RAIRE

Sur ces bases, Saussure avance ainsi sa théorie : le lien entre « signifiant » et « signifié » est arbitraire et immotivé. En d'autres termes, il n'existe aucun lien entre le sens d'un mot ou d'un concept ainsi que l'image que nous avons de ce dernier, et sa sonorité quelle qu'elle soit. Il prend ainsi l'exemple du signifié "bœuf" employé dans les différents cantons suisses, qui a pour signifiant [Böf] dans les cantons qui parlent français, [Oks] (Ochs) dans ceux à dominante allemande, et [Mäzo] (Manzo) pour les Italiens. De nombreux autres exemples issus de langues diverses étayent son argumentaire tout au long de son ouvrage.

Ainsi, pour résumer les trois points fondamentaux qui sont inclus dans la définition de l'ar-

bitraire par Saussure sont : un signe linguistique divisé en deux éléments, le « signifiant » et le « signifié » [1], ces deux éléments sont tous les deux psychologiques ce qui fait d'un signe une entité psychologique à deux faces [2]. Enfin le rapport entre ce « signifiant » et « signifié »

est complètement arbitraire et non-motivé. En d'autres termes, une fois que tous les membres d'une communauté se sont mis d'accord pour utiliser de façon similaire une relation entre un signal sonore et sa signification, ce lien n'est soumis à aucune influence personnelle.



*Musée Rodin  
dans le 7ème  
arrondissement de Paris*



A

## Incompatibilité partielle entre le principe de l'arbitraire et le symbolisme phonétique

Malgré les travaux de Saussure, un certain nombre de phénomènes découverts lors de recherches sur le sujet au XXème siècle, nous montrent qu'il existe un lien psychologique entre la nature articulatoire d'un mot et le « signifié ». Ainsi lorsque l'on demande à un échantillon de sujet quelle est la lettre la plus grande entre le "a" et le "i", la réponse est quasiment unanime pour qualifier le "i" de plus « petit ». On peut aussi évoquer l'expérience Takete-Maluma précédemment décrite, dont Davis montre en 1961 que le phénomène se reproduit à l'identique chez des sujets parlant des langues non indo-européennes

(Swahili et Kitongwe).

En énumérant ces expériences, et sans développer sur celle de Newman en 1933, celle de Chastaing en 1958, celle de Brown, Black et Horowitz en 1955, celle Usnadze en 1924, on se rend compte au premier abord que la théorie Saussurienne est incompatible avec son homologue du symbolisme phonétique.

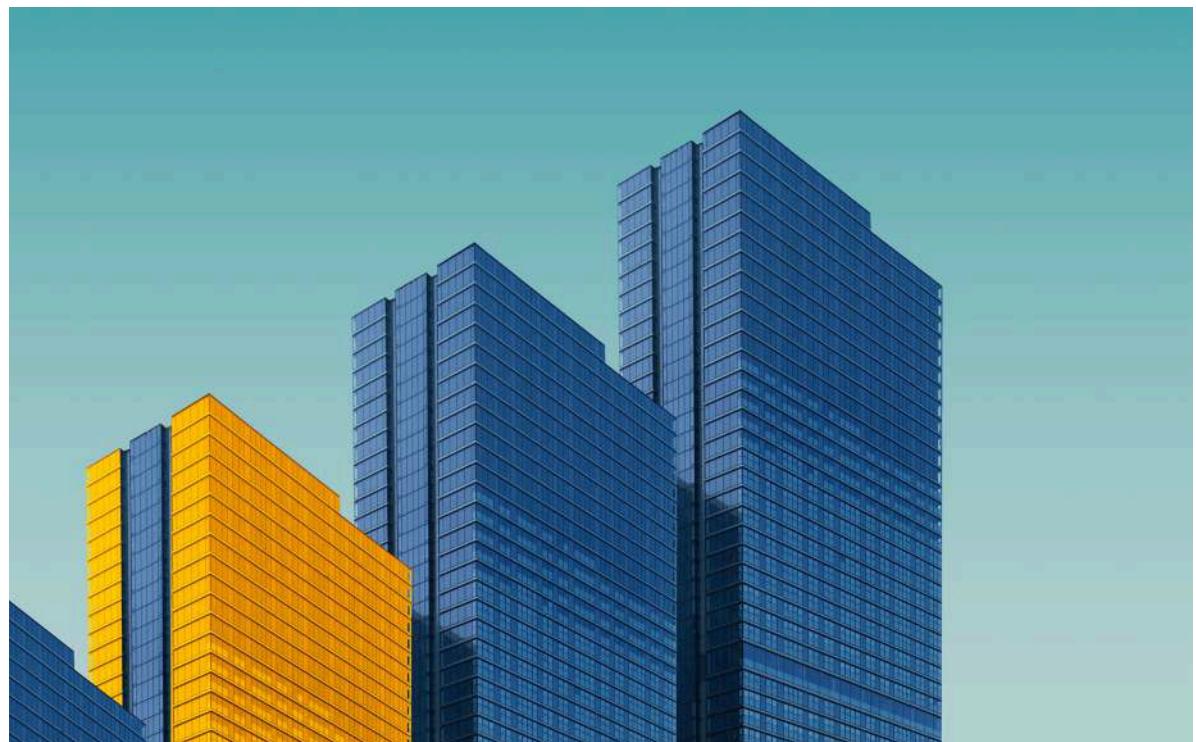
Mais peut-on réconcilier ces deux théories ?

Tout d'abord, commençons par souligner le fait que l'hypothèse élaborée par Saussure est en réalité un véritable outil et instrument de l'analyse linguistique. Le danger principal serait d'essayer d'appliquer la phonosémantique à l'analyse des langues directement ; bornons-nous pour le moment à la constatation du phénomène psychologique. En effet, supposer des le départ qu'il existe un lien entre « signifié » et « signifiant » fausserait

tout le raisonnement, puisqu'on pourrait construire un mode communication où tout « signifiant », c'est-à-dire toute suite sonore, aurait un sens qui lui est propre. En d'autres termes, cela reviendrait à construire le langage en partant de la forme vers le sens, ce qui est en contradiction avec les capacités cognitives finies de l'homme. L'infinie d'expériences vécues serait alors rattachée une à une à une infinité d'associations sonores ce qui dépasserait les capacités mémorielles d'un individu ordinaire.

Commençons donc par supposer que le langage est arbitraire. Si on fait cette hypothèse, alors comment se construisent les mots de notre langue ? A Martinet apporte une réponse à cette question et définit la notion de double articulation qui permet de décomposer la chaîne parlée en unités de langage : les phonèmes et les monèmes.

*Des immeubles de différentes tailles, comme certaines lettres de l'alphabet*



# La Double Articulation

La double articulation est un concept linguistique qui rend compte de l'utilisation de deux types d'unités, celles significatives d'une part (unité de première articulation) et distinctives

d'autre part (unité de deuxième articulation). Selon André Martinet, la langue s'organise donc sur deux niveaux qui ne remplissent pas les mêmes caractéristiques.

Les morphèmes (ou unités de première articulation) constituent les plus petites associations de lettres qui conservent un sens. Ce sont des verbes (courir, danser), des noms communs (feuille, lampe) mais aussi

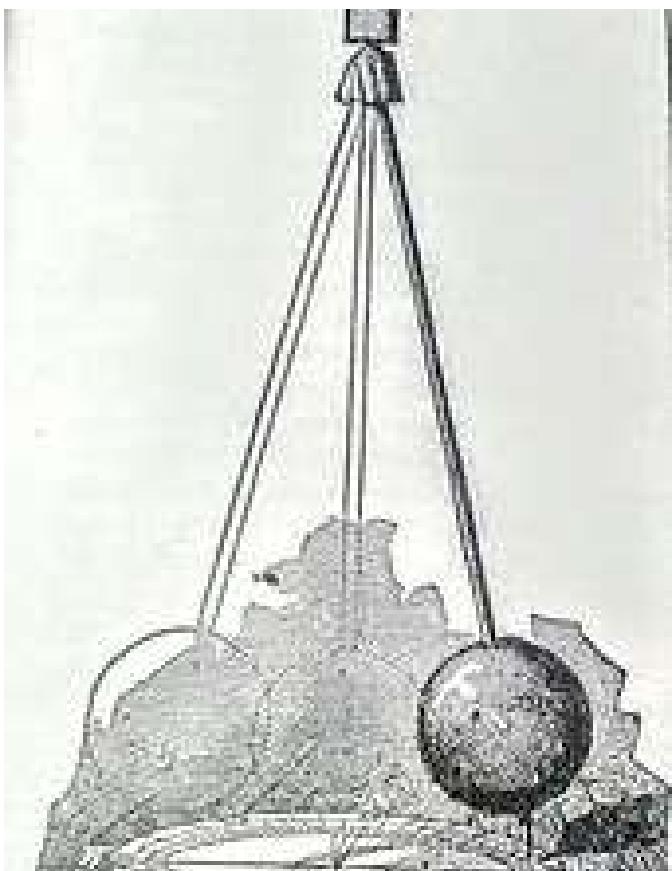
certaines parties de mots comme le “-ons” dans le verbe « assisterons ». En effet, chaque lecteur sait précisément avec ces 3 lettres que le sujet du verbe en question est une première personne du pluriel (nous). On donne ainsi les

exemples suivants, il y a 2 morphèmes dans “je cours” et 3 morphèmes dans “nous lisons”.

Les phonèmes (unités de seconde articulation) sont des sons différents les uns des autres mais qui n'ont pas de sens particulier. Ainsi, ils sont constitués d'un signifiant mais pas de signifié. On peut par exemple trouver 2 phonèmes dans le mot pont et thon respectivement.

Avec ces deux niveaux de lecture, la langue peut optimiser son système en représentant un grand nombre d'images distinctes avec une seule modification de phonème ; on remplace un seul son plutôt que de créer un mot différent. C'est cette organisation du langage qui constitue la différence majeure entre le système de communication humain et ceux employés par les animaux.

Mais revenons à notre sujet, une fois ces bases définies un peu laborieusement, cela n'explique pas comment se lient les mots et les sons, tout en assurant une compatibilité avec la théorie de Saussure.



Pendule à un degré d'articulation. Il existe des pendules à deux degrés d'articulation.



*Un peu comme en montagne,  
la langue française est munie  
de plusieurs niveaux d'organi-  
sation.*



Lorsque l'on étudie le symbolisme phonétique, il est frappant de voir à quel point le rapport entre son et stimuli sensoriel est bien plus fort que celui avec le « signification ». En effet, lors de l'apprentissage du langage dans les plus jeunes années, l'enfant associe chaque mot à une ou plusieurs expériences plutôt qu'à l'image pure de cet objet. Pour illustrer ce propos prenons quelques exemples ; le mot « croquant » renvoie immédiatement à cette sensation que nous éprouvons dans la bouche lorsque nous mangeons dont nous pourrions imiter le bruit par ... « krkrkr ». On retrouve par ailleurs notre histoire de taille des lettres précédente. Le « a » ou le « o »

semblent plus « grands » que la lettre « i » ou « e », ce qui peut s'expliquer plutôt simplement par l'aperture vocale nécessaire pour prononcer la lettre. Il est nécessaire d'ouvrir bien plus la bouche pour prononcer les deux premières lettres que les deuxièmes. Ainsi, intuitivement nous attribuons une taille différente aux lettres, la « motivation » indiquant un chemin vers plus grand à l'inverse de la « petitesse ».

Les deux exemples précédents confirment ce lien fort qu'il existe entre expérience sensorielle et le mot qu'on lui attribue.

## Compatibilité entre le «symbolisme phonétique» et le principe de l'arbitraire



Néanmoins si l'on suit notre raisonnement, nous nous heurtons à un ultime problème. Les expériences adoptent par définition un caractère continu. Cette continuité constituerait donc une propriété de ce symbolisme phonétique ; mais les unités de deuxième articulation, nos phonèmes, sont discontinues. En effet on forme nos syllabes dans les langues les plus communément parlées, toujours de la même manière. On verra ainsi souvent “-mal-” et non “-mla” que nous ne pouvons prononcer facilement et qui sera ainsi absent de notre langage. Les deux caractères semblent donc ne pas se juxtaposer et nous ne pouvons toujours pas réconcilier « symbolisme phonétique » et arbitraire du signe.

*Ensemble de ponts et passerelles, comme celle que l'on essaie de construire en ce moment entre ces deux théories*

Mais ce que montrent Taylor et Taylor en (1962), c'est que l'ordre des lettres importe en réalité bien peu dans le message sonore que porte une syllabe. Prenez par exemple la phrase suivante : « la suele coshe ipmrotnate est que la pmeirère et la drenière sioent à la bnnoe pclae ». En théorie, vous arrivez à lire ce texte, précisément parce que chaque mot conserve sa première et dernière lettre, le cerveau lit chaque mot comme un tout. Eh bien c'est la même chose avec les phonèmes, l'effet de chaque lettre au sein d'une unité de deuxième articulation est entièrement due aux lettres qui l'entourent. Ce caractère à une conséquence très importante, ce ne sont pas les « variantes » du phonème qui reçoivent chacune un caractère symbolique, mais le phonème lui-même.

C'est précisément cet élément qui nous permet de relier nos deux théories. Notre cerveau interprétant les phonèmes comme un tout, il les rend discontinues ! Elles ne le sont en réalité pas, puisqu'on pourrait former à l'infini une association de lettres qui n'aurait aucun sens comparé aux phonèmes précédentes, ce qui suffirait pour désigner cette nouvelle unité de deuxième articulation et assurer un caractère continu à la double Articulation par la même occasion. Mais notre cerveau arrangeant les lettres pour reconnaître ce qu'il connaît déjà, et on retombe ainsi sur un ensemble fini d'associations de lettre ayant un symbolisme phonétique.

## Conclusion de la partie 3:

En fin de compte, résumons notre raisonnement au travers des points suivants. Saussure établit que le lien entre « signifiant » et « signifié » est non motivé. Néanmoins, il existe un symbolisme phonétique, son existence ayant été démontrée expérimentalement, à la physicienne diront nous. Ce dernier est néanmoins cantonné aux secteurs sensoriels, c'est-à-dire que nous ne faisons aucun lien entre le sens « profond » d'un mot et sa sonorité. Mais nous assimilons cette phonétique à notre expérience personnelle que nous avons de cet objet, ce qui est possible grâce au fait que les phonèmes soient interprétés comme un tout par notre cerveau. Ainsi le symbolisme ne joue pas de rôle linguistiquement déterminant, mais agit propre à chacun et à ses capacités mémorielles.

Ouf ...

# ROLLS ROYCE





# 4

## Baladons-nous plus loin

*Chapitre 1. Vérification à  
réseau neuronal*

*Chapitre 2. Ecrivains, ces  
inventeurs de mots*

*Chapitre 3. Une danse finale  
avec Rimbaud*



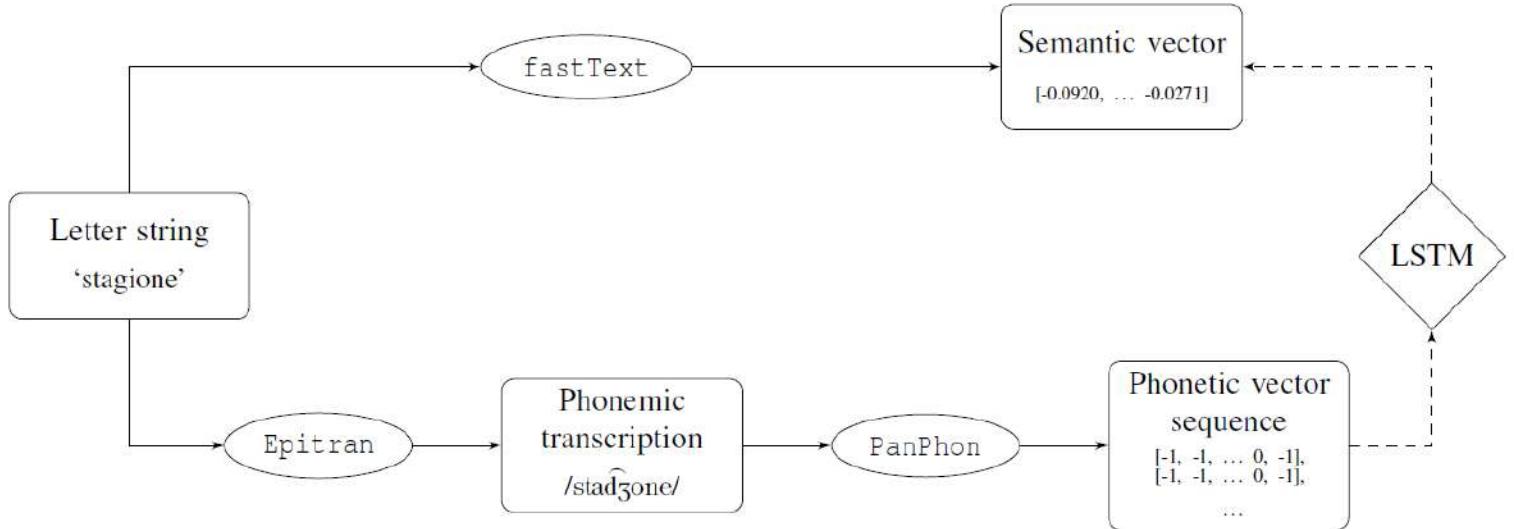


Figure 1: Schematic representation of the experimental pipeline

## Vérification à l'aide d'un réseau neuronal

Avec l'émergence de nouvelles technologies et en particulier les performances exponentiellement grandissantes de l'intelligence artificielle, une équipe de chercheurs a cherché à étudier en 2021 ce lien phono-sémantique mais entre différentes langues. Ils ont pour cela utilisés un réseau neuronal LSTM qui s'est entraîné sur 7 langues différentes sur les liens entre « signifiant » et « signifié » avec de passer au test sur une 8 langue non étudiée.

Dans l'étude considérée, le réseau neuronal LSTM (Long Short-Term Memory) est entraîné pour associer la représentation phonétique d'un mot à sa représentation sémantique. Pour cela, il convertit ces deux représentations en vecteurs et cherche le meilleur rapprochement.

Le vecteur sémantique est déterminé par le réseau en utilisant un outil de Wikipédia Data nommé fastText pour convertir la représentation sémantique du mot en un vec-

teur à 300 dimensions. Cet outil fonctionne sur plus de 15 langue différentes, et appartenant à des familles linguistiques différentes.

Les Vecteurs phonétiques sont obtenus en deux étapes. On utilise dans un premier temps une librairie python nommée Epitran qui convertit le mot dans l'alphabet international phonétique (AIP). Puis le mot en AIP est converti à l'aide du package Panphon en un ensemble de vecteurs à 22 dimensions.

Le rôle de l'architecture neuronale, construite sur la base du modèle de deep-learning sur Python Keras, est alors de créer un vecteur sémantique expérimental, toujours à 300 dimensions. On compare alors notre vecteur sémantique de base et celui expérimental et on observe si les deux sont éloignés ou pas.

L'architecture décrite précédemment est résumée dans le schéma en haut de page.

Photo d'une structure métallique prise en Italie qui rappelle l'architecture d'un réseau de neurones

## Resultats et discussion

Le tableau 2 rapporte les résultats des modèles, comparés à un vecteur sémantique aléatoire. Les deux premières colonnes du tableau précisent la condition expérimentale, les troisièmes et quatrièmes présentent la similarité moyenne cosinus entre la cible sémantique et l'expérimental (exp) ou l'aléatoire (r). Les deux dernières colonnes donnent deux facteurs de confiance. On voit clairement que parmi les 3 langues testées, appartenant à 3 familles linguistiques distinctes, tous les modèles expérimentaux ont surpassé leurs valeurs de référence randomisée respectives. Un modèle entraîné pour repérer des liens de symbolisme phonétique est donc efficace, ce qui valide en grande partie la théorie.

La deuxième partie de l'étude, dont nous ne développerons pas le fonctionnement technique ici est particulièrement intéressante. Elle s'attache à déterminer si ces relations phono-sémantiques sont uniformément réparties dans le vocabulaire ou non. Par une multitude de régressions linéaires et une analyse quantitative basée sur la théorie, l'étude montre ainsi que les mots reliés aux sens présentent un lien beaucoup plus fort avec leur phonétique ! Ceci confirme notre étude théorique du phénomène dans la partie précédente ;)

<b>Model</b>	<b>N<sub>train</sub></b>	<b>N<sub>test</sub></b>	<b>Cos<sub>exp</sub></b>	<b>Cos<sub>r</sub></b>	<b>t</b>	<b>p</b>
Monolingual	691259	172815	0.5105	0.3411	534.15	$\ll 0.001$
Mono-family	2742222	331475	0.3882	0.3383	242.59	$\ll 0.001$
Cross-family	2742222	558791	0.3192	0.2986	126.11	$\ll 0.001$

Table 2: Test results by experimental condition



*Toile de Paul BlenkHorn qui joue  
justement sur les sens et ici le tou-  
cher dans ses œuvres artistiques*

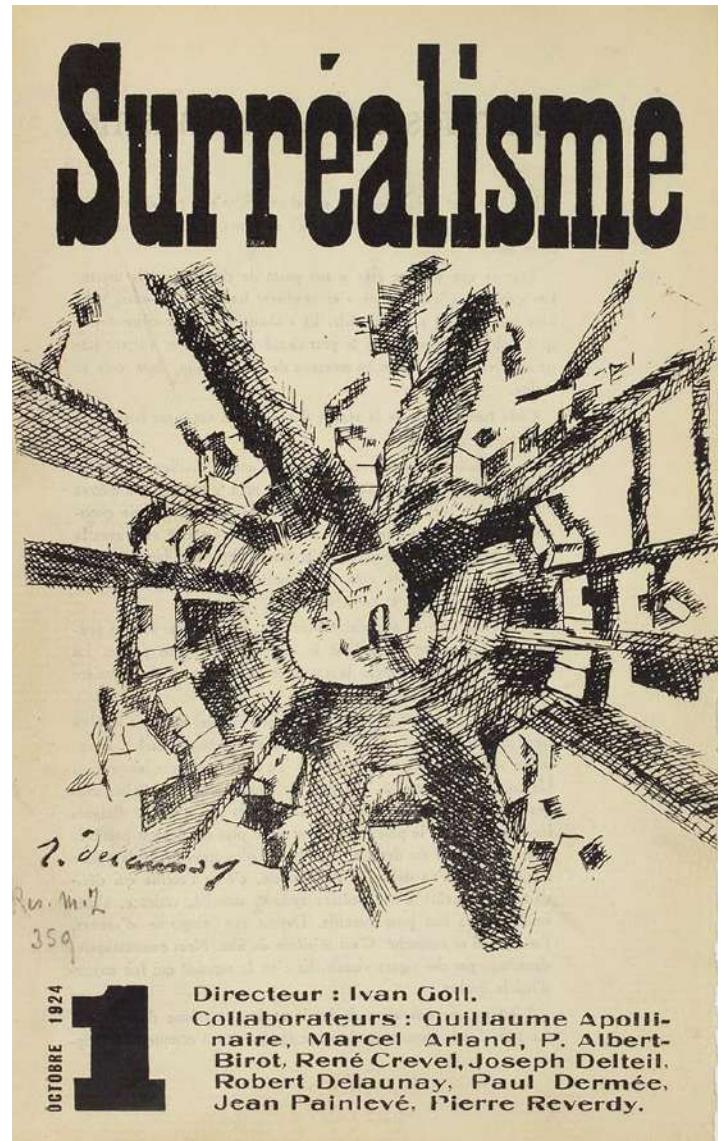
## Quand les écrivains inventent des mots

A toutes les époques, les auteurs n'ont cessé de jouer avec les sons pour créer de nouveaux mots, soit en cherchant à les intégrer volontairement au langage, soit involontairement ceux-ci s'y étant insérés à l'usage. San Antonio écrivait ainsi, « l'avenir du langage, c'est moi ». Revenons sur quelques contributions d'auteurs à l'enrichissement de la langue française.

Dans son ouvrage achevé en 1549, *Défense et illustration de la langue françaize*, Du Bellay encourage ses pairs à « innover quelque terme en un long poème ». « Plus nous aurons de mots dans notre langue, plus elle sera parfaite », renchérissait Ronsard (1524-1585).

Rabelais avait déjà ouvert la voie dans la première moitié du XVI<sup>e</sup> siècle, cherchant son inspiration dans les langues de toutes les époques, et devant ainsi l'un des plus célèbre « inventeur » de mots de la langue française. On peut lui attribuer les expressions « mouton de Panurge », indigène, gymnaste, automate ou encore catastrophe.

Ce dernier mot est mis en lumière dans la pièce « Rabelais, portrait d'un homme qui n'a pas souvent dormi tranquille », actuelle-



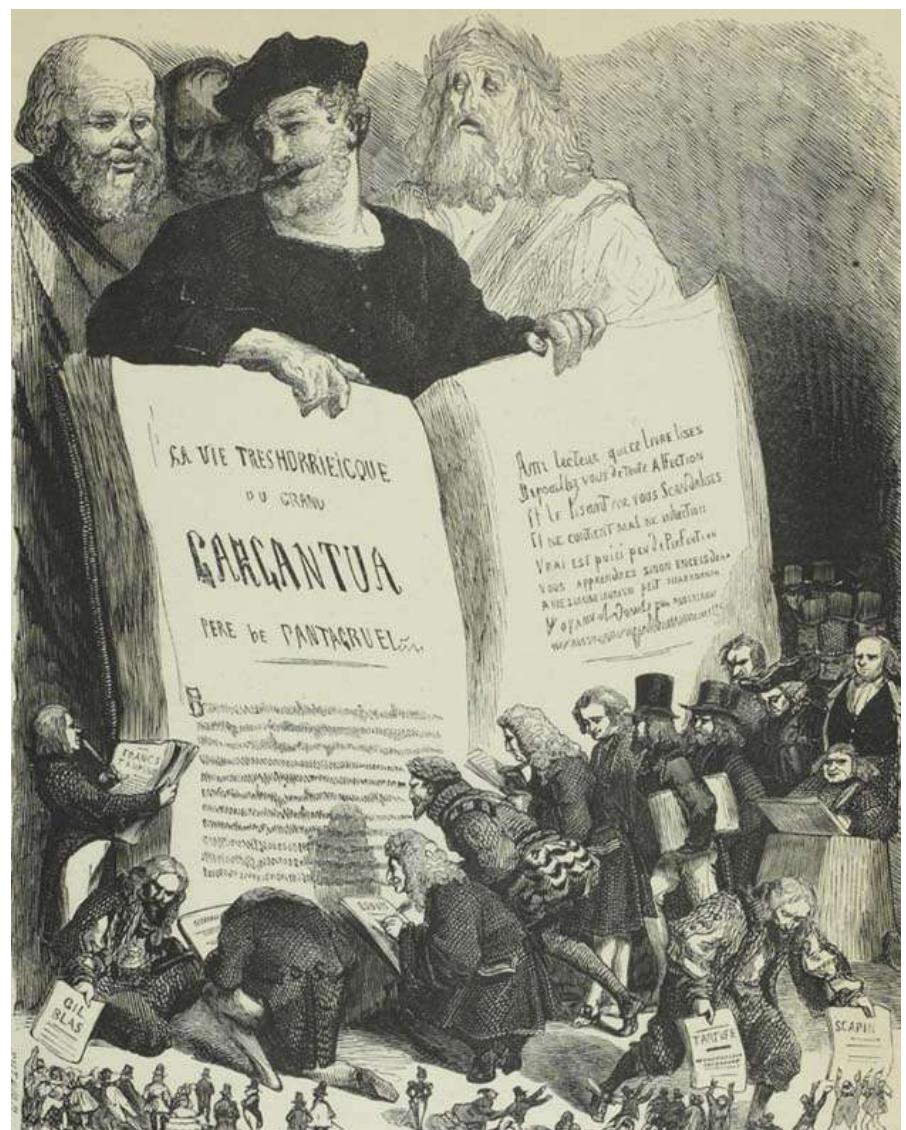
ment au théâtre de l'Essaïon à Paris. Jean-Pierre Andréani y incarne un Rabelais poursuivi par la police morale de la Sorbonne qui veut le faire taire. Ce dernier à la fin de la pièce décrit la création du mot « catastrophe » comme quelque chose qui s'écroule dans un fracas effroyable, la phonétique de ce dernier lui donnant presque son sens naturellement.

Nous vous recommandons cette pièce que nous sommes allés voir et d'une grande qualité.

*Couverture du premier numéro de Surrealisme, illustrée par Robert Delaunay (1924)*



Véritable portrait de Monsieur Ubu.



Depuis, ses héritiers ont suivi le même chemin ... parfois par erreur ! Le mot compatissance né de la plume de Balzac fait réagir André Gide dans son journal : « « Il me semble que compassion suffisait ». Mais l'artiste est maître en la demeure et il a tous les droits, y compris celui de se tromper !

Citons par ailleurs le surréalisme de Guillaume Apollinaire qui voit le jour en 1917 et dont on ne soupçonnerait le si jeune âge aujourd’hui.

Enfin, le mot ubuesque provient du personnage d’Alfred Jarry, le père Ubu. Le personnage d’Ubu est ainsi devenu proverbial, symbole du délire du pouvoir et de l’absurdité des hiérarchies politiques.

## Un petit jeu pour finir

Concluons ce court voyage au pays des mots par un peu de poésie. Après avoir lu ce poème d'Arthur Rimbaud composé en mai 1871, tentez de répondre aux questions suivantes :

1. Quand Rimbaud parle de son cœur, quel est le principal sentiment qui ressort du poème ?
2. Dans le quatrième quatrain, retrouvez-vous un élément de symbolisme phonétique ?
3. Le dernier quatrain vous permet-il de déterminer si Rimbaud parle de sursauts heureux ou malheureux avec la phonétique ?
4. Que vous évoque le dernier vers du poème ? Est-ce un appel à l'espoir ?



## Le Cœur volé

Mon triste cœur bave à la poupe,  
Mon cœur couvert de caporal :  
Ils y lancent des jets de soupe,  
Mon triste cœur bave à la poupe :

Sous les quolibets de la troupe  
Qui pousse un rire général,  
Mon triste cœur bave à la poupe,  
Mon cœur couvert de caporal !

Ithyphalliques et pioupiesques,  
Leurs quolibets l'ont dépravé !  
Au gouvernail on voit des fresques  
Ithyphalliques et pioupiesques.

O flots abracadabantesques,  
Prenez mon cœur, qu'il soit lavé !  
Ithyphalliques et pioupiesques,  
Leurs quolibets l'ont dépravé !

Quand ils auront tari leurs chiques,  
Comment agir, ô cœur volé ?  
Ce seront des hoquets bachiques  
Quand ils auront tari leurs chiques :

J'aurai des sursauts stomachiques,  
Moi, si mon cœur est ravalé :  
Quand ils auront tari leurs chiques  
Comment agir, ô cœur volé ?

Arthur Rimbaud



## Resultats et discussion

Pour apporter une réponse aux questions précédentes, dans le désordre, on peut s'intéresser à la signification probable de ce poème pour Rimbaud. Bien qu'elles ne suscitent pas l'unanimité deux thèses émergent des débats. Le vocabulaire employé correspond pour un bon nombre de mot à de l'argot ayant une connotation sexuelle forte. Une première thèse avance ainsi le souvenir de sévices sexuelles infligées à Rimbaud par les communards. Rimbaud a en effet partagé sa grande déception à l'égard du soulèvement lors de sa visite dans la capitale en avril 1871. Parfois jugée anti-commune, certains avancent une autre hypothèse, celle du viol quelques mois plus tôt de l'auteur à la prison de Mazas. Quoi qu'il en soit, le souvenir douloureux de l'épisode mentionné est marqué par le vocabulaire au sonorité hachées, aux consonnes croquantes et à l'utilisation judicieuse de lettres de « taille » différente pour imager le récit.

**FIN**

