
Calculer la sémantique avec le langage IEML

Semantic Computing with the IEML Language

Pierre Lévy



Édition électronique

URL : <https://journals.openedition.org/revuehn/3836>
ISSN : 2736-2337

Éditeur

Humanistica

Référence électronique

Pierre Lévy, « Calculer la sémantique avec le langage IEML », *Humanités numériques* [En ligne], 8 | 2023, mis en ligne le 01 décembre 2023, consulté le 19 décembre 2023. URL : <http://journals.openedition.org/revuehn/3836>



Le texte seul est utilisable sous licence CC BY 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.



Calculer la sémantique avec le langage IEML

Semantic Computing with the IEML Language

Pierre Lévy

Résumés

Cet article présente IEML (*Information Economy MetaLanguage*), un système de représentation uniforme de la signification et de la connaissance humaine qui peut être lu et traité automatiquement par les machines. Distinguée des sémantiques pragmatique et référentielle, la sémantique linguistique est aujourd'hui formalisée de manière incomplète. Seule sa dimension syntagmatique a été mathématisée sous la forme des langages réguliers. Il restait à formaliser sa dimension paradigmatic. Pour résoudre le problème de la mathématisation complète du langage, y compris sa dimension paradigmatic, je propose de coder le sens linguistique en IEML. IEML a la même capacité expressive qu'une langue naturelle et possède une structure algébrique permettant le calcul de sa sémantique. L'article explique son dictionnaire, sa grammaire formelle et ses outils intégrés de construction de graphes sémantiques. Au titre de ses applications, IEML pourrait être le vecteur d'un calcul et d'une communication fluide du sens – l'interopérabilité sémantique – capable de décloisonner la mémoire numérique et de nourrir les progrès de l'intelligence collective, de l'intelligence artificielle et des humanités numériques. Je conclus en indiquant quelques directions de recherche. Ce texte présente la synthèse de plusieurs décennies de recherches.

This paper presents IEML (*Information Economy MetaLanguage*), a uniform representation of human meaning and knowledge that can be read and processed automatically by machines. Distinguished from pragmatic and referential semantics, linguistic semantics is today incompletely formalised. Only its syntagmatic dimension has been mathematised in the form of regular languages. Its paradigmatic dimension remained to be formalised. To solve the problem of the complete mathematising of lan-

guage, including its paradigmatic dimension, I propose to code the linguistic meaning in IEML. IEML has the same expressive capacity as a natural language and has an algebraic structure allowing the calculation of its semantics. The article explains its dictionary, its formal grammar, and its integrated tools for building semantic graphs. As far as its applications are concerned, IEML could be the vector of a fluid calculation and communication of meaning – semantic interoperability – capable of decompartmentalising the digital memory and feeding the progress of collective intelligence, artificial intelligence, and digital humanities. I conclude by indicating some research directions. This paper presents a synthesis of several decades of research.

Entrées d'index

MOTS-CLÉS : linguistique et sciences du langage, intelligence collective, apprentissage automatique

KEYWORDS: linguistics and language sciences, collective intelligence, machine learning

Introduction

1 Comme l'ont fait l'écriture, l'imprimerie et les médias électroniques dans les siècles passés, la croissance des capacités de mémoire et de calcul amenée par le numérique a augmenté la puissance du langage et, du même coup, les capacités de raisonnement, de narration et de dialogue qu'il supporte. Le projet explicite d'augmenter l'intelligence humaine au moyen des technologies numériques est aussi ancien que le numérique lui-même, comme le montrent Bush (1945), Licklider (1960) et Engelbart (1962). IEML prolonge les travaux de ces pionniers. Il peut aussi être considéré comme la réalisation de projets bien plus anciens tels que la caractéristique universelle de Leibniz (1989) : une langue idéographique formelle, supportant l'arrangement diagrammatique, le calcul et l'invention des concepts. Les références qui précèdent ont pour but de situer mon entreprise dans la tradition intellectuelle de l'augmentation de l'intelligence humaine, par opposition à celle de l'imitation ou de la simulation. Je parlerai comme tout le monde d'intelligence artificielle (IA), mais en entendant cette expression dans une perspective d'augmentation cognitive des humains plutôt que d'autonomie des machines.

2 Sur un plan scientifique, la création d'une langue à la sémantique calculable se place au croisement de la philosophie, de la linguistique, de l'intelligence artificielle dite « symbolique » et de l'intelligence collective. Une telle approche transdisciplinaire peut prêter à de nombreux malentendus. Afin d'en dissiper quelques-uns dès l'abord, je voudrais préciser ce que j'entends par les mots « sémantique » et « concept ». La sémantique ne se laisse pas enfermer dans le champ linguistique et dépend évidemment du contexte pratique, des modes de vérification en jeu et des sujets réels impliqués dans la saisie du sens. Lorsqu'une parole est prononcée, elle prend sens au moins sur quatre plans. Celui de la représentation mentale commandée par sa structure grammaticale et le sens de ses mots (un réseau de concepts), le plan logique de sa référence à un état de choses (elle est vraie ou fausse) et le plan pratique de l'interaction sociale (c'est un coup

dans un jeu de langage). Ces trois sens – linguistique, référentiel et pragmatique – font écho au trivium de l'Antiquité romaine et du Moyen Âge occidental : grammaire, dialectique et rhétorique¹. Finalement, le quatrième plan de la sémantique, que l'on pourrait appeler phénoménologique – c'est celui des schémas moteurs, des images sensibles et des émotions –, s'enracine dans la mémoire et l'expérience personnelle plutôt que dans les structures de la langue ou d'autres conventions sociales. Ces quatre plans codépendants déterminent la sémantique globale d'un énoncé. Quand je parle de « calculer la sémantique », je vise uniquement le plan linguistique puisque les niveaux logiques et pragmatiques sont déjà largement formalisés et que le niveau phénoménologique semble ne pas pouvoir l'être. On pourrait s'étonner de ce que je place les concepts au niveau linguistique. En effet, la conceptualisation est bien attestée dans la cognition animale, elle existe donc indépendamment de la symbolisation. Mais le langage permet aux concepts – c'est-à-dire aux catégories générales – de se réfléchir tout en multipliant leur nombre et leur complexité. Qu'est-ce que le signifié d'un symbole linguistique, sinon la réflexion d'un concept par un signifiant ? Lorsque je parle de concepts, je ne réfère donc pas aux signes naturels fixes d'un langage de la pensée inné à la Guillaume d'Oc-cam ou à la Jerry Fodor, mais aux pendants cognitifs des signifiés d'une langue : étoiles de relations à l'intersection de plusieurs systèmes de différences, qui se transforment selon les contextes ou les ontologies locales. Les mots « concept » et « signifié » fonctionnent ici comme les deux faces – cognitive et linguistique – de la même pièce. Je précise en outre que les concepts ne sont pas seulement évoqués par des mots, mais aussi par des phrases, en particulier dans les définitions.

³ Le problème que je m'étais proposé de résoudre était de formaliser la sémantique linguistique. C'est pourquoi je suis parti d'un modèle opératoire simple. Mettons de côté pour le moment les obstacles de l'ambiguïté et des malentendus pour schématiser le processus principal. Du côté de la réception, nous entendons une séquence de sons que nous traduisons en un réseau de concepts, conférant ainsi son sens à une expression linguistique. Du côté de l'émission, à partir d'un réseau de concepts que nous avons à l'esprit – un sens à transmettre – nous générons une séquence de sons. Selon ce modèle, une langue fonctionne comme une interface entre des séquences de sons et des réseaux de concepts. Les chaînes sonores peuvent être remplacées par des séquences d'idéogrammes, de lettres ou de gestes comme dans le cas de la langue des signes. L'interfaçage quasi automatique entre une séquence d'images sensibles (sonores, visuelles, tactiles) et un graphe de concepts abstraits (catégories générales) reste constant parmi toutes les langues et l'ensemble des systèmes d'écritures.

⁴ Cette traduction réciproque entre une suite d'images (le signifiant) et des réseaux de concepts (le signifié) est le principe même du modèle formel d'IEML, à savoir une catégorie² mathématique organisant une correspondance entre une algèbre et une structure de graphe. L'algèbre règle les opérations de lecture et d'écriture sur les textes séquentiels, tandis que la structure de graphe organise les opérations sur les nœuds et les liens orientés des réseaux de concepts. À chaque texte correspond un réseau de concepts et les opérations sur les textes reflètent dynamiquement les opérations sur les graphes conceptuels.

5 Pour coder mathématiquement des chaînes de signifiants, j'ai utilisé un langage régulier (Chomsky 1957 ; Chomsky et Schützenberger 1963). Un tel formalisme permet de transformer automatiquement les séquences de symboles en arbres syntagmatiques – reflétant la structure de dépendance entre les mots d'une phrase – et *vice versa*. Cependant, si son arbre syntagmatique est indispensable à la compréhension du sens d'une phrase, il n'est pas suffisant. En effet, chaque expression linguistique se trouve à l'intersection d'un axe syntagmatique et d'un axe paradigmatique³. L'arbre syntagmatique représente le réseau sémantique interne d'une phrase, l'axe paradigmatique représente son réseau sémantique externe et en particulier ses relations avec des phrases ayant la même structure, mais dont elle se distingue par quelques mots différents. Pour comprendre la phrase « Je choisis le menu végétarien », il faut bien sûr reconnaître que le verbe est « choisir », le sujet « je » et l'objet « le menu végétarien » et savoir en outre que « végétarien » qualifie « menu ». Mais il faut aussi connaître *le sens des mots* et savoir, par exemple, que végétarien se distingue de carné, de flexitarien et de végétalien, ce qui implique de sortir de la phrase pour situer ses composantes dans des systèmes de taxonomie et d'oppositions sémantiques, ceux de la langue comme ceux de divers domaines pratiques. L'établissement de relations sémantiques entre concepts suppose donc que l'on reconnaisse les arbres syntagmatiques internes aux phrases, mais aussi les matrices paradigmatiques externes à la phrase. C'est pourquoi IEML code algébriquement non seulement les arbres syntagmatiques, mais également les *matrices paradigmatiques* où les mots et les concepts prennent leur sens. En somme, chaque phrase d'IEML se situe à l'intersection d'un arbre syntagmatique et de matrices paradigmatiques. Outre une grammaire régulière, IEML repose sur un dictionnaire d'environ 3 000 mots – sans synonymes ni homonymes – organisés en un peu plus d'une centaine de paradigmes. Ces mots ont été choisis de manière à permettre la construction récursive de n'importe quel concept ou paradigme de concepts au moyen de phrases. Soulignons que, par contraste avec les langues naturelles, qui sont ambiguës et irrégulières, les expressions en IEML sont univoques. Sur la base de la grille syntagmatico-paradigmatique d'IEML, il devient alors possible de générer et de reconnaître des relations sémantiques entre concepts *de manière fonctionnelle* – donc programmable – et c'est pourquoi je parle de calcul sémantique.

6 L'approche du sens comme réseau de concepts se limite au niveau linguistique d'une sémantique à quatre couches. Elle est choisie pour des raisons de méthode parce qu'elle permet la formalisation et qu'elle consonne avec les graphes de connaissances et ontologies qui sont les applications privilégiées d'IEML. Ce choix n'implique nullement l'invalidation des approches interprétatives, éenactives ou phénoménologiques. Puisque mon but était de formaliser le sens linguistique, j'ai choisi un modèle algébrique de type chomskien. Toujours pour la même raison, l'esprit de la grammaire d'IEML est proche de la sémantique de Montague (1970). En effet, la signification d'une phrase IEML est déterminée strictement par la manière dont ses constituants – mots et sous-phrases – sont combinés entre eux (principe de compositionnalité) et les dialectes ou ontologies IEML sont compatibles avec la théorie des types logiques. Néanmoins, il n'a jamais été question pour moi de choisir une école formaliste ou centrée sur la syntaxe *contre* d'autres écoles, mais bien au contraire d'intégrer

les divers apports de la linguistique théorique afin de construire l'outil le plus utile et le plus riche possible. Comme le lecteur le découvrira, la conformité d'IEML à des modèles logico-mathématiques rigoureux n'efface en rien sa dette à l'égard des tendances structuralistes (Saussure 1916 ; Hejlslev 2000 [1968] ; Rastier 1987) et cognitives au sens large (Tessnière 1959 ; Fillmore 1968 ; Langacker 1987 ; Melchuk 1998 et l'école sens-texte ; Lakoff 1987).

7 Quelles pourraient être les applications d'IEML ? Puisque la communication et la coordination pratique des communautés humaines s'effectuent de manière croissante par l'intermédiaire d'une mémoire numérique partagée, ma vision pour IEML est de favoriser l'interopérabilité sémantique des mémoires numériques et d'augmenter ainsi l'intelligence collective (Lévy 1994, 2011, 2021 ; Mulgan 2017 ; Baltzersen 2022). Compte tenu du rôle du langage dans l'intelligence humaine, l'intelligence artificielle pourrait bénéficier d'un saut qualitatif grâce à un perfectionnement de la catégorisation des données et à une nouvelle calculabilité de la sémantique. Une langue compréhensible par les humains tout en étant exécutable par les machines apporterait aux contrats intelligents, et aux divers systèmes de *blockchain*, la transparence qui leur manque aujourd'hui. Dans le prolongement des humanités numériques, IEML a vocation à tenir lieu de système de coordonnées sémantiques introduisant les sciences humaines à un nouveau registre de rigueur et de cohérence. Enfin, le métavers ne serait plus limité à une imitation de l'espace tridimensionnel, mais pourrait rendre sensible le monde des idées sur un mode immersif et social. Il faut penser ensemble toutes ces mutations dans la perspective d'une évolution culturelle à long terme.

8 Sur le plan technique, il s'agit d'un projet agile et décentralisé. Un éditeur *open source* est déjà disponible qui inclut le dictionnaire IEML avec traductions en anglais et en français, un analyseur syntaxique supportant quelques fonctions calculables sur les expressions de la langue et un système de visualisation des paradigmes sous forme de tables. Reste à construire une plateforme d'édition collaborative et de partage des concepts et ontologies. Sur le plan intellectuel et social, en revanche, il s'agit d'un projet ambitieux, car le développement, la maintenance et l'utilisation d'un éventuel protocole sémantique basé sur IEML nécessiteraient un effort coordonné de recherche et de formation à long terme.

9 Je termine cette introduction en répondant par avance à quelques questions d'ordre éthico-politique. IEML est un langage à source ouverte, qui se place délibérément dans la perspective d'une augmentation des communs de la connaissance, et dont le développement fera l'objet d'une gouvernance décentralisée. On pourrait questionner le fait qu'un langage artificiel complexe puisse effectivement servir à démocratiser et rendre plus transparentes l'organisation et l'exploration des données. Mais un métalangage dont la grammaire, entièrement régulière, tient en quinze pages et dont le dictionnaire (et bientôt le trésor des concepts créés par ses utilisateurs) se trouve intégré à l'éditeur n'est peut-être pas si compliqué qu'on pourrait le penser. J'ajoute que la plupart de ses bénéficiaires n'auront pas à l'apprendre puisque les interfaces des applications, y compris l'éditeur lui-même, seront en langues naturelles. La face « code » d'IEML n'est destinée qu'aux ordinateurs. On pourrait également contester l'utilité d'IEML puisque les grands modèles de langue (les *large language models* ou LLM) auraient déjà résolu sur un mode probabiliste les

problèmes liés à la sémantique des langues naturelles. Même si cela était vrai, l'existence des LLM ne remet pas en cause la nécessité, pour un grand nombre d'applications, de catégoriser les données de manière rigoureuse et interopérable, ce qui est la principale finalité d'IEML. D'autre part, il n'est pas nécessaire de penser en termes d'exclusives. Un couplage d'IEML avec les LLM permettrait de perfectionner la formulation et la maîtrise des consignes (*prompts*) et l'explicabilité des résultats, donc la communication entre IA et humains. Il fournirait de plus un langage de communication entre les IA. Les grands modèles de langue pourraient servir d'interfaces utilisateurs à un réseau de bases de connaissances interopérables catégorisées en IEML. J'ajoute, pour terminer sur la fameuse question des biais, qu'une maîtrise transparente des catégories organisatrices des domaines de connaissance par les diverses communautés d'utilisateurs serait préférable au probabilisme statistique à partir des données d'entraînement (forcément entachées de partialité culturelle) ou au réglage massif et souvent opaque par feed-back humain mis en œuvre par les GAFAM (Google, Apple, Facebook, Amazon, Microsoft) et les grandes compagnies d'IA. Cette approche est en phase avec la tendance contemporaine à l'alignement diversifié des grands modèles sur des besoins locaux ou particuliers. L'utilisation d'IEML pourrait commencer par se répandre dans un domaine spécifique, autour d'une ontologie IEML particulièrement utile, puis s'étendre progressivement à d'autres domaines.

¹⁰ Cet article se divise en trois parties. Je présenterai successivement a) les principales données du problème de la mathématisation du langage ; b) la solution que j'ai trouvée à ce problème, IEML ; c) les applications possibles d'IEML, notamment pour l'intelligence collective, l'intelligence artificielle et les humanités numériques.

Le problème : mathématiser la sémantique linguistique

¹¹ Comment créer une langue à la sémantique calculable ? Nous devons d'abord savoir exactement de quoi nous parlons. Nous allons donc commencer par répertorier les grandes *fonctions* des langues afin d'être sûr que notre langue mathématique soit capable de les remplir. Cela fait, nous nous intéresserons à la *structure* des langues et nous identifierons les trois axes – syntagmatique, paradigmatique et sémiotique – qui ordonnent la sémantique linguistique. Je m'attacherai ensuite à bien distinguer la sémantique linguistique des sémantiques référentielle et pragmatique. En effet, la sémantique référentielle est déjà formalisée par la logique mathématique et la sémantique pragmatique relève de la théorie des jeux et de la théorie des systèmes, qui sont en bonne voie de mathématisation. Seul le domaine de la sémantique linguistique – c'est-à-dire celui de la conceptualisation – n'est pas encore modélisé de manière formelle. Nous pourrions alors reconnaître que, dans ce domaine bien délimité, l'axe syntagmatique a déjà été mathématisé sous la forme des langages réguliers. Reste donc à *mathématiser l'axe paradigmatique* au moyen d'un code (IEML) apte à faire correspondre les relations entre les sons (les séquences de phonèmes) aux relations entre les sens (les réseaux de concepts).

Les fonctions des langues

12 Commençons par distinguer les *langues philologiques* des autres types de langues ou de langages. Le concept de « langue philologique » a été introduit par Louis Hejlslev (2000 [1968]) pour désigner les langues capables de « tout dire », qu'elles soient construites ou naturelles. Par exemple, l'espéranto ou IEML sont des langues construites. Néanmoins, parce qu'elles ont les mêmes propriétés expressives qu'une langue naturelle, ce sont des langues philologiques. En revanche, le langage de programmation Python, qui est aussi un langage construit, n'est pas une langue philologique parce qu'il n'est pas capable de tout dire. C'est un langage spécialisé.

La fonction narrative

13 Le premier caractère d'une langue philologique est sa capacité narrative. Pour qu'elle puisse jouer ses rôles pratiques et référentiels, la langue doit être capable d'évoquer des *scènes* complexes dans lesquels plusieurs *actants* interagissent. La citation suivante, extraite de l'œuvre posthume de Lucien Tesnière *Éléments de syntaxe structurale* (1959, chap. 48, 102) explique bien cette propriété des langues philologiques : « Le nœud verbal [...] exprime tout un petit drame. Comme un drame, en effet, il comporte [...] un procès et, le plus souvent, des acteurs et des circonstances. Le verbe exprime le procès. [...] Les actants sont des êtres ou des choses [...] participant au procès. [...] Les circonstants expriment les circonstances de temps, lieux, manière, etc. » En un lieu donné, pendant un certain temps, plusieurs intervenants humains ou non humains, chacun avec ses propres intentions et à sa manière, jouent des rôles distincts déroulant un processus à l'esprit d'un auditeur ou d'un lecteur. La structure de la phrase est une réification grammaticale de cette scénarisation : le verbe détermine l'état ou le processus, le sujet initie l'action, l'objet interagit avec le sujet, etc. Avec la narrativité vient la capacité de description de scènes ou d'interactions réelles, mais aussi la possibilité de la *fiction*.

14 Pour donner libre cours à leur capacité d'évocation de situations complexes, les phrases sont capables de s'emboîter les unes dans les autres comme des poupées russes. Les actants, les circonstances ou les modes d'action peuvent être eux-mêmes décrits par des phrases et cela de manière *réursive*. Marcel Proust a poussé cette propriété des langues à sa limite dans *À la recherche du temps perdu*. De même, les citations et les récits peuvent s'enchâsser indéfiniment, comme dans *Les Mille et Une Nuits* ou dans le *Manuscrit trouvé à Saragosse* de Jan Potocki.

La fonction dialogique

15 Le langage ne doit pas seulement permettre l'évocation des interactions sociales, il doit aussi offrir aux interlocuteurs des moyens de se situer mutuellement dans ces interactions. Toutes les langues possèdent donc des personnes grammaticales, c'est-à-dire des moyens de dire « je » et « nous », « tu » et « vous », « il », « elle », « ils » et « elles ». Une langue vivante est apprise et pratiquée dans le dialogue. Avec le dialogue viennent le questionnement, la réponse, la persuasion, le raisonnement, la démonstration et les rôles variés que peuvent jouer les personnes dans une multitude de jeux sociaux conventionnels (Bakhtine 1978 ; 1981).

16 Le dialogue peut traverser les langues puisque les langues philologiques peuvent *traduire* les autres langues philologiques ainsi qu'une multitude d'autres codes et langages plus spécialisés. Certes, une langue est étroitement adaptée à la vie de ses interlocuteurs, elle reflète leurs univers de référence et leurs jeux sociaux. Elle est pourtant toujours capable d'interpréter les autres langues, avec une inévitable marge de trahison, et de se transformer par des emprunts, des abandons et des créations.

La fonction référentielle

17 Les expressions linguistiques doivent pouvoir pointer vers des références, notamment vers des réalités extralinguistiques, à tous les niveaux de complexité : mot, sous-phrase, phrase complexe, texte. Tel est le rôle des noms propres. Tel est aussi le rôle des démonstratifs et des déictiques (Benveniste 1966 ; 1974, voir notamment les chapitres 20 et 21 du tome 1) : « ceci », « celui-là », « je », « tu », « maintenant », « demain », « ici », « là-bas »...

18 Plus généralement, la langue est un moyen de catégorisation et d'indexation de l'univers de ses interlocuteurs. Toute langue fonctionne comme un système d'adressage sémantique de la mémoire – individuelle et collective – et donc comme un instrument de coordination de l'action.

19 La langue n'est pas seulement un outil pour référencer le monde extralinguistique, elle est aussi capable d'*autoréférence*. C'est notamment le rôle de l'anaphore grammaticale, par exemple lorsqu'un pronom fait référence à un nom cité antérieurement. Cette capacité d'autoréférence de la langue intervient également dans le commentaire et le dialogue. Elle s'élargit jusqu'à l'*auto-explication* : dans un dictionnaire, les mots s'expliquent mutuellement de manière circulaire, si bien qu'une langue philologique est à elle-même son propre métalangage.

La structure des langues

Les composants : phonèmes et mots

20 Nous savons que les langues évoluent, se différencient et s'hybrident mais, pour les besoins de l'analyse, nous allons considérer la *structure* d'une langue à un moment donné. Une langue se construit à partir de quelques unités de *son* (entre une trentaine et une centaine) : les *phonèmes*, qui n'ont pas de signification et sont généralement divisés en *consonnes* et *voyelles*⁴. Les unités de *sens* – appelées *morphèmes* ou *monèmes* – sont formées de séquences de phonèmes. Ce que la plupart des gens appellent des mots sont en fait des agglomérations de quelques unités de sens. Par exemple, le mot « aimeront » contient : le concept d'amour, la forme verbale, le futur, la troisième personne et le pluriel. On considère que les locuteurs d'une langue maîtrisent un vocabulaire compris entre trois mille (compétence minimale) et trente mille (compétence maximale) *lemmes*, ces derniers étant des mots canoniques, indépendants des conjugaisons, des flexions, des cas, etc.

21 Puisque mon but ici n'est autre que de broser à grands traits la structure des langues, et bien que la réalité soit plus complexe, j'appellerai « mots » les unités de sens élémentaires et « phrases » leur composition grammaticale en unités de sens composées.

L'axe syntagmatique

- 22 Selon Roman Jakobson (1963) et l'école structuraliste en linguistique, une langue croise deux axes : syntagmatique et paradigmatique. Commentons par définir un syntagme : c'est une *séquence* de mots – phrase ou membre de phrase – appartenant à une langue dont les règles grammaticales permettent de reconstituer l'arbre de dépendance et le rôle de chaque feuille dans cet arbre. Considérons la phrase : « Je repeins la petite chambre en bleu. » Les règles grammaticales du français nous disent que le sujet (ou premier actant) se place avant, l'objet (ou second actant) après le verbe et le complément de manière souvent après le verbe. Nous en déduisons que le verbe repeindre se trouve à la racine de l'arbre dont nous identifions clairement trois feuilles distinctes : le sujet (« je »), l'objet (la chambre) et la manière (« en bleu »). La chambre est elle-même déterminée (« la ») et qualifiée par un adjectif de taille (« petite »).

Figure 1. L'axe syntagmatique : la chaîne sonore s'organise en un arbre de dépendance.

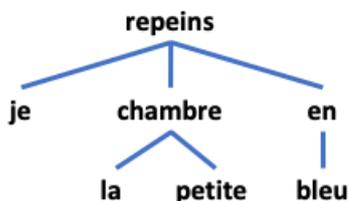


Figure produite par l'auteur

- 23 L'usage de conjonctions, de signes de ponctuation – écrits ou oraux – ainsi que certaines règles grammaticales codent les hiérarchies de dépendance entre phrases et membres de phrases, entre propositions principales et secondaires. Leur place dans la séquence ou l'utilisation de cas permettent de reconnaître le rôle grammatical des mots : verbe ou processus, sujet ou premier actant, objet ou deuxième actant, destinataire ou troisième actant, complément de manière, etc.
- 24 Je ne peux comprendre une phrase que si, à partir d'une séquence de phonèmes, d'une part, et de la connaissance du dictionnaire et du code grammatical propre à ma langue, d'autre part, je reconstitue un schéma d'emboîtement syntactico-sémantique dont la structure topologique est celle d'un arbre. À partir de cette arborescence, dans laquelle le rôle grammatical de chaque mot est déterminé, j'infère la scène ou le modèle mental que mon interlocuteur avait en tête. Inversement, je prononce une phrase sensée lorsque, à partir d'un modèle mental, je construis d'abord un arbre syntaxique de concepts représentés par des mots (une « structure profonde » selon Noam Chomsky) que je transforme ensuite en séquence de sons suivant les règles grammaticales de ma langue. Bien entendu – sauf création littéraire délibérée – ces différentes étapes logiques sont franchies quasi simultanément et de manière inconsciente.

L'axe paradigmatique

25 L'axe paradigmatique est « perpendiculaire » à l'axe syntagmatique. En règle générale, les paradigmes sont des groupes de substitution sur les nœuds d'un arbre syntagmatique. Les exemples de paradigmes les plus connus sont ceux des conjugaisons ou des déclinaisons que nous avons appris à l'école.

26 Considérons de nouveau la phrase « Je repeins la petite chambre en bleu. » En conservant le présent de l'indicatif et l'objet grammatical (deuxième actant), je peux faire varier la personne et le nombre du sujet (ou premier actant) et substituer à « je », la série « tu, il/elle, nous, vous, ils/elles » tout en accordant le verbe. Je peux aussi faire varier les temps en conservant le mode : « repeignis » (passé simple), « repeignais » (imparfait), « repeindrai » (futur). Je peux également, toujours pour la même phrase, conserver le mode, le temps, la personne grammaticale et l'objet, mais remplacer « repeins » par « tapisse », « meuble », « décoore », etc. Dans ce dernier cas, j'ai fait varier le lemme ou, si l'on préfère, le concept de l'action. Je ne peux cependant pas le faire n'importe comment. Comment pourrais-je « endetter » ou « chatouiller » une chambre, par exemple (sinon en un sens hautement métaphorique) ? C'est dire qu'il existe un paradigme des actions – donc des verbes – qui *conviennent* avec le mot « chambre » en rôle grammatical d'objet (ou deuxième actant). De même, je peux certes repeindre un mur, une barrière ou un poteau, mais un peuple, un poème ou un cauchemar seraient nettement plus difficiles à repeindre. Il existe donc un paradigme des entités matérielles en rôle grammatical d'objet qui conviennent avec le verbe « repeindre ».

Figure 2. L'axe paradigmatique : chaque mot d'une phrase est choisi dans une matrice de possibilités alternatives pertinentes dans le contexte.

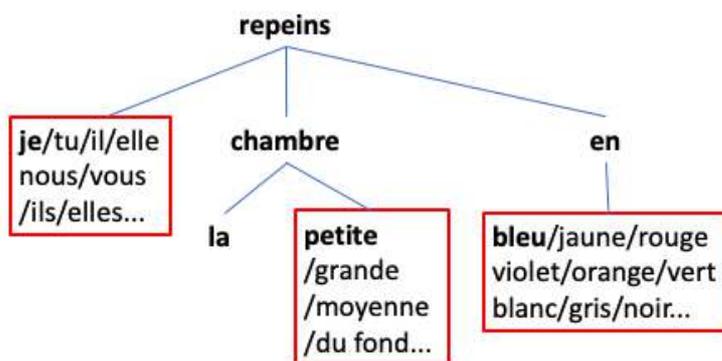


Figure produite par l'auteur

27 Un ensemble de mots qui peuvent se substituer les uns aux autres *dans le même rôle syntagmatique* au sein d'une phrase donnée constitue un paradigme. Les mots du même paradigme *ont en commun* leur convenance sémantique avec les parties constantes de la phrase, mais ils *diffèrent* selon leur système de symétries ou d'oppositions. Un paradigme se présente donc comme un système de variations sur un fond de constantes sémantiques.

28 Beaucoup de mots d'une langue sont associés à d'autres mots qui désignent de manière standard des rôles syntactico-sémantiques complémentaires. Par exemple, le sujet grammatical du verbe *voler* (le délit) est un *voleur* et l'objet du vol est un *larcin*. Lorsque l'on parle d'un *véhicule*, celui qui le fait marcher est un *conducteur* ou un *chauffeur*. Ces systèmes de

convenance entre les mots d'une langue ont été particulièrement étudiés sous les noms de fonctions lexicales (Melchuk 1982 ; Melchuk 1998 ; Polguère 2003), de collocations (Tremblay 2014), de *frames* ou cadres (Fillmore 1968). Les *relations de dépendance entre l'emploi des mots* – relations qui reflètent des connexions sémantiques et pratiques – appartiennent de plein droit aux conventions des langues et devraient figurer systématiquement dans les dictionnaires dignes de ce nom.

29 On peut aussi ranger sous le vaste parapluie des paradigmes les ensembles de mots qui conservent leur concept fondamental – le lemme est identique – tout en changeant de rôle grammatical. Par exemple : « métaphore » (nom qui peut s'utiliser en rôle d'actant), « métaphoriser » (verbe qui s'utilise en rôle d'action), « métaphorique » (adjectif qui s'utilise pour qualifier un nom) et « métaphoriquement » (adverbe qui s'utilise pour qualifier un verbe).

30 En somme, les paradigmes correspondent à ces trois cas :

- groupes de substitution de flexions ou de rôles grammaticaux pour un lemme constant
- groupes de substitution de mots dans une phrase et pour un rôle grammatical constant
- systèmes plus ou moins complexes de cooccurrence de mots (ou de syntagmes) comme dans les fonctions lexicales, les collocations ou les *frames*

31 Retenons que chacun des nœuds d'un arbre syntagmatique appartient à un paradigme, c'est-à-dire à une matrice de possibilités alternatives pertinentes dans un certain contexte. Une matrice paradigmatique décrit un domaine de variation sémantique structuré par un système de symétries ou d'oppositions. J'ajoute que les variables paradigmatiques sont généralement codépendantes : le choix d'un mot dans un rôle syntagmatique conditionne le choix des mots dans les autres rôles. *Dans tous les cas, un paradigme se présente comme un champ de variations sémantiques qui conviennent à un contexte déterminé.*

L'axe sémiotique

32 Certains linguistes ajoutent aux deux axes syntagmatique et paradigmatique un troisième axe, sémiotique, qui concerne le rapport entre signifiant et signifié ou, si l'on préfère, la relation entre le son (le signifiant) et le sens (le signifié) des expressions linguistiques.

33 La signification implique au moins trois actants qui jouent des rôles distincts. Un *signe* (a) indique quelque chose, le *référént* (b), et l'indique pour quelqu'un, l'*interprétant* (c). Le symbole linguistique est un type particulier de signe, divisé en deux parties : le signifiant et le signifié. Nous disons qu'un symbole linguistique est conventionnel parce que, dans la plupart des cas, il n'existe pas d'analogie ou de lien de causalité entre le son et le concept : par exemple, entre le son « arbre » et un genre de plantes. Le *signifié* du son « arbre » est régi par la langue française et n'est pas laissé au choix de l'interprétant. Cependant, c'est dans le contexte d'une énonciation particulière que l'interlocuteur comprend le *référént* du mot : est-ce un arbre syntaxique, un palmier, un arbre de Noël ou autre chose encore ? Souvenons-nous de cette distinction importante : le signifié est déterminé par la langue, mais le référént identifié par l'interprétant dépend du contexte. Nous faisons parfois l'erreur d'attribuer la connexion entre le signifiant et le signifié à l'interprétant parce que les langues natu-

relles sont ambiguës et que l'interprétant utilise souvent sa compréhension du contexte pour déterminer le signifié d'un signifiant équivoque... Mais l'intermédiaire fondamental entre le signifiant et le signifié n'est autre que le système linguistique lui-même.

Figure 3. L'axe sémiotique : une chaîne signifiante correspond à un réseau de concepts, un arbre syntagmatique que croisent des matrices paradigmatiques.

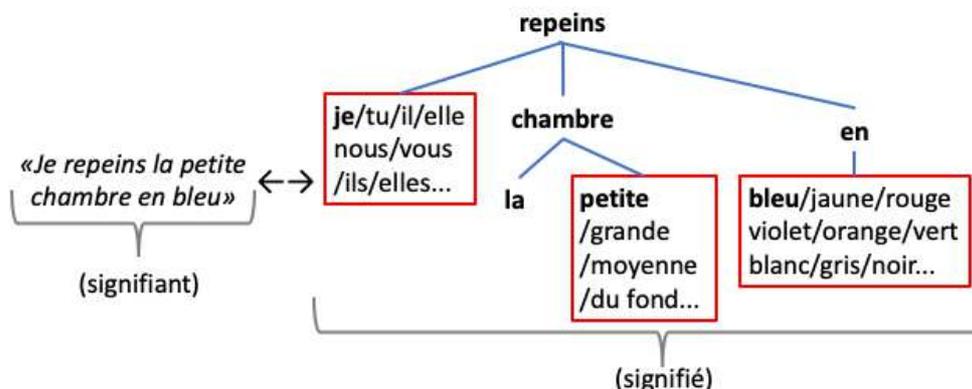


Figure produite par l'auteur

34

On a remarqué depuis longtemps déjà qu'à part quelques rares cas de mots dont le son évoque le sens (le nom « coucou » rappelle le chant de l'oiseau), la phonétique d'un mot n'a pas de rapport direct avec sa sémantique. La relation entre le son et le sens des mots est purement conventionnelle et dépend des langues. Il ne faut cependant pas exagérer l'indépendance du son et du sens dans la parole. Plus la langue est poétique et plus les aspects prosodiques comme les rythmes et les assonances – en particulier les rimes – contribuent à la création du sens. De plus, le déchiffrement de la dimension syntagmatique – c'est-à-dire la reconstitution d'un arbre de dépendance comprenant l'identification du rôle grammatical des feuilles – s'appuie largement sur l'ordre des mots, l'usage de cas sensibles à l'oreille, la ponctuation et l'accent tonique. Il ne s'agit pas ici d'une correspondance terme à terme directe entre unités signifiantes et unités signifiées mais plutôt d'une analogie ou d'une résonance entre structure sonore et structure grammaticale. Enfin, la phonétique joue son rôle dans la reconnaissance de la dimension paradigmatique, par exemple quand elle permet de repérer une partie constante et une partie variable dans un mot. Notons par exemple l'alternance « petit/petite » où la constante sémantique et la variable de genre sont facilement reconnaissables. De même pour l'alternance « tarte/tartelette » : la constante sémantique comme la variation de taille sont audibles. Mais pour désigner une minimontagne, on ne dit pas une « montagnette » mais une colline. Montagne et colline font pourtant partie du même paradigme des accidents topologiques et des descriptions géographiques, elles ont en gros la même forme (pied, pente, sommet) tout en s'opposant sur une échelle de variation petit/grand. Rien de ce qui précède n'est cependant signalé au niveau phonétique. Dans cette veine des paradigmes impossibles à reconnaître au son, on peut évoquer les couleurs : vert, rouge, jaune et bleu s'opposent dans le paradigme des couleurs, mais leur appartenance au même paradigme n'est pas sensible au plan phonétique. Si la phonétique aide souvent à coder et à décoder l'axe syntagmatique dans les langues naturelles, elle ne joue ce rôle que rarement pour l'axe paradigmatique. Retenons ce dernier point, qui va se révéler capital pour la solution du

problème du calcul du sens car, en IEML – et contrairement à la situation qui prévaut dans les langues naturelles –, tous les mots d'un même paradigme ont un invariant phonétique en commun (voir le paradigme des couleurs en IEML à la figure 7).

Sémantique de la parole

35 Abordons maintenant la question de la sémantique en général, de manière plus approfondie que dans l'introduction.

La sémantique pragmatique ou l'interaction

36 Le langage fonctionne d'abord comme un outil d'exercice, de régulation et de représentation des interactions sociales. De fait, même la pensée personnelle – qui ne dépasse pas les limites de notre for intérieur – prend la forme d'une conversation. Selon Lev Vygotsky (1962 [1934]), la pensée résulte même d'une internalisation du dialogue. Lorsque nous nous interrogeons sur le sens d'une phrase, il faut donc d'abord relever ses dimensions pragmatique et dialogique : le type de jeu de langage (Wittgenstein 2014 [1953]) joué par les interlocuteurs, les circonstances de son énonciation, ses effets potentiels. En somme, l'énonciation d'une phrase est un acte (Austin 1962 ; Searle 1969). Et le sens pragmatique de la phrase est l'effet de cet acte tel qu'il s'enregistre dans la mémoire collective des interlocuteurs. La sémantique pragmatique se noue dans le rapport entre un acte de langage – plus généralement un acte symbolique – et une situation sociale. Elle relève principalement de la théorie des jeux, du droit, de la sociologie, voire de la théorie des systèmes.

La sémantique référentielle ou la vérité

37 Au niveau de la sémantique pragmatique, une phrase est plus ou moins pertinente ou efficace, heureuse ou malheureuse, selon ses effets en contexte. Mais pour qu'une énonciation soit capable d'un tel sens pragmatique, il faut le plus souvent qu'elle soit *aussi* capable de décrire une réalité, qu'elle soit exacte ou inexacte, sérieuse ou fictive. Comme les états de conscience, les propositions sont intentionnelles (Husserl 1950 ; Searle 1983), c'est-à-dire qu'elles pointent vers une référence. La sémantique pragmatique enregistre la manière dont une énonciation modifie une situation sociale tandis que la *sémantique référentielle* se focalise sur le rapport entre un énoncé et un état de choses. Une proposition n'est vraie ou fausse qu'au niveau de la sémantique référentielle. C'est aussi à ce niveau que sa valeur de vérité se transmet – ou non – d'une proposition à l'autre dans un *raisonnement logique*. Notons que, si le langage n'était pas susceptible de porter la sémantique référentielle – c'est-à-dire de représenter la réalité et de dire le vrai –, il ne pourrait pas jouer son rôle pragmatique. Il serait par exemple impossible d'énoncer les règles d'un jeu ou d'évoquer des interactions sociales. Relevant plutôt des sciences exactes et de la logique, la sémantique référentielle se détache sur l'arrière-plan d'une réalité objective. Ici, les expressions linguistiques décrivent et indexent le monde des interlocuteurs et permettent le raisonnement logique.

La sémantique linguistique ou la conceptualisation

38 De même que la sémantique pragmatique a pour condition de possibilité la sémantique référentielle, la sémantique référentielle à son tour ne peut se manifester que sur la base de la sémantique linguistique. En effet, il faut d'abord qu'une phrase ait un sens dans une certaine langue et qu'elle évoque une certaine représentation mentale – une *scène* concrète ou métaphorique – pour que l'on puisse ensuite comparer cette conceptualisation avec la réalité et la déclarer vraie, fausse ou à moitié vraie. Le sens linguistique d'une expression vient des mots dont elle est composée et du sens qui leur est assigné dans un dictionnaire. Il découle également des rôles grammaticaux que jouent ces mots dans la phrase. En somme, le sens linguistique émerge des relations d'interdéfinition, de convenance, de ressemblance et de différence entre les mots dans le dictionnaire et des relations grammaticales entre les mots dans la phrase. Un complément circonstanciel de plus ou de moins, le remplacement d'un verbe par un autre, un singulier au lieu d'un pluriel auraient modifié le sens de la phrase et produit un autre récit. Le sens linguistique est différentiel. À cet étage de la sémantique, le sens d'une phrase se détermine dans son rapport à la langue, voire dans le rapport avec sa traduction dans d'autres langues. Antérieur à la question de la vérité, il a plus d'affinités avec la littérature qu'avec la logique.

La sémantique phénoménologique ou l'ancrage sensoriel-émotionnel

39 Une parole qu'on nous adresse, pour peu que nous maîtrisions la langue, fait résonner – en interaction avec les trois niveaux qui précèdent – un complexe de souvenirs, d'impulsions motrices subliminales, d'images sensorielles et d'émotions qui est évidemment variable selon les personnes. Ce niveau de la sémantique correspond à la notion d'ancrage corporel et subjectif du sens et ne semble pas formalisable.

Calculer la sémantique linguistique

40 Nous venons d'évoquer la structure feuilletée du sens. Une énonciation manifeste d'un seul geste les quatre sémantiques : pragmatique, référentielle, linguistique et phénoménologique. Ce n'est que pour l'analyse qu'elles se distinguent. La sémantique référentielle et la sémantique pragmatique sont déjà formalisées et la sémantique phénoménologique ne semble pas pouvoir l'être. Au cours de cette analyse, nous avons distingué soigneusement la sémantique linguistique, dont la formalisation, pour tant possible, reste incomplète.

41 Par « calcul de la sémantique linguistique », j'entends essentiellement deux choses étroitement liées : (1) un codage-décodage automatique entre chaînes phonétiques et réseaux de concepts et (2) un ensemble cohérent de mécanismes fonctionnels pour la génération, la transformation et la reconnaissance de réseaux de concepts. Comme nous allons le voir, le problème du calcul de la sémantique a été résolu par l'invention du code linguistique, IEML, qui le rend possible. Or le calcul sémantique doit en principe prendre une forme algébrique. Qu'est-ce à dire ?

Le calcul algébrique

42 L'algèbre peut être définie comme l'art de mécaniser les opérations symboliques. Elle suppose l'adoption d'un *système de codage* pour les variables et les opérations ainsi que l'utilisation d'*enchaînements réglés d'opérations* – les algorithmes – pour exécuter des calculs. L'application d'un algorithme à un ensemble de variables en entrée mène à la variable de résultat en sortie.

43 Les symboles étant constitués d'une partie signifiante et d'une partie signifiée, il faut souligner que les opérations algébriques s'appliquent aux signifiants de manière mécanique, c'est-à-dire sans tenir compte des signifiés. Les algorithmes sont aveugles aux contenus sémantiques des symboles qu'ils manipulent. Même lorsque nous effectuons une multiplication à la main, nous suivons toujours la même routine, quels que soient les nombres multipliés. Les signifiants manipulés par les opérations algébriques peuvent être assimilés à des pièces matérielles comme des jetons, des billes ou des cailloux. Le mot « calcul » lui-même vient du latin *calculus* signifiant « galet » (*calculus* a aussi donné « caillou » en français) parce que les anciens Romains utilisaient des petits galets pour effectuer des opérations arithmétiques sur des abaques.

44 Le progrès du codage algébrique et le perfectionnement des procédures automatiques de calcul marquent généralement un saut de scientificité du domaine auquel ils s'appliquent.

La modélisation algébrique du langage

45 Qu'en est-il de la modélisation algébrique du langage ? Pour ce qui est de sa dimension pragmatique, la théorie mathématique des jeux et la théorie des systèmes, qui ont pris leur essor à partir de la Seconde Guerre mondiale, sont déjà bien avancées. Nous avons vu plus haut que le *raisonnement logique* relevait de la sémantique référentielle, puisque son objet ultime est la vérité ou la fausseté des propositions. Le codage algébrique de la logique est quasiment achevé et supporte évidemment les progrès de l'informatique.

46 L'enjeu principal de la logique est la vérité des propositions – donc leur correspondance avec des états de choses – ainsi que la validité des raisonnements à partir de propositions vraies. En revanche, la *théorie des langages réguliers*, telle qu'elle a été développée par les pionniers Stephen Cole Kleene (1956), Noam Chomsky et Marcel-Paul Schützenberger (1963), concerne uniquement la structure interne des langues, autrement dit leur *syntaxe*. Un langage régulier est constitué d'un ensemble fini de symboles (son alphabet) et de règles de formation des mots, qui sont des séquences de symboles. Il faut noter que les langages réguliers peuvent se composer en hiérarchies de complexité interne et que leurs « mots » correspondent souvent à ce que l'on aurait envie de nommer des phrases ou des textes. Moyennant une certaine idéalisation, les structures syntagmatiques des langues naturelles peuvent être décrites au moyen de langages réguliers. Le point essentiel du traitement algébrique de la syntaxe est que les expressions bien formées d'un langage régulier peuvent être produites ou reconnues *par des machines* (des automates finis). La théorie des langages réguliers s'est développée en même temps que l'informatique et elle a accompagné la conception des langages de programmation et des compilateurs.

47 La théorie des jeux, le calcul logique et la syntaxe formelle ne suffisent cependant pas à une modélisation algébrique de la sémantique. Le premier concerne la sémantique pragmatique, le second, la sémantique référentielle et le troisième ne formalise que l'axe syntagmatique de la sémantique linguistique. Il reste à coder l'axe paradigmatique.

48 Les philosophes ont remarqué depuis longtemps que les langues naturelles – parce qu'elles sont irrégulières et ambiguës – sont impropres au calcul mécanique sur les concepts. On a essayé (de Ramon Llull à G.W. Leibniz⁵) des appariements entre concepts et nombres afin de soumettre les concepts à des calculs de type arithmétique. Mais si tous les nombres sont des concepts, tous les concepts ne sont pas des nombres. Les codes et les opérations conçus pour faciliter le calcul des uns se sont avérés inadéquats au calcul des autres. L'algèbre purement combinatoire a également prouvé son insuffisance. Il est impossible de reconstruire tous les concepts imaginables et leurs relations à partir d'un petit nombre de concepts élémentaires en n'utilisant *que* des méthodes combinatoires. Pour mener à bien le programme du calcul sémantique, il faut sans doute utiliser une algèbre, au sens le plus abstrait du terme, mais une algèbre spécialement conçue pour coder la matière linguistique des concepts. Puisque les langues ont une dimension syntagmatique, nous devons nécessairement utiliser un langage régulier. Seulement, ce doit être un langage régulier *capable de coder les paradigmes* : les groupes de substitution sur les nœuds des arbres syntagmatiques, les symétries internes à ces groupes de substitution, le jeu des constantes et des variables dans la construction des concepts et de leurs relations. C'est pour répondre à ce défi qu'IEML a été inventé.

Le codage du signifié par le signifiant en IEML

49 Rappelons que les ordinateurs sont des robots qui exécutent des algorithmes. Ils ne font que reconnaître des séquences de 0 et de 1 dans des cases-mémoire, exécuter des opérations arithmétiques et logiques sur ces séquences et ranger les résultats dans d'autres cases-mémoire. Du début à la fin, ils ne manipulent que du signifiant et c'est pourquoi ils sont bien adaptés au traitement des langages réguliers. Les langages réguliers effectivement utilisés aujourd'hui sont des langages de programmation. Mais la « sémantique » des langages de programmation n'est autre que l'exécution des opérations qu'ils commandent. Aucun d'eux n'approche la capacité expressive d'une langue philologique, qui permet de parler de tout et de rien et d'accomplir des actes illocutoires qui vont au-delà des instructions données à une machine. Pour dépasser cette limite, je me suis proposé de construire une langue (1) philologique, (2) régulière et (3) dont les structures paradigmatiques seraient automatiquement décodables ; en somme, une langue à la sémantique calculable.

50 Aussi bien les signifiants que les signifiés d'une langue sont organisés par des systèmes de différences, de symétries et d'oppositions. Puisque les langues sont conventionnelles, rien n'interdit de construire une langue philologique dont les structures syntagmatiques et paradigmatiques soient codées par un arrangement régulier des signifiants. Le système de différences des *signifiants* d'une langue à la sémantique calculable doit refléter la structure de son système de différences des *signifiés*. En suivant cette méthode, on n'obtient pas une phonétique des mots (ou des phrases)

qui ressemble à leur sens – comme dans le *Cratyle* de Platon ou d'autres utopies linguistiques – mais plutôt des *relations* entre les sons qui entretiennent une analogie avec les *relations* entre les sens.

Une solution : IEML

51 IEML permet de calculer la sémantique parce que les symétries sémantiques entre ses signifiés sont codées par des symétries syntaxiques entre ses signifiants. Les rôles *variables* des fonctions morphologiques (pour générer les mots) et syntagmatiques (pour générer les phrases) créent des matrices paradigmatiques régulières. Les lettres de l'alphabet IEML correspondent à des *places dans des systèmes de symétries*, tout comme les cellules des matrices paradigmatiques de ses mots et de ses phrases. Le langage dans son ensemble répond à une structure algébrique régulière (un anneau non commutatif) qui autorise l'automatisation d'un grand nombre de fonctions.

52 Ma description d'IEML va s'étagier en quatre parties. Je vais d'abord présenter l'alphabet du langage et la *fonction morphologique* uniforme qui utilise cet alphabet pour construire les mots du dictionnaire. Dans un second temps, j'expliquerai la *fonction syntagmatique* qui prend en entrée les mots du dictionnaire pour construire des phrases. J'exposerai dans une troisième partie le système de relations syntaxiques – représenté pour la machine sous forme de graphe – qui découle des fonctions morphologiques et syntagmatiques. En plus de ces deux dernières, des *fonctions de création de liens* sont utilisées pour tisser des *réseaux de relations sémantiques entre concepts* (c'est-à-dire entre mots ou entre phrases), ce qui correspond au niveau du texte en linguistique ou aux ontologies et graphes de connaissances en informatique. Finalement, une quatrième partie abordera le traitement de la référence (et de l'autoréférence du langage) en IEML.

Les mots

Le dictionnaire

53 Afin de rendre la langue optimalement calculable et facile à appréhender, le dictionnaire d'IEML ne dépasse pas 3 000 mots. Ces mots ont été soigneusement choisis afin de faciliter au maximum la création récursive de nouveaux concepts au moyen de phrases. J'ai tenu compte du fait que les sciences exactes avaient déjà standardisé leurs codes et langues spéciales et que ces codes (par exemple les éléments chimiques) peuvent être notés en référence. C'est pourquoi, si le vocabulaire de base d'IEML est propre à décrire n'importe quelle signification, il a été plus particulièrement conçu pour outiller les domaines de la littérature, des sciences humaines et des sciences sociales. Une grande variété de points de vue métaphysiques, épistémologiques, anthropologiques et existentiels ont été pris en considération. En particulier, les mots IEML peuvent exprimer aussi bien les philosophies « orientales » que les philosophies « occidentales ».

54 Afin que la langue soit autant que possible *univoque*, il n'y a pas deux mots dans le dictionnaire qui aient le même sens (absence de synonymes) ou la même forme (absence d'homonymes). Les mots IEML sont aujourd'

hui traduits en français et en anglais, mais ils peuvent être traduits dans n'importe quelle langue. Notons que les mots représentent des *concepts* qui pourront se présenter selon leur flexion et leur rôle dans les phrases comme des verbes, des noms, des adjectifs ou des adverbes. Le dictionnaire d'IEML comporte deux types de mots : les mots pleins, ou lemmes, qui expriment des concepts ou des catégories générales, et les mots-outils.

55 Les mots pleins, réunis en paradigmes, couvrent un grand nombre de domaines : climats et paysages, continents et régions du globe, mers et océans, ciel et météorologie, animaux et plantes, fonctions techniques, fonctions anthropologiques, types de relations, unités de temps, valeurs et idées, signes et fonctions sémiotiques, curation de données et pensée critique, sentiments, parties du corps, etc. Je recommande tout particulièrement aux lecteurs de cette revue l'examen des paradigmes liés aux sciences humaines et à leurs objets ⁶.

56 Les mots-outils, également réunis en paradigmes, servent à :

- *fléchir* les lemmes, comme les genres et nombres pour les noms ou les modes et les temps pour les verbes (il existe quatre-vingts flexions distinctes en IEML)
- préciser au moyen de *prépositions* les rôles grammaticaux des phrases, comme « au-dessus », « en dessous », « à gauche », « à droite », pour les compléments de lieu (il existe cent trente et une prépositions en IEML, chacune étant rattachée à un rôle grammatical particulier)
- établir des *jonctions* entre concepts, membres de phrases ou phrases tels que « et », « ou », « mais », « parce que » ou « donc » (il existe vingt-neuf jonctions en IEML)

57 Sa vaste palette de mots-outils permet à IEML d'exprimer des nuances fines et de réduire le nombre de lemmes nécessaires pour définir des concepts complexes.

L'alphabet

58 IEML étant un langage régulier, il repose sur un ensemble de six symboles primitifs ou alphabet : E U A S B T. Toutes les expressions d'IEML se composent de séquences ponctuées de ces six caractères. Chacune des lettres est un mot en soi et représente un signifié particulier lorsqu'elle est utilisée seule. Mais lorsqu'une lettre est utilisée dans la composition d'un autre mot, elle sert à marquer un invariant paradigmatique ou bien une place dans un groupe de symétries.

Figure 4. Arbre des six symboles primitifs (lettres capitales)

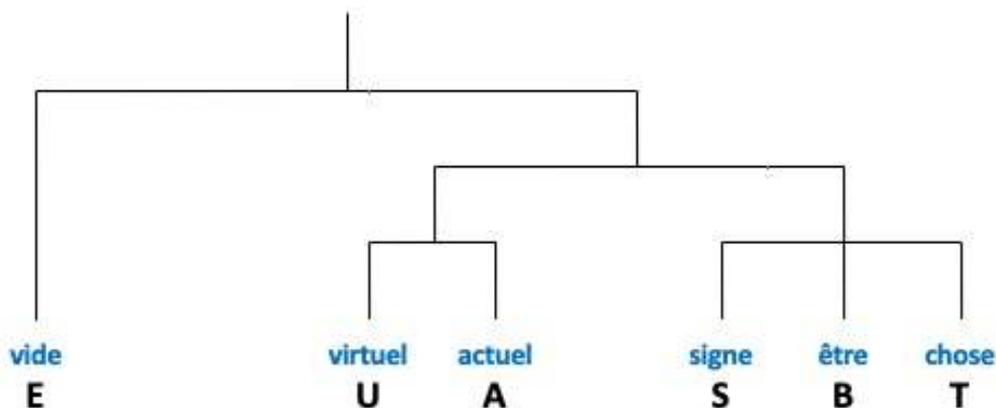


Figure produite par l'auteur

59 La figure 4 ci-dessus ne représente pas une phrase, mais le système de symétries qui organise les variables primitives.

60 E exprime le *vide*, l'absence, le zéro, le silence, etc. Utilisé avec d'autres lettres, il sert à exprimer la « symétrie un ».

61 U, le *virtuel*, dénote le potentiel, l'abstrait, l'immatériel (Lévy 1995).

62 A, l'*actuel*, représente l'effectivité, ce qui est concret, tangible, matériel.

63 Utilisés dans un mot composé, U et A servent à exprimer les deux pôles opposés d'une symétrie binaire quelconque.

64 S, le *signe*, désigne une entité ou un événement qui *signifie* quelque chose pour quelqu'un, tel qu'un symbole ou un document.

65 B, l'*être*, représente l'interprétant pour qui le signe signifie, une entité ou un processus pourvu d'autoréférence et de capacité interprétative.

66 T, la *chose*, dénote un objet ou un *réfèrent*, la réalité dont on parle.

67 Utilisés dans un mot composé, S, B et T expriment les trois places distinctes d'une symétrie ternaire quelconque⁷.

La fonction morphologique

68 Les mots sont générés récursivement par une *opération multiplicative* non commutative à trois rôles qui concatène les mots de la couche inférieure. On ne peut multiplier que des mots de même couche.

69 À la couche *un*, on aura des mots de la forme « U:S:E.: U:B:E.: U:T:E.: A:S:E.: A:B:E.: S:T:E.: », etc. Les mots sont générés par des opérations de la forme « U: ⊗ S: ⊗ E: = U:S:E.: ».

70 À la couche *deux*, on aura des mots de la forme « U:S:E.:U:B:E.:U:T:E.: - A:S:E.:A:B:E.:S:T:E.: - », etc. Les mots sont générés par des opérations de la forme « U:S:E.: ⊗ U:B:E.: ⊗ U:T:E.: = U:S:E.:U:B:E.:U:T:E.: - ».

71 On s'arrête à la cinquième couche, ce qui est amplement suffisant pour générer un dictionnaire de 3 000 mots. Des signes de ponctuation (: . - ' ,) sont utilisés pour marquer les couches successives, de la couche 0 (:) à la couche 4 (,). En hommage à Spinoza, les trois rôles de l'opération morphologique sont appelés substance (la première place), attribut (la deuxième place) et mode (la troisième place).

72 Il existe une opération additive, qui est une simple union des séquences de même couche. L'opération additive est commutative, mais l'ordre des lettres est normalisé afin d'assurer qu'il n'y ait pas deux expressions formellement différentes qui aient le même sens.

73 La multiplication est distributive sur l'addition. Par exemple : « U: ⊗ (S: ⊕ B: ⊕ T:) ⊗ E: = U:S:E.: ⊕ U:B:E.: ⊕ U:T:E.: ».

74 Sur un plan mathématique, l'algèbre qui résulte de l'alphabet de six symboles, de la multiplication ternaire non commutative et de l'addition possède une *structure d'anneau non commutatif*⁸.

Les lettres minuscules

75 Afin de rendre les mots plus lisibles, les vides sont élidés, notamment en fin de mot, et certaines combinaisons de trois primitives sont représentées par des lettres minuscules (voir la figure 5 ci-dessous). Dans le texte explicatif qui suit, on ne notera pas le E: en rôle de mode afin de simplifier la lecture.

Figure 5. Table des 25 lettres minuscules

Les 25 lettres minuscules en IEML					
Attribut →	U	A	S	B	T
Substance ↓					
U	UUE wo. orientation	UAE wa. action	USE y. savoir	UBE o. vouloir	UTE e. pouvoir
A	AUE wu. perception	AAE we. manifestation	ASE u. énoncer	ABE a. s'engager	ATE i. faire
S	SUE j. idée	SAE g. message	SSE s. réflexion	SBE b. langage	STE t. mémoire
B	BUE h. sujet	BAE c. individu	BSE k. société	BBE m. affect	BTE n. monde
T	TUE p. objet	TAE x. corps	TSE d. vérité	TBE f. vie	TTE l. espace

Figure produite par l'auteur

76 **La tétrade wo. wa. wu. we.** est générée par la fonction « (U: ⊕ A:) ⊗ (U: ⊕ A:) ». Pour en comprendre le sens, il faut se rapporter au *cycle sensori-moteur* fondamental. Commençons par l'acte virtuel de l'orientation, qui se déroule dans le for intérieur (wo.). À partir de ce lieu virtuel, nous *agissons* (wa.). En conséquence, notre environnement actuel se transforme et une nouvelle réalité *se manifeste* (we.). Enfin, à partir de l'actualité de l'événement externe, nous *percevons* (wu.) par un retour à l'intériorité. Notez les deux oppositions entremêlées :

- wa. *agir* (dedans → dehors) / wu. *percevoir* (dehors → dedans)
- wo. *s'orienter* (dedans → dedans) / we. *apparaître* (dehors → dehors)

77 **L'hexade y. o. e. u. a. i.** est générée par la fonction « (U: ⊕ A:) ⊗ (S: ⊕ B: ⊕ T:) ». Les mots correspondants représentent six sortes d'actions. La rangée générée par « U: ⊗ (S: ⊕ B: ⊕ T:) » affiche trois actions virtuelles (ou latentes) : *savoir* (y. liée au signe), *vouloir* (o. liée à l'être) et *pouvoir* (e. liée à la chose). La rangée générée par « A: ⊗ (S: ⊕ B: ⊕ T:) » exprime trois actions actuelles (ou manifestes) : *communiquer* ou dire (u.), utiliser des signes pour s'exprimer, *s'engager* (a.) auprès de quelqu'un et/ou à quelque chose et *faire* (i.) quelque chose dans le monde matériel.

78 **L'hexade j. g. h. c. p. x.** est générée par la fonction « (S: ⊕ B: ⊕ T:) ⊗ (U: ⊕ A:) ». Les mots correspondants représentent six sortes d'acteurs. La rangée générée par « S: ⊗ (U: ⊕ A:) » montre deux types d'acteurs liés au signe : les *idées* (j.) dans le monde virtuel et les *messages* (g.) dans le monde actuel. La rangée générée par « B: ⊗ (U: ⊕ A:) » affiche deux types d'acteurs liés à l'être : les *sujets* (h.), ou les formes abstraites de subjectivité, du côté virtuel et les *individus* (c.) qui existent effectivement du côté actuel. La

rangée générée par « T: ⊗ (U: ⊕ A:) » affiche deux types d'acteurs liés à la chose : les *objets* (p.), du côté virtuel, et les *corps* physiques (x.), du côté actuel.

79 **L'ennéade s. b. t. k. m. n. d. f. l.** est générée par la fonction « (S: ⊕ B: ⊕ T:) ⊗ (S: ⊕ B: ⊕ T:) ». Les mots correspondants représentent neuf dimensions de l'expérience humaine.

80 *Réflexion* (« s. = S: ⊗ S: »). Les signes et les symboles alimentent le pouvoir de l'esprit de réfléchir, comprendre, former des jugements et générer d'autres symboles. Déduction, induction, interprétation et imagination s'obtiennent en combinant les signes. Les traditions idéalistes, la phénoménologie qui étudie les structures de la conscience, ainsi qu'une bonne partie des philosophies méditatives de l'Inde, sont fondées principalement sur l'acte de réflexion.

81 *Langage* (« b. = S: ⊗ B: »). Les signes mènent à la communication. Un message est envoyé à un sujet qui lui donne un sens. La philosophie du langage et de la communication se concentre sur la parole, ses pouvoirs, ses limites et les médias dans lesquels elle baigne.

82 *Mémoire* (« t. = S: ⊗ T: »). Par le geste de l'inscription, un signe est réifié en une chose. Sans aucun moyen d'enregistrement, il n'y a pas de mémoire et donc pas de notion de temps, d'histoire ou d'évolution. De même, les grandes traditions religieuses fondées sur des textes sacrés relèvent partiellement de cet archétype conceptuel.

83 *Société* (« k. = S: ⊗ B: »). Une communauté s'autoorganise au moyen de systèmes symboliques. Les membres d'un clan peuvent avoir un animal totemique. Nous combattons sous le même drapeau. Nous échangeons des biens économiques en nous mettant d'accord sur leur valeur. Nous obéissons à la loi. En bons citoyens, nous nous engageons à respecter nos promesses et nos contrats. Nous sommes liés par des conventions. La sociologie a mis en évidence ces fonctions symboliques essentielles qui construisent la société.

84 *Émotion* (« m. = B: ⊗ B: »). Les êtres s'orientent vers les autres êtres. Le désir, l'amour, la haine, la peur, l'indifférence ou la compassion sont des émotions qui circulent entre nous. Les gens pieux, les poètes, les artistes et les comédiens connaissent ce rôle central des émotions dans la vie humaine. La psychanalyse, une grande partie de la psychologie clinique et le bon sens insistent sur l'importance des liens affectifs.

85 *Monde* (« n. = B: ⊗ T: »). Les humains s'expriment dans leur environnement physique. Ils l'habitent, ils y travaillent à l'aide d'outils, ils nomment les choses et leur donnent de la valeur, créant un monde ordonné, un cosmos. La philosophie, l'anthropologie, le droit et d'autres disciplines des sciences sociales et humaines résonnent avec ce concept qui implique simultanément la réalité de l'univers non humain (la chose) et le modelage culturel de son ordre (l'être).

86 *Vérité* (« d. = S: ⊗ T: »). Un message descriptif contient une représentation de la réalité. Une proposition est vraie si c'est une description correcte d'un état de choses. Les traditions de philosophie logique et analytique s'intéressent avant tout à la notion de vérité et aux problèmes de la référence. De même, l'épistémologie et les sciences cognitives placent la construction de la croyance véritable, ou connaissance, au cœur de leur approche.

87 *Vie* (« f. = B: ⊗ T: »). La matière s'anime. Les créatures vivantes sont incarnées. Manger et boire sont parmi les moyens les plus évidents par lesquels une chose est transformée en être. Plusieurs écoles philosophiques mettent l'accent sur l'expérience vécue du sujet. Les traditions empiriques, biologiques et médicales trouvent leurs concepts dans le corps vivant et ses fonctions.

88 *Espace* (« l. = T: ⊗ T: »). Un point se connecte à d'autres points. Les espaces géométriques et physiques sont construits sur les relations et les proximités. Géomètres, topologues, atomistes, matérialistes et physiciens fondent leurs concepts sur l'espace.

89 De même que les autres lettres, s. b. t. k. m. n. d. f. l. peuvent être utilisées dans des mots composés pour représenter des positions dans une *en-néade*, à savoir une dialectique quelconque de neuf concepts.

90 Retenons que, comme les symboles primitifs, les lettres minuscules ont un sens particulier lorsqu'elles sont utilisées seules, mais qu'elles marquent un invariant paradigmatique ou une place dans un groupe de symétries lorsqu'elles servent à composer un mot. En combinant la multiplication et l'addition, il est possible de créer autant de fonctions morphologiques (et de systèmes de symétries) que l'on voudra.

91 On aurait pu choisir d'autres primitives et arriver à d'autres concepts comme résultats des premières multiplications, comme d'ailleurs des suivantes. Mais il s'agit moins ici de démontrer des vérités indiscutables que d'arranger un système de notions primitives pointant vers toutes les directions de l'espace sémantique et capable de générer un grand nombre de notions dérivées aux couches suivantes. Non seulement les langues sont conventionnelles, mais *elles ne peuvent pas ne pas l'être*.

92 Je rappelle que les primitives d'IEML (les lettres majuscules et minuscules) n'ont pas vocation à construire tous les sens possibles au moyen d'une combinatoire de leurs significations particulières (ce qui est impossible), mais à fournir les symétries sémantiques élémentaires à partir desquelles les règles de grammaire peuvent représenter n'importe quelle symétrie sémantique complexe dans les paradigmes que l'on voudrait construire. Ces primitives n'ont pas non plus la prétention d'être « naturelles ». S'il existe des primitives sémantiques réellement universelles, elles ont été découvertes par Anna Wierzbicka (1996) à la suite d'une étude de nombreuses langues naturelles. Notons d'ailleurs que la centaine de concepts élémentaires du *natural semantic metalanguage* de Wierzbicka figure dans le dictionnaire de base d'IEML.

Les paradigmes de mots

93

Chaque mot du dictionnaire IEML fait partie d'un paradigme et d'un seul. Un paradigme est généré par une fonction morphologique dont certains rôles fonctionnels sont occupés par une constante tandis que d'autres rôles sont occupés par des variables, une addition fonctionnant comme groupe de substitution. Il n'y a pas plus de trois rôles variables dans une fonction morphologique, qui se présente donc comme une matrice à une, deux ou trois dimensions. Un paradigme est visualisé par une ou plusieurs *tables à deux dimensions*. Les mots de la même table ont une partie constante (ils ont les mêmes lettres dans les mêmes rôles fonctionnels) et se distinguent par leurs parties variables. Les mots appartenant à la même rangée ou à la même colonne ont également une partie constante et une partie variable. Ces jeux de constantes et de variables signifiantes dans des rôles fonctionnels codent les rapports de symétrie entre les signifiés des mots. On trouvera ci-dessous deux exemples de paradigmes de mots.

Figure 6. Table présentant quelques composantes du développement humain

connaissance organisée, discipline scientifique s.y.-	pensée s.o.-	grammaire, maîtrise des codes s.e.-	icône, signe analogue à ce qu'il représente s.u.-	interprète - entre les langues s.a.-	outil d'interface, outil de transcodage s.i.-
intelligence conceptuelle, capacité d'abstraction b.y.-	mot, mot juste b.o.-	dialectique, capacité de raisonner, raison b.e.-	signe linguistique b.u.-	conteur b.a.-	médium - médias, outil de communication b.i.-
gestion de la mémoire, mnémotechnique t.y.-	temps, moment t.o.-	éloquence, capacité de convaincre, métonymie t.e.-	marque, signe durable t.u.-	scribe t.a.-	contenant, outil de conservation t.i.-
intelligence relationnelle k.y.-	philanthropie k.o.-	maîtrise de soi k.e.-	symbole, signe conventionnel k.u.-	roi, directeur, président, chef, dirigeant k.a.-	don k.i.-
intelligence émotionnelle m.y.-	empathie m.o.-	capacité d'interaction ouverte, prise en compte de l'autre m.e.-	sympôme, signe d'affect m.u.-	parent m.a.-	jouet, outil de séduction m.i.-
premiers principes, métaphysique n.y.-	ouverture au vaste monde n.o.-	capacité de diriger n.e.-	nom, place symbolique n.u.-	jugé n.a.-	feu - outil n.i.-
connaissance factuelle d.y.-	évidence, clarté intellectuelle d.o.-	habiletés sensori-motrices d.e.-	pointeur, signe indiciel d.u.-	chercheur, enquêteur d.a.-	instrument d'observation, mètre, instrument de mesure d.i.-
intelligence du corps l.y.-	plaisir l.o.-	habiletés techniques l.e.-	sensation, signe phénoménal l.u.-	guérisseur, soignant l.a.-	organe l.i.-
conscience des limites l.y.-	disposition dans l'espace l.o.-	compétence en ingénierie l.e.-	trace, signe matériel l.u.-	gardien, protecteur, défenseur l.a.-	nœud, outil de connexion l.i.-

Figure produite par l'auteur

94

Le paradigme de la figure 6 est généré par la fonction « (s. ⊕ b. ⊕ t. ⊕ k. ⊕ m. ⊕ n. ⊕ d. ⊕ f. ⊕ l.) ⊗ (y. ⊕ o. ⊕ e. ⊕ u. ⊕ a. ⊕ i.) ».

95

Les six *colonnes* correspondent à des invariants en attribut : types de connaissance (invariant y.), d'orientations psychiques (invariant o.), de compétences (invariant e.), de signes (invariant u.), de rôles sociaux (invariant a.) et de fonctions techniques (invariant i.)⁹. Les neuf *rangées* correspondent à des invariants en substance et déclinent le sens des lettres s. b. t. k. m. n. d. f. l. selon leurs attributs.

Figure 7. Table des couleurs

<i>blanc éclatant</i> U:.e.-U:.i.-'	<i>gris clair</i> U:.e.-A:.i.-'	<i>blanc bleuté</i> U:.e.-S:.i.-'	<i>blanc beige</i> U:.e.-B:.i.-'	<i>blanc rosé</i> U:.e.-T:.i.-'
<i>gris foncé</i> A:.e.-U:.i.-'	<i>noir d'encre</i> A:.e.-A:.i.-'	<i>noir bleuté</i> A:.e.-S:.i.-'	<i>noir kaki</i> A:.e.-B:.i.-'	<i>noir brun</i> A:.e.-T:.i.-'
<i>bleu clair</i> S:.e.-U:.i.-'	<i>bleu nuit</i> S:.e.-A:.i.-'	<i>bleu roi</i> S:.e.-S:.i.-'	<i>bleu vert</i> S:.e.-B:.i.-'	<i>indigo</i> S:.e.-T:.i.-'
<i>jaune clair</i> B:.e.-U:.i.-'	<i>beige foncé</i> B:.e.-A:.i.-'	<i>jaune vert</i> B:.e.-S:.i.-'	<i>jaune d'or</i> B:.e.-B:.i.-'	<i>jaune orangé</i> B:.e.-T:.i.-'
<i>rose</i> T:.e.-U:.i.-'	<i>brun rouge</i> T:.e.-A:.i.-'	<i>violet</i> T:.e.-S:.i.-'	<i>rouge orangé</i> T:.e.-B:.i.-'	<i>rouge vermillon</i> T:.e.-T:.i.-'

Figure produite par l'auteur

96 Le paradigme représenté à la figure 7 est généré par la fonction « (U:.e.-
⊕ A:.e.- ⊕ S:.e.- ⊕ B:.e.- ⊕ T:.e.-) ⊗ (U:.i.- ⊕ A:.i.- ⊕ S:.i.- ⊕ B:.i.- ⊕ T:.i.-) ».

97 Les lettres invariantes e. à la seconde place et i. à la quatrième place servent à identifier le paradigme des couleurs. Les variables U: et A: représentent le blanc et le noir tandis que les variables S: B: et T: représentent les trois couleurs primaires bleu, jaune et rouge. Les nuances principales se trouvent en substance et les nuances secondaires en attribut.

Les phrases

La structure algébrique

98 On notera que l'anneau des *mots*, défini par un alphabet et deux opérations, comprend en principe un nombre immense de variables, mais que seules trois mille d'entre elles ont un sens lexical. On maintient « petit » le nombre de mots pour faciliter aussi bien le calcul sémantique automatique que la gestion cognitive de la langue pour les humains. Les *phrases* ont également une structure d'anneau non commutatif¹⁰, mais plus complexe que celui des mots. Leur alphabet n'est autre que le dictionnaire de 3 000 mots. Les phrases sont générées récursivement par une opération multiplicative non commutative à neuf rôles et des opérations additives qui correspondent aux jonctions commutatives (et, ou) ou non commutatives (mais, parce...).

La fonction syntagmatique

99 Une phrase se met entre parenthèses et la parenthèse fermante est suivie d'un point.

100 La fonction syntagmatique – qui génère les phrases en IEML – possède neuf rôles, numérotés de 0 à 8.

0. La *racine* contient le verbe (ou le nom en cas de phrase nominale) auquel les autres rôles sont subordonnés.

1. L'*initiateur* correspond au sujet de la grammaire traditionnelle ou au

premier actant de la grammaire moderne.

2. L'*interactant* correspond au complément d'objet direct ou indirect de la grammaire traditionnelle ou au deuxième actant de la grammaire moderne.

3. Le *destinataire* correspond au « datif » ou au troisième actant de la grammaire moderne.

4. La *cause* – en un sens large – est le quatrième actant, qui correspond à un rôle d'instrument, de causalité ou de structure ontologique (type de, partie de, etc.).

5. Le *temps* spécifie la temporalité de la phrase (durée, moment, etc.)

6. Le *lieu* spécifie la dimension spatiale de la phrase.

7. L'*intention* précise le but, la motivation ou le contexte social de l'action.

8. La *manière* précise les dimensions qualitatives, quantitatives, possessives et autres de la racine.

101 Certains rôles, mais non la racine, peuvent être vides et chaque rôle d'une phrase, sauf le dernier, se termine par une virgule. Les rôles 4, 5, 6, 7 et 8 peuvent être précisés au moyen de prépositions appartenant aux paradigmes de mots-outils respectivement affectés à ces rôles.

102 Les phrases ont quatre types de composants : les concepts, les flexions, les prépositions et les jonctions.

103 Les *concepts* – représentés par des mots pleins du dictionnaire ou par des phrases subordonnées – sont précédés d'un croisillon (#).

104 Les *flexions*, qui précèdent les mots pleins qu'elles qualifient et qui sont précédés d'un tilde (~) spécifient la classe grammaticale du concept (verbe, nom, adjectif, adverbe). Elles précisent aussi le genre, le nombre ainsi que d'autres déterminants pour les noms, mais également le mode, le temps, l'aspect et d'autres déterminants pour les verbes.

105 Les *prépositions* ou cas, sont toujours précédés d'une étoile (*). La préposition est analogue à un *paramètre* du rôle grammatical.

106 Les *jonctions* peuvent être commutatives (« et », « ou inclusif », « ou exclusif ») ou non commutatives (comme « mais », « donc », etc.). Elles sont précédées d'une esperluette (&) et les phrases, rôles ou concepts jonctés ¹¹ sont mis entre crochets droits ([]).

107 Dans les exemples qui suivent, les mots et phrases IEML sont remplacés par leurs identifiants en français afin de faciliter la lecture. Le lecteur doit toutefois garder en tête que les phrases IEML sont construites récursivement à partir des mots du dictionnaire et qu'elles se présentent pour la machine comme des séquences ponctuées des six symboles de l'alphabet IEML.

108 « Dans la maison, la maman raconte une histoire d'un livre, à son enfant, avec amour, parmi les rires joyeux et les caresses, avant qu'il ne s'endorme. » La phrase entre guillemets qui précède est une traduction – littéraire – de la phrase IEML qui suit, et non l'inverse.

109 (0 racine ~indicatif ~présent #raconter, 1 initiateur ~défini ~féminin #maman, 2 interactant ~indéfini #récit, 3 destinataire ~défini #enfant, 4 causalité *instrument ~indéfini #livre, 5 temps *avant #dormir, 6 lieu *dans ~défini #maison, 7 intention *coloré par le sentiment #amour, 8 manière *avec &et [~pluriel #rire joyeux ~pluriel #caresse]).

L'accent sémantique

110 Par défaut, l'accent sémantique (Langacker 1987) d'une phrase porte sur sa racine. Il est néanmoins possible de mettre l'accent sémantique sur un autre rôle de la phrase au moyen d'un point d'exclamation. Par exemple la phrase ci-dessous, avec le point d'exclamation au rôle 2, se lit ainsi « *L'histoire* que la maman raconte à son enfant à partir d'un livre, avec amour, parmi les rires joyeux et les caresses, avant qu'il ne s'endorme. » Si le point d'exclamation avait été au rôle 3, la phrase se lirait : « *L'enfant* dont la maman raconte une histoire dans un livre », etc.

```
111 (
    0 racine ~indicatif ~présent #raconter,
    1 initiateur ~défini ~féminin #maman,
    2 interactant ! ~défini #récit,
    3 destinataire ~défini #enfant,
    4 causalité *instrument ~indéfini #livre,
    5 temps *avant #dormir,
    6 lieu *dans ~défini #maison,
    7 intention *coloré par le sentiment #amour,
    8 manière *avec &et [~pluriel #rire joyeux ~pluriel #caresse]
).
```

112 Lorsque l'accent sémantique porte sur un rôle syntagmatique avec un concept vide, il désigne le concept prototypique correspondant à ce rôle. Par exemple, le concept prototypique de l'initiateur (ou sujet) du verbe « commettre un crime » est « criminel ». On peut désigner de la même manière l'objet, la victime ou le lieu d'un crime. Cette notation minimise le nombre de mots nécessaires pour exprimer les concepts. Elle peut servir à expliciter le *frame* (au sens de Fillmore) d'un concept et certaines fonctions lexicales. La notation se présente comme dans l'exemple ci-dessous.

```
113 @node
    fr: commettre un crime
    en: to commit a crime
    (
        0 racine ~verbe #commettre un crime
    ).

    @node
    fr: un criminel
    en: a criminal
    (
        0 racine ~verbe #commettre un crime,
        1 initiateur ! ~indéfini #vide
    ).
```

114 On procède de la même manière pour définir l'objet, la victime ou la scène d'un crime en marquant respectivement d'un point d'exclamation les rôles vides de l'interactant, du destinataire et du complément de lieu.

Les paradigmes de phrases

115 Un paradigme de phrases est une phrase composée de deux parties : une partie constante et une partie variable, qui peut concerner un, deux ou trois (au maximum) rôles. Les variables se présentent comme des #concepts entre accolades et sont séparées par un point-virgule.

116 Dans l'exemple ci-dessous, le rôle 1 comporte un groupe de substitution de cinq variables. « 1d: /#/1 » indique que le paradigme n'a qu'une seule dimension (un seul rôle variable) et que ce rôle est l'initiateur (rôle numéro 1).

```
117 @paranode
    fr: paradigme des fonctions psychiques
    en: psychic functions paradigm
    1d: /#/1
    (
        0 #symptômes en santé mentale,
        1 { #autonomie psychique ; #dimension subjective ; #dimension émotionnelle ;
            #dimension cognitive ; #mémoire personnelle }
    ).
```

Dans le contexte de la construction d'une ontologie de la santé mentale en IEML ¹², j'ai élaboré la figure 8, qui organise les grandes catégories de symptômes dans ce domaine. On voit que le paradigme ci-dessus énumère les cinq rangées principales de la figure 8 ci-dessous.

Figure 8. Table des catégories de symptômes les plus générales dans une ontologie de la santé mentale

répertoire des états mentaux pour la psychiatrie			
Autonomie psychique	auto-représentation -	auto-réparation -	auto-régulation -
	insight en contexte psychiatrique	auto-réparation mentale	symptômes liés au jugement
	présentation de soi en contexte psychiatrique	auto-réparation physique	symptômes relatifs au comportement
Subjectivation	auto-référence -	inter-subjectivité -	pensée -
	auto-référence psychologique	disposition intersubjective -	processus de pensée
	auto-référence somato-sociale	cognition sociale -	contenu de la pensée
Emotion	qualité émotionnelle -	force de l'émotion -	cohérence émotionnelle -
	qualité de l'humeur -	orientation - interne-externe - de l'émotion	cohérence humeur-affect
	qualité de l'objet -	intensité de l'émotion -	cohérence humeur-situation
Cognition	langage -	raisonnement -	sensori-motricité -
	compréhension du langage -	réflexion et abstraction -	problèmes de perception -
	production linguistique -	planification et exécution -	problèmes de motricité -
Mémoire personnelle	conscience -	mémoire intermédiaire -	mémoire à long terme -
	conscience subliminale -	mémoire à court terme -	mémoire déclarative -
	attention -	mémoire de travail -	mémoire non-déclarative

Figure produite par l'auteur

Je me suis servi des symétries fondamentales d'IEML pour ordonner l'univers de discours des symptômes en santé mentale. La rangée du haut correspond aux qualités émergentes du psychisme dans sa dimension la plus virtuelle (l'autonomie), tandis que la rangée du bas correspond à la dimension la plus actuelle ou concrète, qui supporte les autres fonctions psychiques (la mémoire). Entre ces deux dimensions, l'information montante et descendante se trouve codée et décodée sur un mode subjectif (signe), émotionnel (être) et cognitif (chose), comme on le voit dans les trois rangées intermédiaires. Les trois colonnes de la table correspondent en gros à une partition signe/être/chose et la subdivision binaire au sein de chacun des quinze rectangles correspond à une répartition entre les aspects les plus virtuels (pour la cellule du haut) et les plus actuels (pour la cellule du bas) ¹³.

Pour construire une telle table en IEML, il suffit d'organiser les paradigmes au moyen d'une arborescence dans laquelle les variables d'un paradigme de niveau supérieur (plus général) deviennent les constantes d'un paradigme de niveau inférieur (plus détaillé).

Par exemple, pour créer les trois colonnes de la cinquième rangée concernant la mémoire, on écrit le paradigme qui suit.

```
@paranode
fr: paradigme des symptômes liés à la mémoire personnelle
en: paradigm of symptoms related to the personal memory
id: /#/2
(
  0 #symptômes en santé mentale,
  1 #mémoire personnelle,
  2 { #conscience ; #mémoire intermédiaire ; #mémoire à long terme }
).
```

123 Pour préciser le type de symptômes concernant la mémoire à long terme, on écrit le paradigme suivant.

```
124 @paranode
fr: paradigme des symptômes liés à la mémoire à long terme
en: paradigm of symptoms related to the long term memory
id: /#/4
(
  0 #symptômes en santé mentale,
  1 #mémoire personnelle,
  2 #mémoire à long terme,
  4 { #déclarative ; #non déclarative }
).
```

125 Une fonction de l'éditeur IEML permet d'emboîter les paradigmes de phrases les uns dans les autres pour créer des arbres paradigmatiques complexes et générer automatiquement des tables.

126 Par exemple, la cellule « problèmes de perception », dans la troisième colonne de la quatrième rangée de la figure 8, contient à son tour le paradigme à deux dimensions de la figure 9 ci-dessous.

Figure 9. Table des problèmes de perception

problèmes de perception					
hypersensibilité de l'audition	hypersensibilité de la vue	hypersensibilité de l'odorat	hypersensibilité du goût	hypersensibilité du toucher	hypersensibilité de la cenesthésie
handicap de l'audition	handicap de la vue	handicap de l'odorat	handicap du goût	handicap du toucher	handicap de la cenesthésie
douleur auditive	douleur de la vue	douleur de l'odorat	douleur du goût	douleur tactile	douleur de la cenesthésie
illusions auditives	illusions visuelles	illusions olfactives	illusions gustatives	illusions tactiles	illusions cenesthésiques
hallucinations auditives	hallucinations visuelles	hallucinations olfactives	hallucinations gustatives	hallucinations tactiles	hallucinations cenesthésiques

Figure produite par l'auteur

127 Cette table est générée par le paradigme de phrase ci-dessous.

```
128 @paranode
fr: paradigme des symptômes liés à la perception
en: paradigm of symptoms related to perception
id: /#/1
2d: /#/2
(
  0 #problèmes de perception -,
  1 { #hypersensibilité ; #handicap ; #douleur ; #illusions ;
    #hallucinations },
  2 { #sens de l'ouïe ; #sens de la vue ; #sens de l'odorat ;
    #sens du goût ; #sens du toucher ; #cénesthésie }
).
```

Les relations

Les relations syntagmatiques

129 Une phrase IEML est un arbre avec un rôle racine et des rôles feuille. Une phrase crée donc automatiquement un nœud de relations entre les concepts qui occupent ses différents rôles. Les relations entre racine et feuilles sont définies par la syntaxe : #B est le sujet de #A, #C est l'objet de #A, #D est le complément de lieu de #A, etc. Les liens du graphe syntagmatique sont étiquetés par la relation racine-feuille.

130 Les concepts qui occupent les rôles d'une phrase peuvent être des mots ou des phrases. Si ce sont des phrases, ils se décomposent eux-mêmes en rôles syntagmatiques, et ainsi de suite récursivement, jusqu'à ce que l'on arrive à un mot. Ces relations de composition emboîtées sont enregistrées par la syntaxe du langage et permettent une navigation hypertextuelle dans l'arbre conceptuel.

Les relations paradigmatiques

131 Les relations paradigmatiques connectent les concepts qui sont générés par la même fonction, définie par ses opérations, ses constantes et ses variables. Le graphe qui réunit les concepts selon leurs relations paradigmatiques n'est pas hiérarchique, comme dans le cas précédent, mais de type matriciel. Les concepts générés par la même fonction ont en commun une partie invariante et un domaine de variation. Ils appartiennent donc à un même système de symétries. En IEML, le système de symétries des signifiants est analogue au système de symétries des signifiés qui lui correspond. Alors que les relations syntagmatiques connectent les concepts en arbres, les relations paradigmatiques les connectent en matrices « horizontales », perpendiculaires à l'ordre hiérarchique. En somme, la grille syntaxique fondamentale d'IEML croise hiérarchies syntagmatiques et matrices paradigmatiques en un graphe fractal régulier, un réseau sémantique complexe qu'on peut appeler rhizomatique (Deleuze et Guattari 1980).

Les liens conditionnels

132 Sur la base de la grille fondamentale qui vient d'être évoquée, il est possible de générer des relations supplémentaires au moyen de fonctions conditionnelles. Pour générer des relations entre des concepts en IEML, on procède en deux étapes : (1) l'écriture d'une phrase de *lien*, qui représente une relation virtuelle, et (2) la formulation d'une *fonction* qui va créer des relations actuelles entre des nœuds conceptuels déterminés.

133 Le lien, qui peut s'appliquer à un grand nombre de nœuds, se contente de lister des « arguments » (\$A, \$B, \$C...) et de les situer dans la phrase *en position de référence*. Le lien peut être unaire, binaire, ternaire, voire plus puisque la phrase possède neuf rôles distincts.

134 Par exemple, le lien ci-dessous, qui se nomme « est le contraire de », est construit pour réunir deux nœuds. Il est possible de construire des relations comprenant entre un et neuf nœuds (on trouvera dans le dictionnaire des exemples de liens unaires et ternaires).

135 La *fonction* qui actualise le lien mentionne le *domaine* de variation des arguments (généralement un paradigme) et les *conditions* de l'établissement des relations en termes de *contenus des rôles syntaxiques*. Les conditions élémentaires peuvent être connectées au moyen de NON, de ET et de OU (les connecteurs booléens). Voici un exemple concernant la relation : « \$A est le contraire de \$B », avec le lien d'abord et la fonction ensuite.

136

```
@link
  args: ($A, $B)
  fr: est le contraire de
  en: is the contrary of
  template-fr: $A est le contraire de $B
  template-en: $A is the contrary of $B
  (
    0 #signifier,
    1 #mot <$A>,
    8 *contraire #mot <$B>
  ).
```

137 Le même lien peut s'appliquer à un grand nombre de domaines, mais à chaque domaine correspond un ensemble de *conditions* spécifiques. La fonction ci-dessous est appliquée au paradigme des gradients (voir la figure 10 ci-dessous).

```

@function
type: word
link: est le contraire de
domain: ($A in "F:.o.E:+S:+B:+T:.-", $B in "F:.o.E:+S:+B:+T:.-")
condition:
(
  $A.substance == $B.substance
  and
  $A.mode.substance == "S:"
  and
  $B.mode.substance == "T:"
).

```

« mode.substance » se lit : « substance du mode ». Le même lien « est le contraire de » peut être appliqué à d'autres domaines (à d'autres paradigmes) mais avec des fonctions différentes, par exemple ceux qui concernent des relations spatiales (« en haut » est le contraire de « en bas »), des relations temporelles (« avant » est le contraire de « après ») et ainsi de suite. Les fonctions peuvent aussi connecter des liens appartenant à des paradigmes différents.

Figure 10. Table du paradigme des gradients

<i>qualité - bonne ou mauvaise</i> U:.o.E:.-	<i>malice, mal, méchanceté</i> U:.o.S:.-	<i>qualité moyenne</i> U:.o.B:.-	<i>bonté, bien</i> U:.o.T:.-
<i>taille</i> A:.o.E:.-	<i>petitesse</i> A:.o.S:.-	<i>taille moyenne</i> A:.o.B:.-	<i>grandeur</i> A:.o.T:.-
<i>probabilité</i> S:.o.E:.-	<i>impossibilité</i> S:.o.S:.-	<i>occasion, contingence</i> S:.o.B:.-	<i>éternité, constance, nécessité</i> S:.o.T:.-
<i>pattern temporel, fréquence, distribution dans le temps</i> B:.o.E:.-	<i>rareté - dans le temps, basse fréquence</i> B:.o.S:.-	<i>régularité, périodicité</i> B:.o.B:.-	<i>haute fréquence</i> B:.o.T:.-
<i>distribution dans l'espace, pattern spatial</i> T:.o.E:.-	<i>absence</i> T:.o.S:.-	<i>présence partielle</i> T:.o.B:.-	<i>présence</i> T:.o.T:.-

Ce paradigme est généré par la fonction « F:.o.E:+S:+B:+T:.- » (pour mémoire, « F: = U: ⊕ A: ⊕ S: ⊕ B: ⊕ T: »).

Figure produite par l'auteur

La fonction conditionnelle ci-dessus génère automatiquement les cinq phrases qui suivent : « mal est le contraire de bien » ; « petitesse est le contraire de grandeur » ; « impossibilité est le contraire de nécessité » ; « rare est le contraire de fréquent » ; « absent est le contraire de présent ».

Dans la figure 9, on remarquera qu'un mot IEML peut se traduire par plusieurs expressions en langues naturelles, qui décrivent l'extension de son champ sémantique. On n'oubliera pas d'autre part que les concepts ont une traduction nominale (ici dans la table), mais aussi verbale, adjectivale et adverbiale.

142 Notre exemple concernait la création de relations entre des mots. La création de relations entre les phrases obéit aux mêmes principes, à savoir, dans un premier temps, l'écriture d'une phrase de lien avec des arguments en position de référence et, dans un second temps, l'écriture de fonctions qui spécifient (a) les domaines de variation des arguments – généralement des paradigmes de phrases emboîtés – et (b) les conditions d'établissement des relations en termes de contenus des rôles syntagmatiques.

143 Il est enfin possible de spécifier les propriétés mathématiques de chaque relation binaire créée par un lien : est-elle réflexive, symétrique, transitive ?

144 Le point capital est ici que des *relations sémantiques* sont créées automatiquement à partir de *conditions syntaxiques*. Cela n'est évidemment possible que parce que les mots et les phrases appartiennent à des *patterns* de constantes et de variables dans des fonctions algébriques, et que ces *patterns* reflètent des symétries sémantiques.

145 IEML est conçu pour faciliter la génération automatique de concepts et de relations sémantiques entre les concepts. Ce langage est donc particulièrement adapté à la création et à l'exploration de réseaux sémantiques : modèles de données, graphes de connaissance, ontologies et autres hypertextes. Puisque les réseaux de concepts en IEML sont composés à partir du même dictionnaire compact de 3 000 mots et de la même grammaire régulière, ils sont sémantiquement interopérables.

La référence et l'autoréférence

146 En IEML, la référence se met entre chevrons (< >). La référence peut contenir un concept IEML (comme nous venons de le voir), elle peut aussi contenir un « littéral » comme un nom propre, une adresse, une date, une position GPS, un nombre, une unité de mesure, une monnaie, un objet géographique, une URL, une URL pointant vers une image, etc. Le littéral est un signifiant n'ayant pas de signifié en IEML. Son sens est donc purement référentiel. Enfin, le contenu des chevrons peut être de type mixte s'il s'agit d'une phrase IEML qui possède des littéraux en référence.

147 Les douze premiers nombres entiers naturels¹⁴ sont néanmoins considérés comme des concepts IEML. Ils entretiennent des relations sémantiques avec les nombres ordinaux, les symétries et les figures géométriques régulières. Les grandes zones géographiques¹⁵ existent également en IEML. Elles peuvent être assimilées à des codes postaux pouvant donner lieu à des calculs sémantiques. Ces codes permettent notamment de déterminer les positions respectives (au nord, à l'est...) des zones codées, ainsi que de situer et regrouper les pays, les villes et autres objets géographiques.

148 Par exemple, pour dire « l'Italie » en IEML, on écrit : (o ~défini #Europe centre-sud <Italia>). Pour dire « le nombre 292 », on écrit : (o ~défini #nombre <292>). Pour désigner un nom propre : (o ~masculin #sculpteur <Rodin>). Pour désigner un hyperlien : (o ~défini #hyperlien <https://ieml.intlekt.io>).

149

Le lecteur contrastera l'approche d'IEML avec celle du Web sémantique (Berners-Lee 2013), dans lequel les URI¹⁶ sont des désignateurs rigides – des littéraux – et ne peuvent pas faire l'objet de calculs sémantiques à partir de leur forme matérielle. Bien entendu, l'approche d'IEML et celle du Web sémantique ne sont pas incompatibles puisque les expressions IEML valides ou USL (*Uniform Semantic Locators*) ont une forme unique et peuvent se représenter comme des URI.

150

Nous avons donné plus haut quelques exemples d'autoréférence en expliquant la création de relations conditionnelles. Chacun des trois mille mots du dictionnaire d'IEML se définit au moyen de phrases utilisant d'autres mots et chaque expression complexe en IEML (mot, phrase, graphe sémantique) renvoie au noyau circulaire d'interdéfinition du dictionnaire. Cette interdéfinition circulaire est d'ailleurs, comme nous l'avons déjà noté, le propre de toutes les langues. Enfin, indépendamment du dictionnaire, IEML peut faire référence à ses propres expressions afin de définir des relations, de proposer des définitions, de développer des raisonnements ou de faire des commentaires.

IEML est une langue philologique à la sémantique calculable

151

Il est maintenant temps de vérifier qu'IEML tient sa promesse de coder la sémantique sur un mode calculable tout en manifestant les propriétés d'une langue philologique.

Une langue mathématique

152

La syntaxe d'IEML est mathématique. Elle est définie par deux fonctions emboîtées. Pour créer des mots, une fonction *morphologique* à trois rôles multiplicatifs opère sur un alphabet de six variables primitives. Pour créer les phrases, une fonction *syntagmatique* à neuf rôles multiplicatifs opère sur l'alphabet de 3 000 mots généré par la fonction morphologique. Les *paradigmes* et leurs symétries ne sont rien d'autre que les matrices générées par ces fonctions lorsqu'un, deux ou trois de leurs rôles sont variables. Chaque *mot* du dictionnaire d'IEML appartient à un paradigme et un seul. Une *phrase*, en revanche, peut appartenir à plusieurs paradigmes et les paradigmes de phrases sont librement créés par les interlocuteurs. Sur la grille régulière qui croise les arbres syntagmatiques et les matrices paradigmatiques, des fonctions de création de liens permettent de tisser des relations sémantiques à partir de conditions syntaxiques et de créer autant d'hypertextes ou de graphes de connaissance que l'on voudra. En somme, la sémantique d'IEML est calculable.

Une langue philologique

153 La structure de la phrase IEML autorise la description de n'importe quelle interaction complexe et l'usage du point d'exclamation (!) permet de mettre l'accent sur tel ou tel rôle de la phrase. Les verbes peuvent être affirmatifs, négatifs, interrogatifs ou autres et ils déclinent toute la gamme des modalités logiques et des modes grammaticaux. La phrase est évidemment récursive puisque les concepts qui la composent peuvent être actualisés par des mots ou des phrases, phrases qui peuvent à leur tour contenir des phrases, et ainsi de suite. IEML possède donc la puissance narrative des langues naturelles. Le dictionnaire d'IEML contient toutes sortes de déictiques et sa grammaire permet d'explicitier les opérations de référence (extralinguistique) et d'autoréférence (références d'expressions linguistiques à d'autres expressions). Au chapitre de l'explicitation des rôles dialogiques, IEML contient évidemment les trois personnes grammaticales. Enfin, notre métalangage permet en principe de traduire n'importe quel concept, récit ou raisonnement exprimé en langue naturelle. Le parseur de l'éditeur IEML permet d'ores et déjà de rédiger des textes en IEML en n'utilisant que des mots en français ou en anglais.

154 L'actuel éditeur *open source* mono-utilisateur – un simple démonstrateur – devra céder la place à une plateforme où pourront collaborer éditeurs d'ontologies et curateurs de données. Dans cette encyclopédie méta-ontologique d'un nouveau genre, chaque concept pourrait être lu dans toutes les langues et situé dans les paradigmes et les graphes de connaissance auxquels il participe. On pourrait voir les concepts qui le composent, dans quels concepts de couche supérieure il est utilisé et selon quels rôles grammaticaux. Ce réseau conceptuel – vivant, évolutif, en expansion, différencié selon une multitude ouverte de finalités pratiques – servirait à faire face à l'immense flux de données qui monte des activités humaines et le fond mémoriel où il s'accumule.

Quelques applications

155 Je m'inscris dans la perspective d'une intelligence artificielle au service de l'intelligence collective. J'ai conçu IEML comme s'il devait servir de protocole sémantique permettant la communication des significations et des connaissances dans la mémoire numérique tout en optimisant l'apprentissage et le raisonnement automatiques. Il faut imaginer une multitude de communautés aux finalités et aux pratiques hétérogènes. Chacune d'elles s'organiserait à partir d'une base de données indexée en IEML selon un graphe de connaissances original et évolutif. Les modes de comptabilité et d'allocation de la valeur de ces communautés seraient définies par des règles – ou des contrats intelligents – en IEML, donc compréhensibles par des humains et exécutables par des machines. Malgré la diversité de leurs conventions et de leurs savoirs, ces communautés pourraient échanger leurs données publiques¹⁷, leurs modules ontologiques et leurs règles formelles dans un réseau interopérable de bases de connaissances alimentant les apprentissages automatiques.

156 Dans cette dernière section, je m'attache à décrire certaines des applications *rendues possibles* par l'existence d'IEML, mais sans aucune prétention à la *prévision*. Comme toute langue, IEML possède une fonction

communicative. Il peut servir de langage pivot entre langues naturelles et donc entre humains. Il peut également jouer le rôle de langage de communication entre humains et machines, remédiant aux ambiguïtés et aux erreurs qui obèrent les *chatbots* (assistants virtuels) fondés sur l'apprentissage statistique à partir de textes en langues naturelles. Grâce à ses propriétés formelles, IEML permettrait aux ordinateurs de « comprendre » les expressions linguistiques beaucoup mieux qu'ils ne le font aujourd'hui. Notre langue mathématique peut enfin servir de code de communication entre machines. Il va de soi que, dans toutes ses applications, le code IEML brut ne serait manipulé que par les machines et que le contenu serait toujours présenté aux humains – ou fourni par eux – sous forme de texte en langues naturelles, de diagrammes ou d'icônes.

157 Cette section sur les applications d'IEML s'organise en trois parties successives dont les thèmes s'entrelacent étroitement. Je traite d'abord de l'organisation de la mémoire numérique et de ce qu'un système de métadonnées sémantiques basé sur IEML pourrait apporter à l'intelligence collective. J'aborde ensuite la manière dont IEML pourrait augmenter l'intelligence artificielle dans chacune de ses deux branches symboliques et neuronale-statistique, tout en favorisant l'intégration neurosymbolique. Je termine en évoquant la perspective d'un système de coordonnées sémantiques qui unifierait les sciences humaines en ouvrant la voie à un univers de sens calculable, mais inépuisablement complexe.

L'organisation de la mémoire numérique

Métadonnées et interopérabilité sémantique

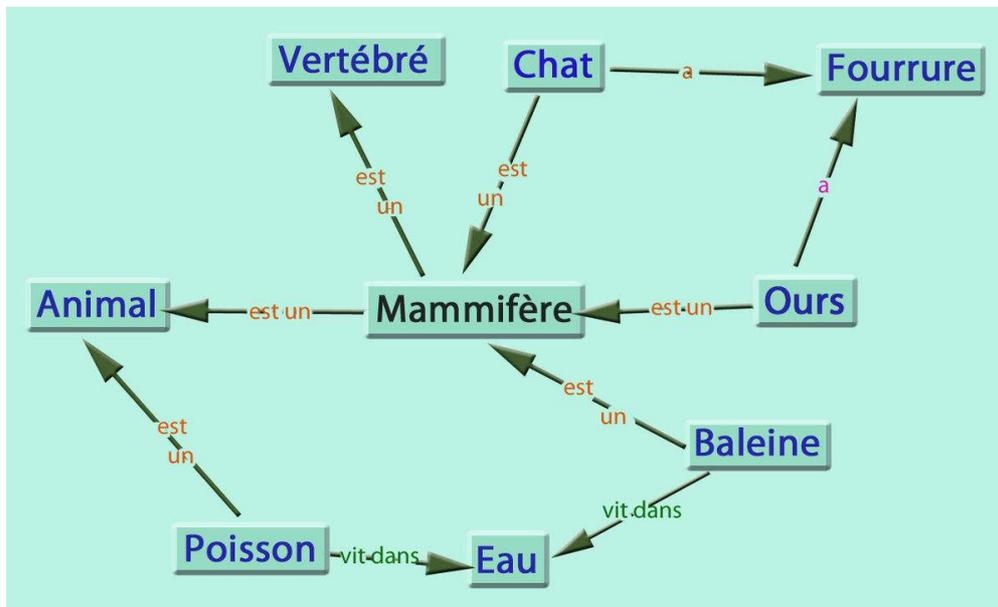
158 L'intelligence collective humaine repose largement sur une communication stigmergique (Bonabeau 1999 ; Dorigo, Bonabeau et Theraulaz 2000) dont le médium s'identifie de manière croissante à la mémoire numérique. Mais il reste à surmonter les cloisonnements dus à la multiplicité des langues, des systèmes de classification, des points de vue disciplinaires et des contextes pratiques. Si l'organisation de la mémoire conditionne la dynamique de notre intelligence collective et si sa fragmentation diminue notre puissance cognitive, nous devons nous intéresser de près aux systèmes de métadonnées.

159 Les systèmes d'information et applications logicielles organisent l'information au moyen de métadonnées. Les métadonnées sont les données qui organisent les données. Dans le labyrinthe de la mémoire, elles font office de fil d'Ariane. Leur fonction est d'identifier les informations afin de mieux les ranger et les retrouver. Elles servent moins à décrire exhaustivement les choses – il ne s'agit pas de faire des cartes à la même échelle que le territoire – qu'à fournir des repères à partir desquels les utilisateurs pourront trouver ce qu'ils cherchent, à l'aide d'algorithmes. La notion de métadonnée, très générale, embrasse toute la gamme des systèmes d'étiquetage qui permettent d'organiser ou de classer les données.

160 On peut distinguer (a) les métadonnées *matérielles*, comme le format d'un fichier, sa date de création, son auteur ou sa licence d'utilisation et (b) les métadonnées *sémantiques* qui concernent le *contenu* d'un document ou d'un ensemble de données : *de quoi* parlent les données ; ainsi que leur dimension pratique : *à quoi* servent les données, *à qui* et *dans quelles circonstances*.

Nous nous intéressons ici principalement aux métadonnées sémantiques. Un système de métadonnées sémantiques peut être aussi simple qu'un vocabulaire. Au niveau de complexité supérieur, cela peut être une classification hiérarchique ou taxonomie. Au niveau le plus complexe, c'est une « ontologie », c'est-à-dire la modélisation d'un domaine de connaissance ou de pratique, qui peut contenir plusieurs taxonomies avec des relations transversales, y compris des relations causales et des possibilités de raisonnement automatique. La notion d'ontologie cède aujourd'hui la place à celle de graphe de connaissances (*knowledge graph*) dans lequel on accède à des ressources numériques au moyen d'un modèle de données organisé en réseau de relations et d'un vocabulaire contrôlé. Les données ne sont plus contenues dans les schémas tabulaires fixes des bases de données relationnelles, mais dans les nouvelles bases de données de graphes, plus souples, plus faciles à faire évoluer, mieux à même de représenter des modèles complexes et autorisant plusieurs vues différentes. Comme une ontologie, un graphe de connaissances se prête au raisonnement automatique (relevant de l'IA symbolique classique), mais aussi à l'apprentissage automatique s'il est bien conçu et si les données sont en nombre suffisant.

Figure 11. Exemple de représentation d'un réseau sémantique



Lamiot, Wikipédia, 2010, https://fr.wikipedia.org/wiki/Fichier:SemanticNetArbre_sémantique_fr.jpg · CC BY-SA 3.0

Les métadonnées sémantiques constituent donc un élément essentiel des dispositifs d'intelligence artificielle. Elles sont notamment utilisées comme squelettes des graphes de connaissances mis en œuvre par les GAFAM afin d'organiser leurs bases de connaissances géantes et, de plus en plus, dans de grandes et moyennes entreprises. Elles sont aussi utilisées – sous le nom de *labels* (étiquettes) ou d'*annotations* – pour catégoriser les données d'entraînement des modèles de *deep learning*.

Parce qu'ils structurent la connaissance contemporaine, dont le support est numérique, les systèmes de métadonnées représentent un enjeu considérable aux niveaux scientifique, culturel et politique. Afin de mettre un peu d'ordre et de transparence dans la mémoire numérique, on a évidemment élaboré des standards. Pour saisir le paysage contemporain des standards de métadonnées, il importe de distinguer entre les *formats* et les *modèles*.

Les formats. Le système de formats et de langages standards proposé par le World Wide Web Consortium¹⁸ pour atteindre le « Web sémantique » existe depuis la fin du xx^e siècle. Les quatre standards principaux sont : XML, RDF, OWL et SPARQL. XML décrit la structure hiérarchique d'un document ; RDF organise des ressources sous la forme de graphes de relations ; OWL programme des ontologies pour le raisonnement automatique ; SPARQL permet d'interroger des graphes RDF. Il faut bien comprendre que le système de standards du W3C concerne les *formats* de fichiers et de programmes traitant les métadonnées sémantiques mais *pas la sémantique proprement dite*, à savoir les catégories, concepts, propriétés, événements et relations qui sont exprimés en langues naturelles, avec toutes les ambiguïtés, multiplicités et incompatibilités que cela implique.

Les modèles. Au-dessus de ce système de formats standards existent des *modèles* standards pour traiter le contenu proprement sémantique des concepts et de leurs relations. En effet, un même *modèle conceptuel* (Matta et Aussenac-Gilles 1996) peut être utilisé pour structurer des données différentes favorisant ainsi la communication entre mémoires distinctes. Les modèles tels que schema.org¹⁹ pour les sites Web, soutenu par Google, ou CIDOC CRM²⁰, développé par les institutions de conservation des patrimoines culturels en sont de bons exemples. Des modèles standards ont été publiés pour de nombreux domaines, de la finance à la médecine en passant par la logistique. Hélas, il existe souvent plusieurs modèles concurrents pour un domaine²¹ et les modèles eux-mêmes sont hyper-complexes, au point que même les spécialistes d'un modèle n'en maîtrisent qu'une petite partie. Notons en outre que ces modèles sont exprimés en langues naturelles et le plus souvent en anglais. Les outils d'édition et de gestion de modèles de métadonnées les plus sophistiqués du marché (tels que PoolParty²², Ab Initio²³ ou Synaptica²⁴) n'ont pas moyen de distinguer clairement entre les mots ou termes dans une langue naturelle particulière et les concepts ou catégories, qui sont des notions plus abstraites et translinguistiques. Cela pose problème parce que le même concept peut être exprimé par différents mots dans la même langue et *a fortiori* dans différentes langues. De plus, le même mot peut correspondre à plusieurs concepts, y compris dans la même langue. Le cardinal est-il un dignitaire ecclésiastique, un oiseau ou un glaïeul ? Une taupe est-elle un animal fouisseur ou un agent ennemi infiltré ? Les mots des langues naturelles sont ambigus et multiples, mais reconnaissables par des humains immergés dans un contexte particulier. Or les machines ont bien du mal à effectuer des inférences contextuelles. D'où la nécessité d'identifier de manière unique les concepts formels sous-jacents pour qu'ils soient interprétables par les machines.

En pratique, les taxonomies, ontologies, graphes de connaissances et autres systèmes de métadonnées mis en œuvre dans des applications réelles pour organiser des ensembles de données n'utilisent les formats et modèles standards que de manière partielle. Les utilisateurs se soumettent tant bien que mal à ces couches de standards dans l'espoir que leurs données et applications deviendront les heureux sujets d'un royaume de l'interopérabilité sémantique. Mais leurs espoirs sont souvent déçus. L'idéal du Web intelligent et décentralisé de la fin des années 1990 a cédé la place au *Search Engine Optimization* (SEO) plus ou moins aligné sur le *knowledge graph* (secret !) de Google.

Aujourd'hui encore, une grande partie de la mémoire numérique se trouve dans des bases de données dites relationnelles, sans distinction claire des données et des métadonnées, organisées selon des schémas mutuellement incompatibles, mal optimisées pour les besoins de connaissance, de coordination ou d'aide à la décision de leurs utilisateurs. À cause de la multitude des formats, des modèles et des applications disparates, sans parler des différences de langues, il reste bien difficile de faire communiquer des cultures, des domaines d'affaires ou des disciplines différentes. Il n'existe malheureusement pas d'interopérabilité sémantique au sein de domaines aussi importants pour le développement humain que la santé, l'éducation ou le droit. Même chose pour les compétences et le marché de l'emploi, chaque secteur et chaque pays cultivant des nomenclatures différentes. La situation dans les chaînes d'approvisionnement, le transport, le commerce international et l'Internet des choses n'est pas moins chaotique. La perspective des villes intelligentes s'éloigne lorsque l'on enquête sur le désordre des métadonnées urbaines. Il faudrait citer tous les secteurs, mais je voudrais souligner particulièrement l'absence d'interopérabilité sémantique dans le monde des encyclopédies, des musées, des bibliothèques et de la recherche en sciences humaines. La plupart des domaines de connaissance et de pratique possèdent donc aujourd'hui une mémoire numérique fragmentée. Quant à l'interopérabilité sémantique entre secteurs d'activités et administrations, si importante dans les situations d'urgence, de catastrophe ou de mobilisation des populations, elle traîne loin derrière.

Vers un protocole sémantique

Des images, des sons, des odeurs, des lieux et des émotions balisent la mémoire humaine, comme celle des animaux. Mais c'est le langage qui unifie, ordonne et réinterprète à volonté notre mémoire symbolique. La chose est vraie non seulement à l'échelle individuelle, mais aussi à l'échelle collective, par la transmission des récits et l'écriture. Grâce au langage, l'humanité a pu accéder à l'intelligence réflexive (Bickerton 1995 ; Deacon 1997). Par analogie, je pense que nous n'accéderons à une intelligence collective réflexive qu'en prenant acte de l'évolution vers une communication stigmergique et en adoptant une langue-écriture adéquate à l'organisation de la mémoire numérique. Il reste à accomplir dans la sphère sémantique ce qui a déjà été mené à bien pour l'espace, le temps (Galison 2003) et diverses unités de mesure : établir un système de coordonnées universel et régulier qui favoriserait la modélisation formelle et l'harmonisation des opérations. La puissance d'enregistrement et de calcul de centres de données gigantesques rend désormais cet idéal matériellement atteignable. L'histoire du calcul automatique est certes scandée par l'évolution du matériel et des applications, mais aussi par une impressionnante stratification de systèmes d'exploitation, de codes, de protocoles et de standards. Il n'existe aucune raison pour que ce millefeuille de normes logicielles cesse d'évoluer.

Pourquoi ne pas adopter IEML comme protocole sémantique, puisque ses propriétés formelles favorisent le raisonnement automatique et l'exploration de données par proximité sémantique ? IEML est indépendant des formats et peut traduire n'importe quel modèle. La mémoire numé-

rique pourrait alors s'organiser de toutes les manières possibles selon la multitude des besoins pratiques et la diversité des points de vue, tout en autorisant des échanges fluides entre ses différentes parties et facettes.

L'intelligence artificielle neurosymbolique

IA neuronale et IA symbolique

170

Les deux branches de l'IA – neuronale et symbolique – existent depuis le milieu du xx^e siècle et elles correspondent à deux styles cognitifs également présents chez l'humain²⁵. Du côté neuronal, nous avons la reconnaissance de formes (*pattern recognition*) qui correspond à des modules sensorimoteurs réflexes, que ces derniers soient d'origine génétique ou appris. Du côté symbolique, nous avons une connaissance conceptuelle explicite et réfléchie, souvent organisée en modèles causaux et qui peut faire l'objet de raisonnements. Comme l'intelligence humaine est elle-même neurosymbolique, il n'existe aucune raison théorique pour ne pas tenter de faire coopérer les deux branches de l'IA dans des systèmes d'intelligence artificielle hybrides combinant apprentissage *et* raisonnement (D'Avila Garcez, Lamb et Gabbay 2009 ; D'Avila Garcez et Lamb 2020). Les bénéfices sont évidents et, en particulier, chacun des deux sous-systèmes peut remédier aux problèmes rencontrés par l'autre. Dans une IA hybride, la partie symbolique surmonte les difficultés de conceptualisation, de généralisation, de modélisation causale et de transparence de la partie neuronale. Symétriquement, la partie neuronale amène les capacités de reconnaissance de forme et d'apprentissage à partir d'exemples qui font défaut à l'IA symbolique.

Le rôle du langage dans l'intelligence

171

Mais les deux types d'IA ont aujourd'hui en commun un certain cloisonnement logique, des difficultés d'accumulation, d'intégration et de recombinaison des connaissances. Réunir les deux branches de l'IA ne nous aidera donc pas magiquement à rendre accessibles aux machines les connaissances humaines accumulées depuis des siècles, ni à faciliter les transferts de savoir d'un système à l'autre. Pourtant, grâce au langage, les sociétés humaines sont capables de transmettre des traditions d'une génération à l'autre. Au sein d'une même génération, le langage naturel permet de transformer des perceptions muettes et des savoir-faire issus de l'expérience en connaissances partageables. À force de dialogue, un spécialiste d'un domaine finit par se faire comprendre d'un spécialiste d'un autre domaine et va peut-être même lui enseigner quelque chose. Comment reproduire ce type de transferts cognitifs dans des collectifs mixtes de machines et d'humains ? Qu'est-ce qui pourrait jouer le rôle intégrateur du *langage naturel* dans des sociétés d'agents logiciels ? C'est ici qu'IEML entre en jeu.

172

Bien des gens pensent que, le cerveau étant le support organique de l'intelligence, les modèles neuronaux sont l'unique clé de sa simulation. Mais de quelle intelligence parle-t-on ? N'oublions pas que tous les animaux ont un cerveau, or ce n'est pas l'intelligence du moucheron ou de la baleine que l'IA veut simuler ou augmenter, mais celle de l'humain. Et si nous sommes « plus intelligents » que les autres animaux (au moins de notre point de vue), ce n'est pas à cause de la taille de notre cerveau. L'éléphant possède

un plus gros cerveau que l'homme en termes absolus et le rapport entre la taille du cerveau et celle du corps est plus grand chez la souris que chez l'humain. C'est principalement notre *capacité linguistique*, soutenue notamment par un grand cortex préfrontal ainsi que par les aires de Broca, Wernicke et quelques autres – uniques à l'espèce humaine –, qui distingue notre intelligence de celle des autres vertébrés supérieurs. Or ces modules de traitement du langage ne sont pas fonctionnellement séparés du reste du cerveau, ils informent au contraire l'ensemble de nos processus cognitifs et notre mémoire est largement organisée par un système de coordonnées sémantiques fourni par le langage.

173

Pourquoi l'IA n'utilise-t-elle pas les langues naturelles pour représenter les connaissances ? La réponse est bien connue : parce que les langues naturelles sont ambiguës. Un mot peut avoir plusieurs sens, un sens peut s'exprimer par plusieurs mots, les phrases ont plusieurs interprétations possibles, la grammaire est élastique et changeante. Les ordinateurs ne sont pas capables de désambiguïser correctement les énoncés en langue naturelle. Pour ses locuteurs humains, une langue naturelle étend un filet de catégories générales prédéfinies qui s'expliquent mutuellement et ce réseau sémantique commun permet de décrire et de faire communiquer aussi bien les multiples situations concrètes que les différents domaines de connaissance. Mais, du fait des limitations des machines, l'IA ne peut pas faire jouer ce rôle à une langue naturelle. C'est pourquoi elle reste aujourd'hui fragmentée en microdomaines de pratiques et de connaissance, chacun avec sa sémantique particulière.

Sémantique des noms communs et sémantique des noms propres

174

Pourtant, l'IA symbolique n'est-elle pas censée simuler les capacités humaines d'expression et de raisonnement, y compris la faculté linguistique ? Malgré son nom de « symbolique », avant l'invention d'IEMML, l'IA ne disposait pas d'un modèle univoque et calculable du langage. Comme nous l'avons vu plus haut, depuis les travaux de Chomsky, nous savions certes calculer la dimension *syntactique* des langues, mais leur dimension *sémantique* restait hors de portée de l'informatique. Ce point apparaît clairement dans la manière dont le « Web sémantique » du W3C traite le sens des mots. Du point de vue de l'étude scientifique du langage, la sémantique du mot se décompose en deux parties : la sémantique des noms communs et celle des noms propres. Or nous allons voir que l'IA symbolique classique du W3C n'a accès qu'à la sémantique des noms propres.

175

La sémantique des noms communs ou sémantique mot-mot. Rappelons qu'un symbole linguistique (mot ou phrase) possède deux faces : le *signifiant*, qui est une image visuelle ou acoustique, et le *signifié*, qui est un concept ou une catégorie générale. Par exemple, le signifiant « arbre » a pour signifié : « végétal ligneux, de taille variable, dont le tronc se garnit de branches à partir d'une certaine hauteur ». La relation entre signifiant et signifié étant fixée par la langue, le signifié d'un mot ou d'une phrase se définit comme un nœud de relations avec d'autres signifiés. Les verbes et les noms communs (par exemple, « arbre », « animal », « organe » ou « manger ») représentent des catégories qui sont elles-mêmes connectées par un dense réseau de relations sémantiques : « est une partie de », « est un genre de », « appartient au même contexte que », « est la cause de », « est antérieur à », etc. Nous ne pouvons penser et communiquer à la manière

humaine que parce que nos mémoires collectives et personnelles sont organisées par des catégories générales connectées par des relations sémantiques.

176

La sémantique des noms propres ou sémantique mot-chose. Par contraste avec la sémantique linguistique, la sémantique référentielle connecte un symbole linguistique (signifiant et signifié) avec un référent (un individu réel). Lorsque je dis que « les platanes sont des arbres », je précise le sens conventionnel du mot « platane » en le mettant en relation d'espèce à genre avec le mot « arbre » et je ne mets donc en jeu que la sémantique linguistique. Mais si je dis « Cet arbre-là, dans la cour, est un platane », alors je pointe vers un état de choses réel et ma proposition est vraie ou fausse. Ce second énoncé met évidemment en jeu la sémantique des noms communs puisque je dois d'abord connaître le sens des mots (et la grammaire du français) pour le comprendre. Mais s'ajoute à la dimension intralinguistique une sémantique extralinguistique ou référentielle, puisque l'énoncé se rapporte à un objet particulier dans une situation concrète. C'est ainsi que certains signifiants, comme les noms propres, n'ont pas de signifiés. Leur signifiant renvoie directement à un référent, il ne pointe pas vers une catégorie générale, mais vers un objet singulier. Par exemple, le signifiant « Alexandre le Grand » désigne un personnage historique et le signifiant « Tokyo » désigne une ville (Kripke 1980 ; Récanati 1983 ; Frege 1971).

177

Dans un dictionnaire classique, chaque mot est situé par rapport à d'autres mots proches (le thésaurus) et il est expliqué par des phrases (la définition) utilisant des mots eux-mêmes expliqués par d'autres phrases, et ainsi de suite de manière circulaire. Un dictionnaire classique relève principalement de la *sémantique des noms communs*. Par contraste avec un dictionnaire ordinaire, qui définit des concepts ou des catégories, un dictionnaire encyclopédique contient des descriptions d'*individus* réels ou fictifs pourvus de noms propres tels que divinités, héros de roman, événements historiques et objets géographiques. Sa principale fonction est de répertorier et de décrire des objets externes au système d'une langue. Il enregistre donc une *sémantique des noms propres*.

178

La distinction simple que nous venons de faire nous suffit à évaluer dans quelle mesure l'IA symbolique pré-IEMML maîtrisait la sémantique. Si l'on considère les ontologies du Web sémantique (le standard en IA symbolique), on découvre que les sens des mots et des phrases n'y dépendent pas de la circularité autoexplicative d'une langue (comme dans un dictionnaire classique), mais d'un renvoi à des URI (*Uniform Resource Identifiers*) qui fonctionnent sur le mode de la *sémantique des noms propres* (comme un dictionnaire encyclopédique). Au lieu de reposer sur des concepts déjà donnés dans une langue et qui se présentent dès l'origine comme des nœuds de relations avec d'autres concepts, les échafaudages du Web sémantique s'appuient sur des concepts définis séparément les uns des autres au moyen d'identifiants uniques (Berners-Lee 2013). La circulation du sens dans un réseau de signifiés est escamotée au profit d'une relation directe entre signifiant et référent, comme si tous les concepts étaient représentés par des noms propres. En l'absence d'une sémantique des noms communs fondée sur la grammaire et le dictionnaire d'une langue, les ontologies restent donc cloisonnées. En somme, avant IEMML, l'IA symbo-

lique n'avait pas accès à la pleine puissance cognitive et communicative du langage, mais seulement à une sémantique référentielle rigide combinée à la logique.

Ce qu'IEML amène à l'intelligence artificielle

179 Une langue mathématique, standardisée et univoque comme IEML pourrait ouvrir à l'intelligence artificielle de nouveaux horizons d'apprentissage, de raisonnement et de communication. Du côté de l'IA neuronale, aujourd'hui en plein essor, le nouveau sous-domaine de recherche du *semantic machine learning* montre que l'étiquetage ou l'annotation au moyen de bonnes ontologies favorise la généralisation en apprentissage machine (voir par exemple Kulmanov *et al.* [2021]). *A fortiori*, l'usage d'ontologies en IEML pourrait soutenir un apprentissage plus efficace, car les unités ou « *tokens* » pris en compte par les machines ne seraient plus des unités de son – caractères, syllabes ou mots – en langues naturelles mais des *concepts* générés par une algèbre sémantique. On atteindrait par ce moyen une IA neuronale plus transparente et mieux explicable. Du côté de l'IA symbolique, l'usage d'IEML amènerait un nouveau calcul sémantique (la création automatique de nœuds et de liens conceptuels) et une véritable sémantique linguistique *en plus* des fonctions logiques et de la sémantique référentielle (les URI) déjà en usage. Cette langue faciliterait également l'intégration neurosymbolique et permettrait aux modèles de se connecter et aux connaissances de s'accumuler.

180 Posons que, pour une machine, « comprendre » signifie projeter des données sur des graphes conceptuels. Dès lors, moyennant codage en IEML des flux d'information, il est désormais possible de concevoir des IA génératives capables de comprendre (1) leurs données d'entraînement, (2) les *prompts* ou commandes des utilisateurs, (3) les textes, codes, images et les autres données qu'elles produisent en réponse. On passerait ainsi d'une approximation probabiliste des résultats attendus à une véritable génération contrôlée.

181 Vu le rôle de la parole dans l'intelligence humaine, il est surprenant qu'on ait espéré atteindre une intelligence artificielle générale sans disposer d'un modèle calculable du langage et de sa sémantique²⁶. La bonne nouvelle est que nous en avons finalement un. L'intelligence artificielle ne pourra durablement augmenter la cognition humaine qu'à la condition d'être sensée, interopérable, cumulable, intégrable, échangeable et distribuée. C'est dire qu'on ne fera pas de progrès notable en intelligence artificielle sans viser en même temps une intelligence collective capable de se réfléchir et de se coordonner dans la mémoire mondiale.

Les humanités numériques

182 Les sciences humaines sont aujourd'hui balkanisées en disciplines et divisées en théories qui souvent ne s'entendent même pas sur leurs désaccords. Nous sommes ici fort loin du découpage des disciplines selon des lignes objectives de niveaux de complexité (physique, chimie, biologie...) qui prévaut dans les sciences de la nature et de l'accord général des sciences exactes sur leurs méthodes et leurs catégories organisatrices. Malgré quelques exceptions, le degré de formalisation des modèles et de testabilité des théories des sciences humaines reste faible. Cela rappelle la situation qui prévalait dans l'alchimie avant l'émergence de la chimie moderne à la fin du XVIII^e siècle. Mais les conditions sont mûres pour une révolution scientifique (Kuhn 1962).

183 Les archives sont progressivement numérisées et la quasi-totalité des activités humaines produit un flot montant de données. Pour traiter ces données, nous disposons d'algorithmes d'analyse et de visualisation efficaces ainsi que d'une puissance de calcul gigantesque. Enfin, les modes de communication du XXI^e siècle ouvrent aux équipes de chercheurs de nouvelles avenues de dialogue, de partage de corpus et d'outils, de collaboration internationale et de publication. Les chercheurs en sciences humaines constituent des bases de données pour l'analyse, la fouille et le partage. Les résultats de la recherche sont de plus en plus disponibles en ligne sur les blogs des chercheurs, certains réseaux sociaux, les sites Web des revues, mais aussi dans des moteurs de recherche spécialisés comme Isidore²⁷.

184 Mais les nouvelles conditions de la recherche en sciences humaines posent de façon cruciale le problème d'une catégorisation partageable et interopérable des données et des documents. Nous l'avons vu plus haut, il existe aujourd'hui de nombreuses ontologies et systèmes de métadonnées sémantiques, différents selon les langues, disciplines, traditions et théories. Ces systèmes sont d'ailleurs souvent hérités de l'ère de l'imprimerie. Dans ce contexte, IEML propose un outil de modélisation et d'indexation programmable, capable d'assurer l'interopérabilité sémantique sans uniformiser les points de vue. Il devient possible d'échanger collaborativement des modèles et sous-modèles entre chercheurs parlant des langues différentes et venant de disciplines distinctes.

185 De même que les corps matériels baignent dans l'espace et le temps, l'esprit humain vit dans l'élément intangible de la signification et de la valeur. La compréhension unifiée de l'univers physique grâce à la modélisation mathématique et à de nouveaux outils d'observation et de mesure ne date que de quelques siècles (Koyré 1972). Elle a d'ailleurs suivi de peu l'usage massif de l'imprimerie à caractères mobiles. L'adoption de systèmes de coordonnées mathématiques et d'unités de mesure communes pour l'espace et le temps a joué un rôle capital dans les avancées scientifiques et techniques des trois derniers siècles, comme pour le perfectionnement des possibilités d'action concertées de nos sociétés. L'émergence du médium algorithmique pourrait avoir sur les sciences humaines un effet semblable à celui de l'imprimerie sur les sciences de la nature, à condition que nous reconnaissons leur unité foncière et que nous adoptions un système de coordonnées sémantiques soutenant leur mathématisation. Précisons que la mathématisation n'implique pas nécessairement la réduction au quantitatif. De nombreux domaines de l'algèbre et de la

géométrie, par exemple, formalisent des structures et des symétries plutôt que des quantités. De même que l'idée d'une physique mathématique était impensable avant la fin du xv^e siècle, l'idée d'une sémantique mathématique peut paraître invraisemblable aujourd'hui. IEML ouvre pourtant la voie à un nouveau cosmos scientifique, accueillant à la signification, cohérent et computable, quoique inépuisablement complexe. Dès lors, une boucle de rétroaction plus fluide entre les écosystèmes d'idées et les communautés qui les entretiennent nous rapprocherait de l'idéal d'une intelligence collective réflexive au service du développement humain. Il ne s'agit pas de nourrir quelque illusion sur la possibilité d'une transparence totale et encore moins de prétendre parvenir à une seule théorie vraie, mais plutôt d'ouvrir la voie à l'exploration critique d'un univers de sens infini.

Conclusion

186

Il existe désormais au moins une langue qui se prête au calcul de sa sémantique. Les fondations scientifiques sont posées pour l'établissement d'un protocole apte à résoudre le problème de l'interopérabilité sémantique, avec toutes les conséquences évoquées plus haut sur le perfectionnement de l'intelligence collective, de l'intelligence artificielle et des sciences humaines. Un tel système de métadonnées sémantiques a pour fonction d'augmenter les traitements automatiques au service d'une intelligence collective plurielle, mais interdépendante. Le chemin qui reste à parcourir pour atteindre une transcroissance effective de la mémoire numérique est immense. Souhaitons qu'un réseau multidisciplinaire de chercheurs, de linguistes et de développeurs se mette au travail et qu'un consortium d'universités, de gouvernements et d'entreprises soutienne leurs efforts.

187

Indépendamment du développement d'une nouvelle génération de modèles d'IA neurosymboliques qui « comprennent » ce qu'on leur dit et ce qu'elles font, ou de la perspective à long terme de son adoption comme protocole, IEML ouvre dès maintenant la voie à une nouvelle *ingénierie sémantique* qui soutient plusieurs directions de recherche. L'une d'elles concerne le *design* de systèmes d'information dans lesquels la conceptualisation rigoureuse de modèles de données au moyen de ce langage mathématique permettrait le calcul d'un grand nombre de vérités analytiques (découlant de la construction même des concepts) et synthétiques *a priori* (découlant des relations entre concepts), avant même le chargement des données²⁸. Cette couche cognitive préalable à l'enregistrement des données pourrait augmenter de manière considérable les raisonnements symboliques et les apprentissages statistiques menés sur les données empiriques. Une direction de recherche complémentaire porte sur le développement d'une *méthode* de production des ontologies ou graphes de connaissances en IEML, propre à accompagner le mieux possible les communautés utilisatrices dans l'atteinte de leurs finalités et l'augmentation de leurs pratiques. Comment cartographier un univers de discours en évitant les angles morts ? Comment reconceptualiser les données (et donc faire évoluer les ontologies) de manière souple au cours du cycle normal d'utilisation d'un système d'information ? Rendre observables les processus d'intelligence collective qui s'appuient sur un système d'information commun nous ferait progresser vers une intelligence collective plus réflexive.

Bibliographie

- Austin, John, Langshaw. 1962. *How to Do Things with Words*. Oxford : Oxford University Press.
- Bakhtine, Mikhaïl. 1978. *Esthétique et théorie du roman*. Paris : Gallimard.
- Bakhtine, Mikhaïl. 1981. *The Dialogic Imagination. Four Essays*. Austin : University of Texas Press.
- Baltzersen, Rolf. 2022. *Cultural-Historical Perspectives on Collective Intelligence. Patterns in Problem Solving and Innovation*. Cambridge : Cambridge University Press.
- Benveniste, Émile. 1966. *Problèmes de linguistique générale*. Vol. 1. Paris : Gallimard.
- Benveniste, Émile. 1974. *Problèmes de linguistique générale*. Vol. 2. Paris : Gallimard.
- Berners-Lee, Tim. 2013. « Philosophical Engineering and Ownership of URIs ». W3C – Design Issues. <https://www.w3.org/DesignIssues/PhilosophicalEngineering.html>.
- Bickerton, Derek. 1995. *Language and Human Behavior*. Seattle : University of Washington Press.
- Bonabeau, Eric. 1999. « Special Issue on Stigmergy ». *Artificial Life* 5 (2). <https://direct.mit.edu/artl/issue/5/2>.
- Bush, Vannevar. 1945. « As We May Think ». *The Atlantic Monthly*. <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>.
- Chomsky, Noam. 1957. *Syntactic Structures*. La Hague et Paris : Mouton.
- Chomsky, Noam et Marcel P. Schützenberger. 1963. « The Algebraic Theory of Context-Free Languages ». Dans *Computer Programming and Formal Systems*, édité par Paul Braffort et David Hirschberg, 118-161. Amsterdam : North Holland.
- Connes, Alain. 1990. *Géométrie non commutative*. Paris : InterÉditions.
- D'Avila Garcez, Artur et Luis Lamb. 2020. « Neurosymbolic AI : The 3rd Wave ». arXiv. <http://arxiv.org/pdf/2012.05876.pdf>.
- D'Avila Garcez, Artur, Luis Lamb et Dov Gabbay. 2009. *Neural-Symbolic Cognitive Reasoning*. Berlin : Springer.
- Deacon, Terrence William. 1997. *The Symbolic Species : The Co-Evolution of Language and the Brain*. New York : W.W. Norton.
- Deleuze, Gilles et Félix Guattari. 1980. *Mille Plateaux*. Paris : Les Éditions de minuit.
- Dorigo, Marco, Éric Bonabeau et Guy Théraulaz. 2000. « Ant Algorithms and Stigmergy ». *Future Generation Computer Systems* 16 (8) : 851-871. [https://doi.org/10.1016/S0167-739X\(00\)00042-X](https://doi.org/10.1016/S0167-739X(00)00042-X).
- Engelbart, Douglas. 1962. « Augmenting Human Intellect ». Stanford : Stanford Research Institute. <https://www.dougelbart.org/pubs/augment-3906.html>.
- Fillmore, Charles. 1968. « The Case for Case ». Dans *Universals in Linguistic Theory*, édité par Emmon Werner Bach et Robert Thomas Harms, 1-88. New York : Holt, Rinehart and Winston.
- Fillmore, Charles. 1982. « Frame Semantics ». Dans *Linguistics in the Morning Calm*, édité par The Linguistic Society of Korea, 111-137. Séoul : Hanshin.
- Frege, Gottlob. 1971. *Écrits logiques et philosophiques*. Traduit par Claude Imbert. Paris : Le Seuil.
- Galison, Peter. 2003. *Einstein's Clocks, Poincaré's Maps*. New York : Norton.
- Greimas, Algirdas Julien. 1970. *Du sens. Essais sémiotiques*. Paris : Le Seuil.
- Hejlslev, Louis. 2000 [1968]. *Prolégomènes à une théorie du langage – La Structure fondamentale du langage*. Paris : Les Éditions de minuit.
- Husserl, Edmund. 1950. *Idées directrices pour une phénoménologie*. Paris : Gallimard.
- Jakobson, Roman. 1963. *Essais de linguistique générale*. Vol. 1 et 2. Paris : Les Éditions de minuit.
- Kant, Emmanuel. 1976 [1787]. *Critique de la raison pure*. Paris : Garnier-Flammarion.

Kleene, Stephen Cole. 1956. « Representation of Events in Nerve Nets and Finite Automata ». Dans *Automata Studies*, édité par Claude Elwood Shannon et John McCarthy, 3-41. Princeton : Princeton University Press.

Koyré, Alexandre. 1972. *Du monde clos à l'univers infini*. Paris : Gallimard.

Kripke, Saul. 1980. *Naming and Necessity*. Oxford : Blackwell.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago : University of Chicago Press.

Kulmanov, Maxat, Fatima Zohra Smaili, Xin Gao et Robert Hoehdorst. 2021. « Semantic Similarity and Machine Learning with Ontologies ». *Briefings in Bioinformatics* 22 (4). <https://doi.org/10.1093/bib/bbaa199>.

Lakoff, George. 1987. *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*. Chicago : University of Chicago Press.

Langacker, Ronald. 1987. *Foundations of Cognitive Grammar*. Stanford : Stanford University Press.

Leibniz, Gottfried Wilhelm. 1989. « On the General Characteristic ». Dans *Philosophical Letters and Papers*, édité par Leroy E. Loemker. Dordrecht : Springer. https://doi.org/10.1007/978-94-010-1426-7_25.

Lévy, Pierre. 1994. *L'Intelligence collective. Pour une anthropologie du cyberspace*. Paris : La Découverte.

Lévy, Pierre. 1995. *Qu'est-ce que le virtuel ?* Paris : La Découverte.

Lévy, Pierre. 2010. « From Social Computing to Reflexive Collective Intelligence : the IEML Research Program ». *Information Sciences* 180 (1) : 71-94. <https://doi.org/10.1016/j.ins.2009.08.001>.

Lévy, Pierre. 2011. *La Sphère sémantique. Computation, cognition, économie de l'information*. Paris : Hermès-Lavoisier.

Lévy, Pierre. 2021. « Pour un changement de paradigme en intelligence artificielle ». *Giornale di Filosofia* 2 (2). <https://mimesisjournals.com/ojs/index.php/giornale-filosofia/articolo/view/1693>.

Licklider, Joseph. 1960. « Man-Computer Symbiosis ». *IRE Transactions on Human Factors in Electronics* 1 : 4-11. <https://doi.org/10.1109/THFE2.1960.4503259>.

Marcus, Gary et Ernest Davis. 2019. *Rebooting AI. Building Artificial Intelligence We Can Trust*. New York : Pantheon.

Matta, Nada et Nathalie Aussenac-Gilles. 1996. « Le schéma du modèle conceptuel, étape dans la modélisation des connaissances ». Dans *Acquisition et ingénierie des connaissances. Tendances actuelles*, édité par Nathalie Aussenac-Gilles, Philippe Laublet et Chantal Reynaud, 29-48. Toulouse : Éditions Cépaduès. <https://www.irit.fr/publis/IC3/livjacAusMatta96.pdf>.

McLuhan, Marshall. 2005. *The Classical Trivium. The Place of Thomas Nashe in the Learning of His Time*. Berkeley : Gingko Press.

Melchuk, Igor. 1982. « Lexical Functions in Lexicographic Description ». Dans *Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society*, édité par Monica Maccaulay et Orin D. Gensler, 427-444. Berkeley : Berkeley Linguistics Society.

Melchuk, Igor. 1998. « Collocations and Lexical Functions ». Dans *Phraseology. Theory, Analysis, and Applications*, édité par Anthony Paul Cowie, 23-53. Oxford : Clarendon Press.

Montague, Richard. 1970. « Universal Grammar ». *Theoria* 36 : 373-98. <https://doi.org/10.1111/j.1755-2567.1970.tb00434.x>.

Mulgan, Geoff. 2017. *Big Mind. How Collective Intelligence Can Change Our World*. Princeton : Princeton University Press.

Polguère, Alain. 2003. « Collocations et fonctions lexicales : pour un modèle d'apprentissage ». Dans *Les Collocations. Analyse et traitement*, édité par Francis Grossmann et Agnès Tutin, 117-133. Amsterdam : De Werelt.

Rastier, François. 1987. *Sémantique interprétative*. Paris : Presses universitaires de France.

Récanati, François. 1983. « La sémantique des noms propres : remarques sur la notion de "désignateur rigide" ». *Langue française* 57 : 106-118. <https://doi.org/10.3406/lfr.1983.5159>.

Riley, Jenn. 2018. « Seeing Standards : a Visualization of the Metadata Universe ». Borealis. <https://doi.org/10.5683/SP2/UOHPVH>.

- Saussure, Ferdinand (de). 1916. *Cours de linguistique générale*. Paris : Payot.
- Searle, John. 1969. *Speech Acts*. Cambridge : Cambridge University Press.
- Searle, John. 1983. *Intentionality*. Cambridge : Cambridge University Press.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris : Klincksieck.
- Tremblay, Ophélie. 2014. « Les collocations : des mots qui font la paire ». *Québec français* 171 : 74-76. <https://id.erudit.org/iderudit/71229ac>.
- Vygotsky, Lev. 1962 [1934]. *Thought and Language*. Cambridge : MIT Press.
- Wierzbicka, Anna. 1996. *Semantics. Primes and Universals*. Oxford : Oxford University Press.
- Wittgenstein, Ludwig. 2014 [1953]. *Recherches philosophiques*. Paris : Gallimard.

Notes

- 1 Dans sa thèse de 1943, McLuhan (2005) défend l'idée que la tradition européenne à partir du Moyen Âge s'est trop concentrée sur la dialectique (c'est-à-dire la logique) au détriment de la grammaire (c'est-à-dire la linguistique et la littérature). Dans le prolongement de la pensée de McLuhan, je pense aussi qu'un rééquilibrage des intérêts scientifiques au profit de la « grammaire » bénéficierait à la compréhension de la sémantique.
- 2 « A category is formed by two sorts of objects, the objects of the category, and the morphisms, which relate two objects called the source and the target of the morphism. » (« Category Theory », *Wikipedia*, 7 novembre 2022, https://en.wikipedia.org/w/index.php?title=Category_theory&oldid=1120594744.) Je paraphrase : la théorie des catégories met l'accent sur les morphismes et les processus qui préservent la structure entre deux objets. (« Théorie des catégories », *Wikipédia*, 6 septembre 2022, https://fr.wikipedia.org/w/index.php?title=Th%C3%A9orie_des_cat%C3%A9gories&oldid=196746532.) Dans notre cas, un morphisme bidirectionnel relie le système des séquences de phonèmes (objet A) à celui des réseaux de concepts (objet B).
- 3 La notion de paradigme en linguistique a d'abord été développée par Saussure (1916). L'idée d'une intersection des deux axes a été proposée par Jakobson (1963).
- 4 Certaines langues ont aussi des consonnes « à clic » (en Afrique orientale et australe) et d'autres utilisent différents tons sur leurs voyelles (comme le chinois mandarin).
- 5 On confond souvent le projet leibnizien de logique mathématique avec le projet complémentaire – mais distinct – de caractéristique universelle. Leibniz se contenta d'esquisser les fonctions de cette langue idéale dont, comme je l'ai dit au début de cet article, IEML peut être considéré comme une réalisation.
- 6 Pour consulter le dictionnaire, aller à <https://ieml.intlekt.io/login/> et sélectionner « read without account », puis « published projects ». On peut choisir entre les traductions en français et les traductions en anglais à partir de l'interface de requête. En cliquant sur un mot ou l'une de ses traductions, on accède à un menu contextuel dans lequel on peut choisir de visualiser le paradigme (la table) auquel le mot appartient. Voir aussi : <https://intlekt.io/change-time/>.
- 7 Pour plus de détails sur le contexte philosophique et historique des six primitives, voir <https://intlekt.io/semantic-primitives/>. Pour comprendre les têtes de tables, de rangées et de colonnes, il faut savoir que quatre autres lettres majuscules peuvent être utilisées pour représenter certaines additions remarquables, notamment dans les têtes de tables, de rangées et de colonnes qu'on verra dans le dictionnaire à l'adresse <https://ieml.intlekt.io> : « O : = U : ⊕ A : », « M : = S : ⊕ B : ⊕ T : », « F : = O : ⊕ M : », « I : = F : ⊕ E : ».
- 8 Voir la démonstration sur une version d'IEML de 2015 : <https://hal.science/hal-04318801> en particulier le chapitre 5. Signalons les travaux d'Alain Connes (1990) sur l'algèbre non commutative.
- 9 « n.i.- », le feu-outil, se distingue en IEML du feu-élément, « n.i.-S.:A.-' ».
- 10 https://fr.wikipedia.org/w/index.php?title=G%C3%A9om%C3%A9trie_non_commutative&oldid=204130048.
- 11 En utilisant le mot « joncté », je suis ici la pratique de Lucien Tesnière (1959).
- 12 Je prévois de publier un article portant sur une méthode pratique de construction d'ontologies en IEML, qui s'appuiera notamment sur cet exemple de la santé mentale.

13 À noter, dans la rangée « subjectivation », que le sujet se rapporte à lui-même dans la première case, à l’alter ego dans la seconde, et à la réalité dans la troisième. La « pensée » représente la manière qu’a le sujet de se rapporter à la réalité, d’où son affinité avec la chose. Par ailleurs, dans la troisième case de la rangée « émotion », l’humeur est l’émotion sentie et l’affect est l’émotion exprimée.

14 <https://intlekt.io/quantities/>.

15 <https://intlekt.io/geography/>.

16 https://en.wikipedia.org/w/index.php?title=Uniform_Resource_Identifier&oldid=1181931955.

17 Les données personnelles seraient sécurisées dans des pods selon les normes du projet *Solid*. Voir [https://en.wikipedia.org/wiki/Solid_\(web_decentralization_project\)](https://en.wikipedia.org/wiki/Solid_(web_decentralization_project)).

18 « Semantic Web – W3C » : <https://www.w3.org/TR/rdf12-concepts/>.

19 <https://schema.org>.

20 <https://cidoc-crm.org>.

21 Voir Riley (2018) et <http://jennriley.com/metadatamap/>.

22 <https://www.poolparty.biz>.

23 <https://www.abinitio.com/fr/>.

24 <https://www.synaptica.com>.

25 Je reprends dans cette section certaines idées et une partie du texte de Lévy (2021).

26 Pour une critique constructive de l’IA contemporaine et en particulier du « tout statistique », voir Marcus et Davis (2019).

27 <https://isidore.science>.

28 Les notions de proposition analytique et de proposition synthétique *a priori* ont été particulièrement illustrées par Emmanuel Kant (1976 [1787]) au début de son *magnum opus*.

Auteur

Pierre Lévy

Membre de la Société royale du Canada et chercheur associé à l’Université de Montréal (Chaire de recherche du Canada sur les écritures numériques et Centre de recherche interuniversitaire sur les humanités numériques). Montréal, Canada.

ORCID : 0000-0003-1599-221X

pierre.levy@mac.com

Droits d’auteur



Le texte seul est utilisable sous licence [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/). Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.