

BULLETIN D'INFORMATIQUE APPROFONDIE ET APPLICATIONS

COMPUTATION - INFORMATION

N° 80 – JUIN 2008

COMITE SCIENTIFIQUE ET DE REDACTION

Patrick Abellard
Françoise Adreit
France Chappaz
M'hamed Charifi
Roger Cusin
Jean - Claude Fumanal
Alain de Gantès
Bernard Goossens
Sami Hilala
Patrick Isoardi
Robert Jacquier
Jean - Michel Knippel
Jean - Philippe Lehmann
Agathe Merceron
Nadia Mesli
Patrick Sanchez
Rolland Stutzmann
André Tricot

CORRESPONDANTS

Afrique
Mohamed Tayeb Laskri

Amériques
Sylvie Monjal

Asie
Moussa HadjAli

Europe
José Rouillard

Océanie
Kalina Yacef

1 EDITORIAL
Lettre ouverte aux responsables de l'évaluation scientifique

par des directeurs de collections d'Hermès Science

3 Informatique et linguistique
Extraits du cours de Jean Véronis

par Jean – Michel Knippel

5 Informatique et linguistique
Introduction, historique

par Jean Véronis

45 VOZZAVEDIBISAR
L'électeur Gaullicoque

par Edmond Bianco

<http://www.univ-provence.fr/biaa>

Publication trimestrielle, gratuite, de l'Université de Provence

BULLETIN D'INFORMATIQUE APPROFONDIE ET APPLICATIONS

COMPUTATION - INFORMATION

N° 80 – JUIN 2008

DIRECTEUR

Jean - Michel Knippel

FONDATEUR

Edmond Bianco

SERVEUR DE PUBLICATION

Christian Blanvillain

SECRETARIAT

Kalassoumi Adjilani

Université de Provence
Equipe Hermès. Case 33
3 place Victor Hugo
F - 13331 Marseille Cedex 3
Téléphone: (0)4 91 10 62 30
Télécopie : (0)4 91 50 91 10

DEPOSITAIRE

Université de Provence
Bibliothèque Universitaire
1 place Victor Hugo
F - 13331 Marseille Cedex 3
Téléphone : (0)4 91 10 85 29
Télécopie : (0)4 91 95 75 57

IMPRIMEUR

Université de Provence
Service Reprographie
3 place Victor Hugo
F - 13331 Marseille Cedex 3
Téléphone : (0)4 91 10 60 48

1 EDITORIAL
Lettre ouverte aux responsables de l'évaluation
scientifique

par des directeurs de collections d'Hermès Science

3 Informatique et linguistique
Extraits du cours de Jean Véronis

par Jean - Michel Knippel

5 Informatique et linguistique
Introduction, historique

par Jean Véronis

45 VOZZAVEDIBISAR
L'électeur Gaullicoque

par Edmond Bianco

<http://www.univ-provence.fr/biaa>

Publication trimestrielle, gratuite, de l'Université de Provence

EDITORIAL

Lettre ouverte aux responsables de l'évaluation scientifique

Les scientifiques doivent-ils continuer à écrire en français ?

des directeurs de collection d'Hermès Science

Il est largement admis que la *lingua franca* de la recherche scientifique est aujourd'hui l'anglais. Pourtant, il existe au moins trois bonnes raisons de penser qu'il est indispensable que les scientifiques continuent d'écrire en français.

– Puisque la recherche repose essentiellement sur des financements publics, une considération élémentaire voudrait que les contribuables aient un accès en français à ce qu'ils ont soutenu par le biais de leurs impôts.

– La deuxième raison concerne l'enseignement. La production de livres de synthèse et de manuels en français est une tâche extrêmement honorable et même nécessaire pour compléter un enseignement dispensé en français. Comment faire aimer une discipline en n'offrant que des livres en anglais qui ne sont en général pas adaptés, ni au niveau, ni aux habitudes que nous avons de structurer nos enseignements ?

– La troisième raison relève de l'apprentissage. Il faut un grand entraînement pour pouvoir s'exprimer dans une autre langue que sa langue maternelle avec le même sens de la nuance, avec la même richesse. Quel meilleur moyen d'accéder à la pensée d'un auteur que de discuter avec lui dans sa propre langue ?

La publication en français apparaît donc comme une nécessité. Pour que cette production continue, il est urgent de valoriser notre activité de recherche dans notre langue. En effet, les systèmes de référencement des publications (dont le principal est une filiale d'un éditeur privé) reconnaissent prioritairement les publications en anglais ! Soit notre système national valorise cette production, soit cette dernière disparaîtra.

Reconnaître à leur juste valeur les publications en français, suppose que les sections du comité national du CNRS, du CNU et de l'AERES prend en compte, en fonction de leurs exigences de qualité, les revues en français mais aussi les livres et les manuels. Il n'est pas normal qu'à l'heure actuelle, un bon livre écrit en français ne soit pas considéré dans l'évaluation d'un chercheur.

L'objectif de cette pétition adressée à L'Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur (AERES) est de montrer que notre communauté scientifique française ou francophone a encore la capacité de penser par elle-même et qu'il ne faut pas rejeter, comme insignifiant, tout ce qui s'écrit en français.

Nous vous remercions pour votre soutien.

Les deux premiers signataires de la lettre :

***Bernard DUBUISSON,
Président du Conseil scientifique d'Hermès Science,
Université de Technologie de Compiègne, 60200 Compiègne,
France***

***Jean-Charles POMEROL,
Rédacteur en chef de RIA et RSTI,
Vice-président du conseil scientifique Hermès Science,
Université P. et M. Curie, 75252 Paris,
France***

Informatique et linguistique
Extraits du cours de Jean Véronis

Jean - Michel Knippel

Jean Véronis est Professeur de linguistique et informatique à l'université de Provence. Il a déjà écrit dans notre revue au siècle dernier avec Nancy Ide un article consacré à la mise au point d'un ensemble de normes pour la préparation et l'échange de textes électroniques (*).

Ici, il nous propose d'éditer les trois premières parties de son unité d'enseignement consacrée à l'histoire de l'informatique et de la linguistique « du rêve ... à la réalité ». La suite se trouve sur son site :
<http://sites.univ-provence.fr/veronis/cours/index.html> .

Je le remercie d'avoir accepté cette nouvelle contribution. Ces cours resituent le terreau historique de certains articles parus dans notre périodique, que nous soyons dans « le rêve » ou « la réalité ». Citons quelques-uns d'entre eux.

« Du rêve... » :

Traitement informatique de la notation du jeu instrumental pour instrument de musique mécanique

Robert Hopp et Patrick Sanchez

B.I.A.A. N°43 de mars 1996

Hamlet – Machine. Ensemble N°22'

Clyde Chabot et Jean – Michel Knippel

B.I.A.A. N°60 de décembre 2001

«... A la réalité » :

Une preuve directe de l'équivalence : Machine de Turing, machines à cases adressables

Jean – Philippe Lehmann

B.I.A.A. N°7 de mars 1984

La littérature animée programmée : du mixage des sens à une réévaluation de la lecture

Philippe Bootz

B.I.A.A. N°62 de juin 2002

(*) Encodage des dictionnaires électroniques :

problèmes et propositions de la TEI

Nancy Ide et Jean Véronis

B.I.A.A. N°42 de décembre 1995

Informatique et linguistique

Cours 1

Introduction

Jean Véronis

1. Thème du cours

Les machines peuvent-elles dialoguer avec les hommes?

- HAL. L'ordinateur HAL¹ du célèbre film de Stanley Kubrick 2001, *L'Odyssée de L'Espace*² (Figure 1) a popularisé l'image de l'ordinateur doué de la capacité de langage : il est capable de suivre sans faute les conversations des humains et de dialoguer avec eux. On le voit même lire sur les lèvres lorsque deux des protagonistes — qui ont découvert les plans machiavéliques de HAL — ont cru s'isoler hors de portée de ses microphones.

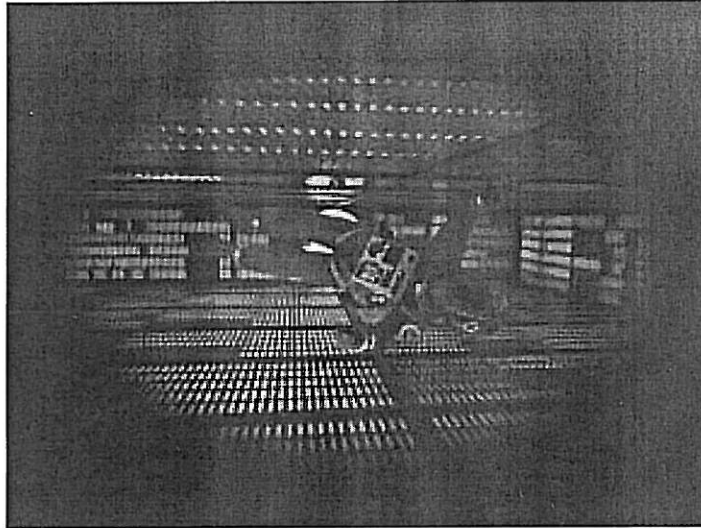


Figure 1. Dave désactive l'ordinateur HAL.
« I'm afraid. I'm afraid Dave. Dave, my mind is going. I can feel it. I can feel it. »

- **Une erreur de prédiction.** Ce qui différencie HAL des cyborgs, droïdes et robots et autres machines intelligentes en tous genres qui ont peuplé les films de science-fiction depuis (par exemple l'adorable C3PO de la *Guerre des Etoiles*, Figure 2), c'est qu'il était une projection tout à fait sérieuse de ce que les scientifiques de l'époque (le film date de 1968) considéraient comme probable trente ans plus tard en matière d'**intelligence artificielle**. Kubrick, comme l'expliquait la *MGM* lors de la sortie du film, avait essayé de peindre de la façon la plus réaliste possible les choses telles qu'elles seraient en 2001, en s'appuyant sur l'avis de scientifiques éminents³. Du point de vue astronautique, l'homme a mis le pied sur la lune très peu de temps après la sortie du film, et, des explorations du système solaire telles que celle décrite dans le film ne sont pas inconcevables dans un futur de quelques décennies. En ce qui concerne l'intelligence de HAL et sa capacité de langage, par contre, les prédictions étaient complètement erronées. Aucune machine approchante n'existe, même dans les cartons. Pire, plus personne ne se hasarde à proposer une date à laquelle la construction d'une telle machine intelligente serait envisageable.

- **Pourquoi une telle difficulté ? Le traitement automatique des langues (T.A.L.)** est né à la fin des années 40. A titre de comparaison, la découverte du code génétique a été à peu près contemporaine des débuts du T.A.L., et l'on peut jalonner le demi-siècle de travaux en génétique de résultats clés, ayant abouti récemment à la cartographie grossière du génome humain. De nombreuses autres disciplines ont évolué de façon spectaculaire (informatique, physique, etc.). Pourquoi le langage humain est-il aussi difficile à comprendre ? Pourquoi ses règles sont-elles si difficiles à déterminer et à fournir à des ordinateurs ? Ce sont les questions de fonds qui seront abordées dans ce cours.

- **Quel est l'état de l'art ?** Même si HAL n'existe pas, des voitures et autres machines parlent grâce à la **synthèse de la parole**, on peut écrire ses lettres grâce à des logiciels de **dictée vocale**, chacun a un **correcteur orthographique** intégré dans son système de traitement de textes (qui nous promet désormais de corriger notre grammaire et notre style), l'on peut tester des systèmes de **traduction automatique** sur le Web. Quelle est la qualité véritable de ces services ? Sont-ils efficaces, ou bien gadgétiques,

voire inutilisables ? Quelle est l'évolution prévisible des systèmes de T.A.L. ? Nous essaierons aussi de répondre à ces questions, tout en analysant les principes de fonctionnement des systèmes existants, ainsi que les raisons de leurs limitations.



Figure 2. « Bonjour, je suis C3PO, cyborg et droïde de protocole. Je pratique couramment six millions de formes de communication... » (*La Guerre des Etoiles*)

2. Questions pratiques

2.1. Statut du cours

Cette unité d'enseignement peut être suivie :

- soit en **option**, pour toutes disciplines et tous niveaux (elle fait partie de la liste 9 du guide des études);
- soit comme unité obligatoire de la licence des Sciences du Langage, mention **Traitement automatique des langues**.

2.2. Organisation

2.2.1. Cours/TD

Cette unité d'enseignement est semestrielle (1er semestre) et se compose :

- d'un **cours théorique**, d'une heure par semaine :
 - horaire à préciser, salle C212 (Jean Véronis)
- d'un **TD/TP**, de deux heures par semaine :
 - horaire et enseignants à préciser, salle A276

Le cours et les TD/TP sont disponibles sur l'Intranet du CILSH, accessible à partir du libre-service (ci-dessous) : voir en X:\INFZ18.

2.2.2. Télé-enseignement

L'unité d'enseignement est disponible par **télé-enseignement** (renseignements au CTE, bureau 63).

2.2.3. Libre-service

Les étudiants suivant cette unité d'enseignement (y compris par télé-enseignement) ont accès au **libre-service** salle C206, où ils peuvent effectuer leur travaux en temps libre.

Lors du premier accès, les étudiants doivent faire tamponner leur carte d'étudiant, qu'ils devront présenter à chaque utilisation du libre-service.

Les règles de fonctionnement du libre-service sont précisées dans une « charte » que chaque étudiant s'engage à respecter. Tout étudiant qui dérogera aux règles de bonne utilisation sera définitivement exclus.

2.3. Modalités de contrôle

2.3.1. Régime général

Cette unité d'enseignement est soumise au **contrôle continu**, composé de:

- **deux travaux** à remettre à une date fixée en cours de semestre, comptant chacun pour 30% de la note ;

- un devoir sur table (questions de cours) à la fin du semestre, comptant pour 40% de la note.

2.3.2. Cas particuliers

Les étudiants (salariés, etc.) bénéficiant du régime dérogatoire ainsi que les étudiants inscrits au télé-enseignement (CTE) peuvent passer un examen final au lieu du contrôle continu. Ils sont invités à se faire connaître auprès de l'enseignant, ainsi que que les étudiants handicapés désireux d'un aménagement personnalisé des modalités de contrôle.

2.4. Contacts

Le responsable du cours est M. Jean Véronis, professeur, qui reçoit le jeudi de 14 à 15h, bureau C202 (ou sur rendez-vous auprès du secrétariat, bureau A293).

Courrier électronique: <jean.veronis@up.univ-mrs.fr>

Des informations complémentaires seront disponibles sur le Web :

<http://www.up.univ-mrs.fr/~wcoilsh/infz18>

3. Bibliographie

Cette section indique les ouvrages généraux, se rapportant à la totalité du cours. Chaque chapitre du cours fournit des références additionnelles sur des points particuliers.

- Bouillon, P. (Ed.) (1998). *Traitement automatique des langues naturelles*. Aupelf-Urelf et Éditions Duculot, Paris.
- Carré, R., Dégremont, J.-F., Gross, M., Pierrel, J.-M., Sabah, G. (1991). *Langage humain et machine*. Presses du CNRS, Paris.
- Danlos, L., Véronis, J. (Eds.) (1997). *Etat de l'Art*. Numéro spécial de la revue *Traitement Automatique des Langues*, 38(2).
- Fuchs, C., Danlos, L., Lacheret-Lajour, A., Luzzati, D., Victorri, B. (1993). *Linguistique et traitements automatiques des langues*. Hachette, Paris.
- Sabah, G. (1988). *L'intelligence artificielle et le langage*. Hermès, Paris.
- Wehrl, E. (1997). *L'analyse syntaxique des langues naturelles: problèmes et méthodes*. Masson, Paris.

4. Filmographie

- Stanley Kubrick (1968). *2001, A Space Odyssey*.

5. Notes

¹ HAL est un acronyme pour «Heuristically Programmed ALgorithmic computer» (ordinateur à programmation heuristique). On a vite remarqué qu'en décalant chaque lettre de HAL d'une position dans l'alphabet on obtenait... IBM. Stanley Kubrick et Arthur C. Clarke (co-auteur du scénario) ont assuré que ce n'était pas intentionnel, mais dans la version française, sortie quelque temps après, HAL est rebaptisé CARL (Cerveau Analytique de Recherche et de Liaison).

² Sur 2001 voir le site :

<http://www.underview.com/30years.html>

et sur HAL en particulier, les pages :

<http://www.underview.com/home98/hal.html>

<http://www.underview.com/home98/haltrans.html> (transcription des dialogues)

<http://www.underview.com/home98/viewshal.html>

On trouvera un site consacré à HAL à l'adresse :

<http://matia.stanford.edu/~stork/HALTalks.html>

³ Par exemple, Marvin Minsky, l'un des pères de l'intelligence artificielle.

Informatique et linguistique

Cours 2

Historique (1)

Du rêve...

Jean Véronis

1. Introduction

L'ordinateur HAL de 2001, *L'Odyssée de l'Espace* est sans doute l'héritier d'un mythe ancestral : depuis les temps les plus reculés, l'homme semble fasciné par la possibilité de créer des **automates**¹ à son image.

Encyclopaedia Universalis. AUTOMATE. Un automate (du grec αυτοματον) est une machine imitant les mouvements, les fonctions ou les actes d'un corps animé. [...]

Ce mythe a peuplé de multiples légendes, et à chaque étape du développement technologique, des tentatives ont été faites pour le transformer en réalité. Les aspects mécaniques simples ont sans doute été les premiers à faire l'objet de tentatives de reproduction, mais au fil des siècles on a tenté d'imiter la plupart des fonctions de l'être humain, que ce soit dans un but magique, récréatif ou scientifique. La **capacité de langage**, qui est propre à l'homme, est l'une des fonctions les plus complexes, car elle repose sur la compréhension, donc l'intelligence, et on ne pouvait guère envisager d'en reproduire autre chose que les aspects purement mécaniques jusqu'à la naissance de l'ordinateur, au milieu du XX^{ème} siècle. Il n'est pas certain que les espoirs d'une **intelligence artificielle** que l'ordinateur a fait naître aient été pour l'instant comblés, ni même qu'ils puissent l'être un jour.

HAL reste pour l'instant un rêve, mais de nombreuses **applications** bien réelles et concrètes ont vu le jour, qui commencent à être utilisées par le grand public : correction orthographique, dictée vocale, etc. Bien que portant sur des aspects limités du langage, elles rendent des services appréciables et progressent rapidement, dans un domaine en constante expansion, celui des **industries de la langue**.

Nous examinerons successivement ci-après les différentes étapes du passage du rêve à la réalité :

Cours n°2 (ce cours) :

- Mythes et légendes
- Les automates de l'Antiquité à la Renaissance
- La mécanique des langues au XVII^{ème} et XVIII^{ème} siècles

Cours n°3 :

- La naissance de l'ordinateur et les espoirs naïfs
- La linguistique computationnelle

2. Mythes et légendes

Bibliographie

Encyclopaedia Universalis.

- Article AUTOMATE.

Breton, Philippe (1990). *Une histoire de l'informatique*. Édition du Seuil, Collection Points Sciences. [Chapitre 1. Automates et êtres artificiels]

2.1. Antiquité

- Héphaïstos moule dans la glaise un corps de vierge
- Vulcain crée des trépieds ambulants, forge une femme en or
- Aphrodite anime la statue de **Pygmalion**
- Athéna apprend aux Rhodiens à fabriquer des statues mobiles
- Dédale anime si bien ses œuvres qu'on doit les enchaîner

Encyclopaedia Universalis. PYGMALION. Dans la mythologie grecque, roi de Chypre qui tomba amoureux d'une statue d'Aphrodite. Dans ses *Métamorphoses*, Ovide donna de l'histoire une version plus élaborée : le sculpteur Pygmalion fit une statue d'ivoire représentant son idéal de la femme et s'éprit de son œuvre. En réponse à ses prières, Aphrodite donna vie à la statue. George Bernard Shaw a développé sur le mode ironique le thème de l'artiste créateur dans son *Pygmalion* (1913) dont fut tirée, en 1956, la comédie musicale *My Fair Lady*, de Lerner et Loewe.

2.2. Moyen-Âge

- Légende juive du **Golem**

Encyclopaedia Universalis. GOLEM. Être, le plus souvent de forme humaine, le *golem* est créé par un acte de magie grâce à la connaissance des dénominations sacrées. [...] [Dans la tradition hassidique d'Europe Centrale des XII^e et XIII^e siècles] ceux qui, au cours d'une réunion mystique, participaient à «l'acte de création» prenaient un peu de terre vierge et en faisaient une idole ; puis ils tournaient autour d'elle en une sorte de danse en prononçant les lettres sacrées et le nom secret de Dieu, selon un ordre et des protocoles détaillés. Le golem prenait alors vie ; quand les initiés inversaient le sens de leur danse ainsi que l'ordre des lettres sacrées, le golem s'écroulait et perdait la vie. Selon d'autres légendes, le mot *Emet* (la Vérité ou le Sceau du Dieu unique) devait être écrit sur le front du golem ; quand la lettre *alef* était effacée, ne demeurait plus que le mot *met* (mort) et le golem s'anéantissait. Dans la tradition populaire du hassidisme ashkénaze du XV^e siècle, le golem devint une créature réelle, capable de servir ses maîtres et de remplir les tâches qu'ils lui fixaient. Cette tradition, qui devint extrêmement populaire au XVII^e siècle, se rattache à la très ancienne croyance en la possibilité de ressusciter un mort en lui mettant dans la bouche (ou sur le bras) un morceau de parchemin sur lequel est inscrit le tétragramme. D'autre part, elle se rapproche beaucoup de nombreuses légendes ésotériques non juives concernant la création d'homoncules (comme on le voit chez Paracelse, par exemple). Enfin, selon cette croyance, le golem, être servile, peut se changer en un être maléfique qu'il convient de détruire pour éviter qu'il ne sème la terreur et la mort. [...]

2.3. Littérature

- Hoffmann (1776-1822) : *Coppelia*
- Mary Shelley (1797-1851): *Frankenstein*
- Villiers de l'Isle-Adam (1838-1889) : *L'Eve future.*
- George Bernard Shaw (1856-1950) : *Pygmalion* => *My Fair Lady*
- Karel Capek (1890-1938) : *R.U.R. (Rossum's Universal Robots)* (donne le terme robot)
- Science-fiction : Isaac Asimov, etc.

3. Les automates de l'Antiquité à la Renaissance

Bibliographie

Encyclopaedia Universalis.

- Article AUTOMATE.

Breton, Philippe (1990). *Une histoire de l'informatique*. Edition du Seuil, Collection Points Sciences. [Chapitre 1. Automates et êtres artificiels]

3.1. Egypte

Automates à vocation magique ou religieuse.

- Statues animées (mentionnées par Hérodote, etc.)
- Masque articulé d'Anubis (dieu au corps d'homme et à la tête de chacal), dont la mâchoire était actionnée par des fils cachés et qui semblait prononcer des paroles dictées par les prêtres (Musée du Louvre).

3.2. Grèce et Alexandrie

Période des μηχανοποιοι, constructeurs de machines ; utilisation de mécanismes simples (siphon, roue dentée, came, ressort) et de techniques hydrauliques (valve, piston).

- Archytas (438-365 av. J.-C.) : construisit une colombe volante qui se mouvait « par l'air qui était enfermé et caché »
- Ingénieurs alexandrins : Ktésibios (III^{ème} siècle av. J.-C., Philon de Byzance (230 av. J.-C.), Heron d'Alexandrie (125 av. J.-C.).

Encyclopaedia Universalis. AUTOMATE. [...] Heron d'Alexandrie (125 av. J.-C.) indiqua les procédés de construction d'un grand nombre d'automates. Il y a lieu de noter une de ses machines ludiques : l'eau d'une vasque coule dans un piédestal creux, cependant que des oiseaux posés au bord de la vasque chantent jusqu'à ce que le piédestal soit plein. Un siphon le vide alors dans un seau suspendu à une corde avec poulie et contrepoids. Les oiseaux reprennent leur chant. Le seau plein descend en faisant tourner le perchoir d'un hibou, qui reprend, une fois le seau siphonné, sa position première : ses déplacements semblent commander le chant des oiseaux. Heron est connu pour ses machineries scéniques notées dans son *Traité des pneumatiques*.

Avec Heron, l'automate atteint à un grand degré de perfectionnement :

1. Les mécanismes agissent en vertu de leur structure interne.
2. L'action relève d'un aménagement des forces motrices, naturelles avec la pesanteur (matières pondéreuses, eau), artificielles avec la vapeur ou l'air comprimé.
3. Les plus perfectionnés sont mobiles, l'ensemble pouvant se déplacer seul (trépieds de Vulcain).

On ignore si ces automates furent construits. C'est avec beaucoup de réserve qu'il faut retenir les prétentions de Heron.

L'activité des mécaniciens alexandrins n'eut qu'un temps et le goût des automates passa chez les Grecs de Byzance.

3.3. Rome

Les neurospates, sortes d'effigies manœuvrées grâce à un système de cordages, comportaient un dispositif permettant au prêtre de parler à travers la statue.

3.4. Moyen-Âge

Héritage byzantin et arabe. Roger II de Sicile (1130-1154) et Frédéric II (1194-1254) accueillent savants et mécaniciens chrétiens, juifs, musulmans. Moteurs mus par l'eau ou le vent. Mécanismes d'horlogerie.

- Jardins à automates
- Horloges monumentales dotées de jacquemarts (Saint-Paul à Londres, 1286, Beauvais, 1320, Paris, 1370). À Strasbourg (1373), les rois mages venaient s'incliner devant la Vierge pendant que le chant d'un coq battant des ailes répondait au cantique égrené par le carillon.
- Automates androïdes et têtes parlantes :
 - le dominicain Albert le Grand (1204-1282) aurait construit un homme artificiel de métal, bois, cire et cuir. Crée une tête qui répond aux questions et résout des problèmes (elle fut détruite par Thomas d'Aquin, un de ses anciens étudiants, qui y voyait l'incarnation du diable).
 - Le franciscain anglais Roger Bacon (1214-1292) aurait, lui aussi, réalisé un « homme de fer » ; on l'accusa de magie.

3.5. Renaissance

Nouvel essor des techniques (vulgarisation du système bielle-manivelle, miniaturisation des mécanismes d'horlogerie). Vision mécaniste du corps humain.

- Leonardo da Vinci (1452-1519) : dans ses dessins anatomiques, décompose les mouvements des membres dans un jeu de fils associés à des leviers osseux.
- Rabelais (1494-1563) : assimile le corps à un groupement de leviers, de poulies, de cordons et de filtres.
- L'anatomiste Vésale (1514-1564) décrit la « fabrique » du corps humain comme « un agencement mécanique effectuant les travaux pour lesquels il a été construit et dont les parties ne méritent pas d'intérêt si on les sépare de l'ensemble ».

4. La mécanique des langues aux XVII^{ème} et XVIII^{ème} siècles

Bibliographie

Encyclopaedia Universalis.

- Article AUTOMATE.
- Article DESCARTES (section 4. « La science cartésienne – La biologie »).
- Article LEIBNIZ (section 4. « L'activité analogique – La combinatoire »).
- Article VIE (section « La vie comme mécanisme »).

Auroux, Sylvain (Ed.) (1992). *Histoire des idées linguistiques*. Liège.

Descartes, René (1637). *Discours de la méthode*. Partie V.

Eco, Umberto (1994). *La recherche de la langue parfaite dans la culture européenne*. Ed. du Seuil, Coll. Points/Essais, Paris [traduit de l'italien].

Séris, Jean-Pierre (1995). *Langages et machines à l'âge classique*. Hachette, Paris.

Webographie

Sites consacrés aux automates :

- *Delightful Machines* : <http://www.database.com/~lemur/dm-delightful-machines.html>

Site consacré aux machines parlantes (*Haskins Laboratories*) :

- <http://macserver.haskins.yale.edu/Haskins/HEADS/talkingheads.html>
- (en particulier <http://macserver.haskins.yale.edu/Haskins/HEADS/simulacra.html>)

MAITRE DE PHILOSOPHIE. Il y a cinq voyelles ou voix : A, E, I, O, U.

MONSIEUR JOURDAIN. J'entends tout cela.

MAITRE DE PHILOSOPHIE. La voix A se forme en ouvrant fort la bouche : A.

MONSIEUR JOURDAIN. A, A, oui.

MAITRE DE PHILOSOPHIE. La voix E se forme en rapprochant la mâchoire d'en bas de celle d'en haut : A, E.

MONSIEUR JOURDAIN. A, E, A, E. Ma foi ! oui. Ah ! que cela est beau.

MAITRE DE PHILOSOPHIE. Et la voix I en rapprochant encore davantage les mâchoires l'une de l'autre, et écartant les deux coins de la bouche vers les oreilles : A, E, I.

MONSIEUR JOURDAIN. A, E, I, I, I, I. Cela est vrai. Vive la science !

MAITRE DE PHILOSOPHIE. La voix O se forme en rouvrant les mâchoires, et rapprochant les lèvres par les deux coins, le haut et le bas : O

MONSIEUR JOURDAIN. O, O. Il n'y a rien de plus juste. A, E, I, O, I, O. Cela est admirable ! I, O, I, O.

MAITRE DE PHILOSOPHIE. L'ouverture de la bouche fait justement un petit rond qui représente un O.

MONSIEUR JOURDAIN. O, O, O. Vous avez raison. O. Ah ! la belle chose que de savoir quelque chose !

MAITRE DE PHILOSOPHIE. La voix U se forme en rapprochant les dents sans les joindre entièrement, et allongeant les lèvres en dehors, les approchant aussi l'une de l'autre sans les joindre tout à fait : U.

MONSIEUR JOURDAIN. U, U. Il n'y a rien de plus véritable : U.

MAITRE DE PHILOSOPHIE. Vos deux lèvres s'allongent comme si vous faisiez la moue : d'où vient que si vous la voulez faire à quelqu'un, et vous moquer de lui, vous ne sauriez lui dire que : U.

MONSIEUR JOURDAIN. U, U. Cela est vrai. Ah ! que n'ai-je étudié plus tôt, pour savoir tout cela ?

MAITRE DE PHILOSOPHIE. Demain, nous verrons les autres lettres, qui sont les consonnes.

Molière, *Le Bourgeois Gentilhomme* (1670).

4.1. Une science mécaniste

L'influence des automates et de la mécanique sur la science s'accroît aux XVII^{ème} et XVIII^{ème} siècles. On bâtit une science du vivant sur le modèle des machines. René Descartes (1596-1650) (Figure 1) compare explicitement les êtres vivants aux automates aperçus « dans les jardins de nos rois ». Pour lui le corps est une machine : les nerfs sont comparés aux tuyaux d'une machine, les muscles « aux divers engins qui servent à (la) mouvoir », la respiration et les autres fonctions « aux mouvements d'une horloge ou d'un moulin »... Descartes dessinera « une perdrix artificielle » qu'un épagneul fait lever, et imaginera une femme automate, Francine, qui ne sera jamais construite. Tout un courant de « iatomécaniciens » apparaît, qui expliquent le fonctionnement des organes par des modèles mécaniques inspirés de la mécanique galiléenne ou cartésienne.

« Je désire que vous considériez que ces fonctions suivent toutes naturellement, en cette Machine, de la seule disposition de ses organes, ne plus ne moins que font les mouvements d'une horloge ou autre automate, de celle de ses contrepoids et de ses roues ; en sorte qu'il ne faut point à leur occasion concevoir en elle aucune autre Âme végétative ni sensitive, ni aucun autre principe de mouvement et de vie, que son sang et ses esprits agités par la chaleur du feu qui brûle continuellement dans son cœur et qui n'est point d'autre nature que tous les feux qui sont dans les corps inanimés. »

René Descartes, *Traité de l'homme* (1633, publié seulement en 1662-1664)

Au-delà de la science du vivant, le « mécanisme » s'impose comme philosophie dominante, en particulier au XVIII^{ème} siècle. Il conquiert même le terrain du langage et de la grammaire, comme nous le verrons ci-dessous.



Figure 1. René Descartes (1596-1650)

4.2. La mécanique de la parole

4.2.1. L'étude de la phonation

Correspondant de Galilée, Descartes et de nombreux autres savants et philosophes, le père Marin Mersenne (1588-1648) (Figure 2) aborde l'étude de la phonation, d'un point de vue articulatoire, acoustique et mécanique. Il décrit la mécanique du conduit vocal et des différents organes mis en œuvre dans la parole, et tente d'expliquer comment se produisent les différents sons du langage :

Les sons ne feraient aucune voyelle si on n'avait point de langue, dont le plus simple abaissement qui se fait au bout forme la première voyelle A, lorsqu'elle s'étend et soutient le son ; l'O se fait quasi par la même situation de la langue, car elle se retire et s'enfle fort peu vers les milieu du palais. Mais les lèvres n'ont pas tant d'ouverture pour faire l'A que pour faire l'O, car elles se rétrécissent nécessairement, et il n'est pas possible de former l'O avec la même ouverture des lèvres dont on use pour former l'A. D'où il s'ensuit que les lèvres sont nécessaires pour former les voyelles.

Marin Mersenne, *Harmonie Universelle* (1636-1637). (Proposition XLIII).



Figure 2. Marin Mersenne (1588-1648)

La découverte de cette mécanique des sons et de la phonation émerveille les esprits du XVII^{ème} siècle et de nombreux savants et philosophes, tels que Cordemoy ou Lamy, publient des traités de phonétique et de physique de la parole. Le retentissement en est si grand que Molière s'en moque quelque peu dans la « leçon d'orthographe » du *Bourgeois Gentilhomme*, (1670) qui est un pastiche d'un passage du *Discours physique de la parole* de Cordumoy paru deux ans plus tôt (mais est-ce le Maître de Philosophie qui est ridiculisé ou Monsieur Jourdain ?). Les études mécanistes se continuent au XVIII^{ème} siècle avec les traités de Dodart sur la voix chantée (1700 à 1707), de Ferrein qui fait « chanter les morts » pour étudier la formation de la voix (1741), etc.

Tout le monde croyait [...] qu'on ne pouvait rendre l'organe de la voix humaine visible en action, ni le faire sonner quand il n'est plus animé par le principe de la vie, cependant je résolus de le tenter. Je pris un cadavre, je soufflai à plusieurs reprises de bas en haut dans la trachée artère, le larynx fut muet en cette occasion. Je fis réflexion dans la suite que la voix ne demande pas seulement un vent plus fort, mais encore un nouveau degré de rétrécissement dans le larynx ; je pris celui d'un chien, le rapprochai les lèvres de cette glotte et soufflai fortement dans la trachée artère ; à ce coup, l'organe parut s'animer, et fit entendre, je ne dis pas seulement un son, mais une voix éclatante, plus agréable pour moi que les concerts les plus touchants. J'avais un cadavre humain destiné à des usages publics, je ne pus m'empêcher de le sacrifier à mon impatiente curiosité, elle fut pleinement satisfaite ; les moyens dont j'ai parlé ayant été mis en œuvre, le larynx du cadavre répondit par un éclat qui étonna les assistants, et c'est, je pense, la première fois qu'on ait vu pareil phénomène [...]

Antoine Ferrein, *De la formation de la voix de l'homme* (1741).

4.2.2. Les machines parlantes

Mersenne évoque la possibilité de construire un orgue prononçant des sons de la parole :

L'on peut aussi adjoindre toutes sortes d'instruments à vent aux orgues ordinaires, par exemple toutes sortes de Musettes et de Hautbois, et même les instruments à corde, et ceux de percussion, qui feront des harmonies très particulières, par le moyen de plusieurs roues et d'autres ressorts, dont l'invention est assez facile, car elle ne dépend que de l'hydrauliques et des autres parties des Mécaniques, dont on peut tirer des secrets assez avantageux pour faire parler les tuyaux d'orgue et pour leur faire prononcer toutes sortes de syllabes, et conséquemment toutes sortes de dicions. Mais cet ouvrage consiste en tant de ressorts et de différents mouvements, qu'il désire l'esprit, le travail et la vie de plusieurs hommes, c'est pourquoi j'en laisse le soin et l'invention à la postérité.

Marin Mersenne, *Harmonie Universelle* (1636-1637). (Proposition XXI).

Dans *l'Art de Parler* (1675), l'oratorien **Bernard Lamy** (1640-1715), élabore ce thème et discute du nombre de tuyaux qu'il faudrait pour réaliser toutes les combinaisons de sons formant les syllabes. Le père **Athanase Kircher** (1601-1680), inventeur d'une machine à écrire, réalise une tête automate qui profère des sons.

Le XVIII^{ème} siècle voit apparaître une floraison d'automates. Pierre Jaquet Droz (1721-1790) et **Jacques de Vaucanson** (1709-1782) sont sans doute les plus connus des constructeurs d'automates du siècle des Lumières. En héritiers du mécanisme de Descartes, ils cherchent à reproduire les principales fonctions de la vie (digestion, circulation, respiration). Parmi les automates les plus célèbres figurent « Le Joueur de Flûte », automate androïde d'1m50 qui pouvait jouer une douzaine d'airs, et le « Canard », qui battait des ailes, mangeait du grain et le rendait digéré (la digestion s'est avérée être une supercherie), présentés à Paris en 1738. Il est probable que Vaucanson envisageait la construction d'un automate parleur.

Le grand mathématicien **Leonard Euler** (1707-1783) a compris l'importance d'une telle invention :

Ce serait sans doute une des plus importantes découvertes, que de construire une machine qui fût propre à exprimer tous les sons de nos paroles avec toutes les articulations. Si l'on réussissait jamais à exécuter une telle machine, et qu'on fût en état de lui faire prononcer toutes les paroles au moyen de certaines touches, comme d'un orgue ou d'un clavecin, tout le monde serait avec raison surpris d'entendre qu'une machine prononçât des discours entiers, ou des sermons, qu'il serait possible d'accompagner avec la meilleure grâce. Les prédicateurs et les orateurs dont la voix n'est pas assez agréable pourraient alors jouer leurs sermons et discours sur une telle machine, tout de même que les organistes jouent les pièces de musique. La chose ne me paraît pas impossible.

Léonard Euler, *Lettre à une princesse d'Allemagne* (1761).

En 1780, le Danois **Kratenstein** remporte le prix annuel de l'Académie Impériale de Saint-Pétersbourg en construisant un orgue composé d'une série de résonateurs capables de prononcer les cinq voyelles (Figure 3).

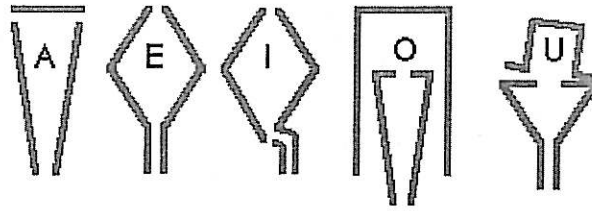


Figure 3. Résonateurs de Kratenstein (1770)

L'abbé Mical construit en 1778 une « Tête d'airain » capable de prononcer une phrase, et présente à l'Académie des Sciences en 1783 une machine composée de deux têtes parlantes, qui séduit l'écrivain et journaliste Rivarol :

M. Mical applique deux claviers à ses *Têtes parlantes* ; l'un en cylindre par lequel on n'obtient qu'un nombre déterminé de phrases, mais sur lequel les intervalles des mots et leur prosodie sont marqués correctement, l'autre clavier contient dans l'étendue d'un ravalement tous les sons et tous les tons de la langue française, réduits à un petit nombre par une méthode ingénieuse et particulière à l'auteur. Avec un peu d'habitude et d'habileté, on parlera avec les doigts comme avec la langue, et on pourra donner au langage des *Têtes* la rapidité, les repos et toute la physionomie enfin que peut avoir une langue qui n'est point animée par les passions. Les étrangers prendront la *Henriade* ou le *Télémaque* et les feront réciter d'un bout à l'autre en les plaçant sur le clavier vocal comme on place des partitions d'opéra sur les clavecins ordinaires.

Rivarol. *Lettre à M. le Président de ****, sur le globe aérostatique, sur les têtes parlantes et sur l'état présent de l'opinion publique à Paris.

C'est le baron **Wolfgang von Kempelen** (1734-1804) qui a réalisé à Vienne en 1791 la machine parlante la plus perfectionnée (Figure 4), après avoir construit un automate plus rudimentaire dès 1778. Von Kempelen est connu pour son automate joueur d'échec qui fut présenté dans divers pays et remporta des parties contre les grands de ce monde, mais qui s'avéra être une supercherie (celle-ci fut dénoncée par Edgar Poe dans ses *Histoires grotesques et merveilleuses*), un joueur humain de petite taille étant caché dans la machine. Cette supercherie porta sans doute du tort à la crédibilité de sa machine parlante, qui était pourtant une invention remarquable pour l'époque. Un soufflet faisait vibrer une anche qui excitait un résonateur unique dont on faisait varier la forme pour les différentes voyelles avec une main. Les consonnes, y compris les nasales étaient produites par quatre passages d'air que l'on contrôlait avec les doigts de l'autre main.

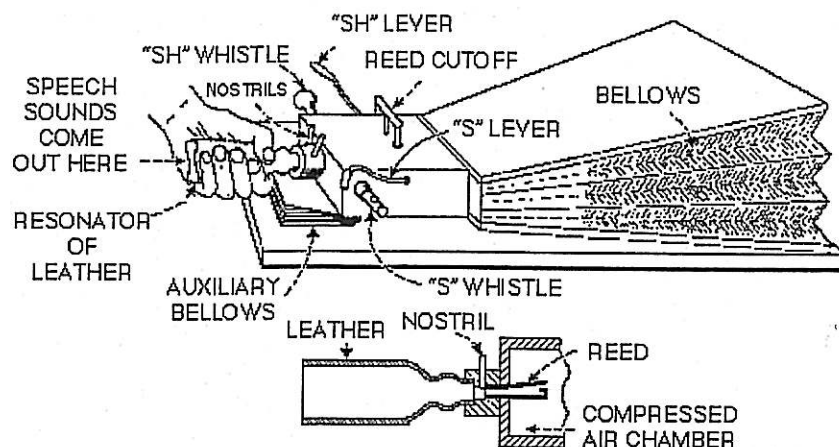


Figure 4. La machine parlante de von Kempelen d'après la description de Wheatstone au XIX^{ème} siècle (James L. Flanagan, "Speech Analysis, Synthesis and Perception", Springer-Verlag, 1965, pp. 166-167)

4.3. La mécanique grammaticale et sémantique

4.3.1. La grammaire

Il était sans doute logique que le contexte mécaniste du XVIII^{ème} siècle influence l'étude et la simulation de la parole, qui relève après tout du domaine physique et mécanique, du moins dans ses aspects acoustiques et articulatoires. Il est plus surprenant cependant que les mêmes idées aient conquis le domaine de la grammaire, qui relève essentiellement du domaine abstrait et logique.

Depuis l'Antiquité, les régularités et la systématisme du langage ont été observées par les grammairiens, mais la façon d'en parler change entre le XVII^{ème} et le XVIII^{ème} siècle. Des règles au mécanisme il n'y a qu'un pas, et l'on voit apparaître les expressions de « mécanisme du langage », « **mécanisme des langues** », sous la plume de Du Marsais, Beauzée, Diderot, D'Alembert et bien d'autres. L'Abbé Noël Antoine de Pluche publie même en 1751 un ouvrage intitulé *La mécanique des langues et l'art de les enseigner*, et Charles de Brosses, en 1765, un *Traité de la formation mécanique des langues et Principes physiques de l'Étymologie*. Les règles de construction grammaticale ou de formation des langues (étymologie) sont vues comme des systèmes mécaniques, **assimilables au calcul** numérique.

Locke, et depuis, M. l'abbé de Condillac, ont montré que le langage est véritablement une espèce de calcul, dont la grammaire et même la logique en grande partie, ne sont que les règles ; mais ce calcul est bien plus compliqué que celui des nombres, sujet à bien plus d'erreurs et de difficultés.

Turgot. *Réflexions sur les langues* (1750).

À côté des préoccupations grammaticales, il faut noter depuis le XVII^{ème} siècle un engouement extraordinaire pour la création de **langues artificielles** ou « langues universelles ». Séris ne recense pas moins de 16 ouvrages parus entre 1627 et 1668 sur le sujet, depuis la *Proposition présentée au Roi d'une écriture universelle admirable parmi ses effets, très utile à tous les hommes de la terre* de Jean Drouet (1627). Cet essor « logotechnique » se double d'un courant très important d'études de **cryptographie** (commencé au XVI^{ème} siècle : Cardan, Vigenère, etc.).

Les idées qui président aux essais de construction de langues artificielles sont les mêmes que celles qui régissent la cryptographie (et les ouvrages traitent souvent simultanément des deux) : dans les deux cas, les techniques mises en œuvre sont des mécanismes et des calculs. Les langues artificielles sont plus complexes, à cause d'exceptions et d'accidents divers, mais elles relèvent fondamentalement des mêmes mécanismes. La conjonction de la logotechnique et de la cryptographie fait émerger l'idée de **traduction mécanique**. Pour Sérís, « le projet cryptographique et le projet de langue universelle reposent sur un assomption commune, celle de traductibilité "tous azimuts" pour parler de façon familière. [...] Au point que l'on peut voir les langues existantes comme n'étant que des chiffres différents, pour celui qui parvient à l'archi-écriture ». Il est intéressant de noter que c'est précisément cette idée, redécouverte, qui donnera naissance aux recherches en traduction automatique deux ou trois siècles plus tard, comme nous le verrons dans le cours suivant.

4.3.2. La sémantique

Si les projets de langue universelle sont souvent conçus plutôt comme recherche d'une **écriture universelle**, permettant de servir de moyen de communication entre peuples, Descartes, puis Leibnitz font le projet de **langues philosophiques** qui supposent le « dénombrement de toutes les pensées des hommes et la découverte de l'ordre naturel de leur composition, à partir des éléments primitifs² ».

Si quelqu'un avait bien expliqué quelles sont les idées simples qui sont en l'imagination des hommes, desquelles se compose tout ce qu'ils pensent, et que cela fut reçu par tout le monde, j'oserais espérer ensuite une langue universelle fort aisée à apprendre, à prononcer et à écrire, et, ce qui est le principal, qui aiderait au jugement, lui représentant si distinctement toutes choses, qu'il lui serait presque impossible de se tromper.

Descartes, *Lettre à Mersenne*, 20 novembre 1629.

Le pas est considérable, car l'on passe de la mécanisation de la grammaire à la mécanisation de la sémantique (le terme n'est apparu qu'au XIX^{ème} siècle) et du raisonnement. On peut sans doute rétrospectivement juger quelque peu naïve la sous-estimation des difficultés³, mais dès 1661, **Dalgarno** propose « une langue artificielle basée sur l'attribution d'une lettre à chacune des catégories irréductibles de la pensée, et des règles de composition et de variation »⁴. **Wilkins** propose en 1668 un système plus complexe, basé sur quarante catégories, et utilisant au lieu de lettres des symboles « composables et modifiables par des traits diversement tracés correspondant aux particules » (Knecht). Il souhaite que sa langue soit une langue à part entière, et en particulier veille à ce qu'elle soit prononçable.

Hobbes (1588-1679) ramène la pensée au calcul :

Par raisonnement (*ratiocinatio*) j'entends un calcul. Calculer, c'est ou bien composer la somme totale de plusieurs choses ajoutées les unes aux autres, ou bien une chose étant soustraite à une autre, connaître ce qui reste. Reasonner est donc la même chose qu'additionner ou soustraire. Je ne m'opposerais pas à qui voudrait y adjoindre la multiplication et la division, la multiplication n'étant rien d'autre que l'addition en quantités égales, et la division, que la soustraction, autant de fois qu'il se peut, de quantités égales. Tout le raisonnement revient donc à deux opérations de l'esprit, l'addition et la soustraction.

Hobbes, *De corpore*. (1655).

En même temps, Hobbes précise que le calcul ne porte pas nécessairement sur des nombres :

Il ne faut pas croire que calcul et raisonnement n'ont lieu que sur des nombres, comme si l'homme ne se distinguait des autres animaux (comme on dit que Pythagore l'a soutenu) par rien d'autre que par la faculté de nombrer ; car on peut ajouter ou soustraire un grandeur à une grandeur, un corps à un corps, un mouvement à un mouvement, un temps à un temps, un degré de qualité à un degré de qualité, une action à une action, une conception à une conception, une proportion à une proportion, un discours à un discours et un nom à un nom (en quoi consiste toute sorte de philosophie).

Hobbes, *De corpore*. (1655).

Ces citations préfigurent de façon frappante les principes de ce qui va être l'informatique (c'est-à-dire le traitement de l'information, pas forcément numérique, par le calcul), trois siècles plus tard.

C'est **Gottfried Wilhelm von Leibniz** (1646-1716) (Figure 5), influencé par Hobbes, qui pousse le plus loin l'analogie entre langue et raisonnement d'une part, et calcul d'autre part. Il cherche à développer un **alphabet des pensées humaines**, qui servirait de base à une **langue algébrique** (*lingua characteristica universalis*) qui permettrait de communiquer comme les langues naturelles, mais serait mécanique et systématique comme les mathématiques.



Figure 5. Gottfried Wilhelm von Leibniz (1646-1716)

Les courants de pensée qui se développent au XVII^{ème} et au XVIII^{ème} siècle contiennent de façon étonnante le germe des idées qui serviront de base au traitement automatique des langues dans la deuxième partie du XX^{ème} siècle. Les principes de la grammaire générative, de la représentation des connaissances et du calcul du sens, fondamentaux pour le traitement automatique des langues trouvent leurs racines chez les grammairiens et les philosophes de l'âge classique. Les chercheurs en traitement automatique des langues, plus souvent informaticiens que philosophes ou historiens des sciences, ont rarement reconnu cette filiation.

5. Notes

¹ Etymologie : le mot **automate** vient du grec *αυτοματον* (« qui se meut soi-même »). Le mot semble attesté depuis 1532 (Rabelais, *Gargantua*).

² Séris (1995), p. 176.

³ Le philosophe John Locke (1632-1704) est l'un des rares à émettre quelques réserves :

Je ne suis pas assez vain pour m'imaginer que qui que ce soit puisse songer à tenter de réformer parfaitement, je ne dis pas toutes les langues du monde, mais même celles de son propre pays, sans se rendre lui-même ridicule. Car exiger que les hommes employassent constamment les mots dans un même sens, et pour n'exprimer que des idées déterminées et uniformes, ce serait se figurer que tous les hommes devraient avoir les mêmes notions, et ne parler que des choses dont ils ont des idées claires et distinctes ; ce que personne ne doit espérer, s'il n'a la vanité de se figurer qu'il pourra engager les hommes à être fort éclairés ou fort taciturnes.

John Locke, *Essai sur l'Entendement Humain* (1690).

⁴ Knecht, H. (1981). *La logique chez Leibnitz*. Lausanne, p. 148. Cité Par Séris (1995) p. 179.

Informatique et linguistique

Cours 3

Historique (2)

... à la réalité

Jean Véronis

1. La naissance de l'ordinateur et les espoirs naïfs

L'erreur de prédiction des années 60 à propos de l'intelligence des machines et de leur capacité de langage est d'une certaine façon compréhensible. La conception du film *2001* se situe temporellement à peu près à mi-parcours entre la naissance de l'informatique (dans les années 40) et l'an 2001. On pouvait donc légitimement extrapoler les progrès fulgurants effectués en un quart de siècle sur le quart de siècle suivant, et prévoir des capacités extraordinaires pour les ordinateurs. L'extrapolation était partiellement juste : la puissance des ordinateurs a bien suivi la courbe prévue, et l'on peut actuellement acheter en supermarché des micro-ordinateurs largement plus puissants que le plus puissant ordinateur des années 60. Toutefois, la simulation de l'intelligence n'a absolument pas suivi la même progression, et les ordinateurs actuels, bien que plus puissants, ne sont strictement pas plus « intelligents » que ceux des années 60. C'est cette difficulté qui n'avait pas été anticipée par les scientifiques de l'époque.

1.1. Naissance de l'ordinateur

Bibliographie

Breton, Philippe (1990). *Une histoire de l'informatique*. Edition du Seuil, Collection Points Sciences.

Webographie

Sur l'ENIAC :
• <http://homepage.seas.upenn.edu/~museum/>

L'ordinateur est né après la deuxième guerre mondiale. La première démonstration publique d'un ordinateur (ou plutôt calculateur, comme on l'appelait à l'époque¹) eut lieu à Philadelphie en 1946. Il s'agissait de l'ENIAC (Electronic Numerical Integrator and Computer). L'ENIAC était un monstre, composé de 18000 lampes à vide, consommant 150 kW et pesant trente tonnes (Figure 1).

C'est en fait la **deuxième guerre mondiale** qui a précipité la naissance de l'informatique, à cause des besoins massifs de calcul en cryptographie, en balistique et en physique nucléaire. L'ENIAC, par exemple, a été financé par le *Ballistic Research Laboratory* sous le nom de « projet PX »². Les progrès ont été rapides et considérables, puisque les calculateurs électroniques n'existaient pas avant la guerre. En 1937-38, Konrad Zuse à Berlin et George Stibitz aux *Bell Telephone Laboratories* à New York en étaient seulement à assembler de façon artisanale des relais de téléphones pour faire des calculs simples. L'un travaillait dans son salon, l'autre dans sa cuisine, et ces projets furent au départ considérés par leur collègues et supérieurs comme de gentils bricolages de week-end.³

C'est l'invention du **transistor** en 1948 qui a constitué une étape décisive vers l'ordinateur moderne, grâce à la miniaturisation qu'il permettait en remplaçant les lampes à vide. Le premier ordinateur transistorisé (l'IBM 650) est produit en 1955. Le reste de l'histoire de l'informatique est une course incessante à la miniaturisation, qui permet un accroissement de la puissance de calcul. L'étape importante suivante a été constituée par l'apparition des **circuits intégrés**, qui donnent naissance à la série 360 d'IBM en 1965 (constituant le fleuron de la technologie informatique au moment de la conception de *2001*), dont les caractéristiques peuvent désormais faire sourire (64 à 256 Ko de mémoire centrale, stockage sur bandes magnétiques, etc.), mais qui a rendu possible le développement massif de l'informatique (banques, administrations, etc.). On était passé en trente ans des quelques relais de

téléphones assemblés par des bricoleurs éclairés, à un raz-de-marée technologique qui annonçait un véritable changement de société.

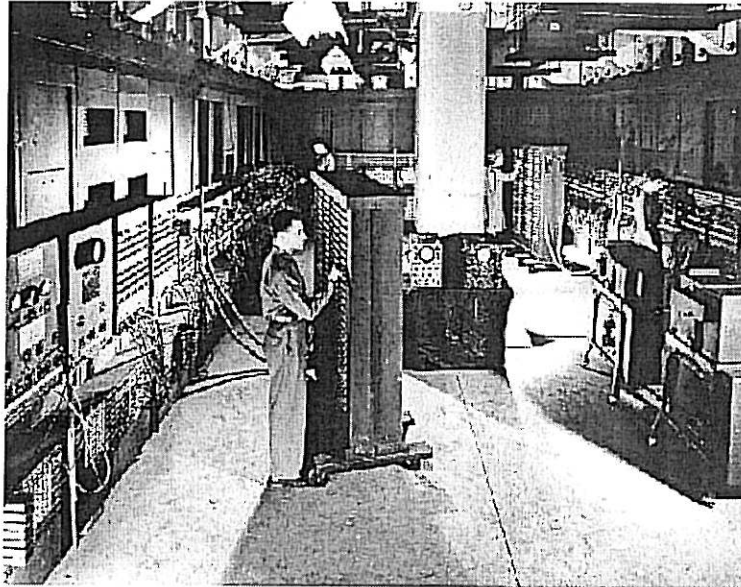


Figure 1. L'ENIAC

1.2. La traduction automatique

Bibliographie

- Arnold, D.J., Balkan, L., Meijer, S., Humphreys, R.L., Sadler, L. (1994). *Machine Translation: an Introductory Guide*, Blackwells-NCC, London. Disponible sur le Web : <http://clwww.essex.ac.uk/~doug/book/>
- Booth, A. and Locke, William N. (1955). *Historical introduction*. In Donald Locke, William N. and Booth, A. Donald (Eds.), *Machine translation of languages*. John Wiley & Sons, New York, 1-14.
- Delavenay, E. (1959). *La machine à traduire*. Presses Universitaires de France, Collection « Que sais-Je ? », Paris.
- Hutchins, W.J., Somers, H.L. (1992). *An Introduction to Machine Translation*. Academic Press, London
- Weaver, Warren (1949). *Translation*. Mimeographed memorandum. Reprinted in Donald Locke, William N. and Booth, A. Donald (Eds.) (1955), *Machine translation of languages*. John Wiley & Sons, New York, 15-23.

Webographie

- Traductions en ligne: <http://babelfish.altavista.com/>

1.2.1. Visions pionnières

Dans ce climat d'euphorie technologique, il était donc normal de placer dans l'informatique naissante les espoirs les plus fous. Le traitement automatique des langues est né en même temps que l'ordinateur : en 1945-46, c'est à dire le moment où naissait le fameux ENIAC, Warren Weaver, de la fondation Rockefeller commençait à penser à l'application des « calculateurs électroniques » à la traduction.⁴ Il écrit à son ami le mathématicien Norbert Wiener⁵ le 6 Mars 1947 pour lui exposer son idée :

One thing I wanted to ask you about is this. A most serious problem, for UNESCO and for the constructive and peaceful future of the planet, is the problem of translation, as it unavoidably affects the communication between peoples. Huxley⁶ has recently told me that they are appalled by the magnitude and the importance of the translation job.

Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (were the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worth while.

Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography — methods which I believe succeed even when one does not know what language had been coded — one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Lettre de Warren Weaver à Norbert Wiener, 4 mars 1947

L'approche est très fortement influencée par le contexte de la deuxième guerre mondiale, où des efforts très importants avaient dû être fournis en **cryptographie**. Weaver a été très intéressé par ce domaine, et il compare explicitement le processus de traduction à un processus de chiffrement : un texte traduit en russe peut être vu comme un chiffrement de sa version anglaise à l'aide d'un code particulier. On retrouve ici, bien que Weaver ne semble pas l'avoir réalisé, le lien qui avait déjà été fait aux XVII^{èmes} et XVIII^{èmes} siècles entre les langues et la cryptographie (voir Cours 2) !

Le choix de la traduction parmi d'autres tâches linguistiques imaginables pour les ordinateurs s'explique aussi par le contexte de l'époque, où la préoccupation était forte d'améliorer la **communication entre les peuples** pour éviter la reproduction de désastres planétaires comme celui de la deuxième guerre mondiale (voir encadré ci-dessus).

1.2.2. Un essor fulgurant

Les recherches en traduction automatique (T.A.) prennent au début des années 50 un essor fulgurant, à grand renfort d'annonces prématurées et exagérées dans les médias, qui promettent la « machine à traduire » :

- Le **premier chercheur** à temps plein en traduction automatique, le logiciel israélien Yehoshua Bar-Hillel, est recruté en 1952 au Massachusetts Institute of Technology (MIT).
- En 1954, des chercheurs de l'Université de Georgetown et d'IBM conduisent la **première expérience** de traduction automatique du russe en anglais, avec un vocabulaire de 250 mots et 6 règles de grammaire.
- La revue *Mechanical Translation* commence à paraître à partir de 1954.
- Le **premier ouvrage** sur la traduction automatique paraît en 1955 (Booth et Locke).

De 1955 à 1960, la recherche s'intensifie :

- De nombreux groupes d'y consacrent à travers le monde (U.R.S.S., Grande-Bretagne, France) .
- La première conférence internationale a lieu en 1956 au MIT.
- En France, une **association** est créée, l'Association pour l'étude et le développement de la Traduction Automatique et de la Linguistique Appliquée (ATALA)⁷.
- La revue *La Traduction Automatique*⁸ est publiée à partir de 1960 par l'ATALA..
- Etc.

1.2.3. Le déclin

Lors de cette période, la plupart des problèmes ont été dépistés, et de multiples solutions ont été proposées. Mais la technologie informatique très sommaire de l'époque était largement insuffisante, de même que les connaissances linguistiques, et une prise de conscience de la **difficulté** de la traduction s'effectue, venant tempérer l'enthousiasme, voire la naïveté, initiaux.

En 1960, **Bar-Hillel** publie un rapport qui met en lumière les difficultés énormes que pose la traduction, tant sur le plan technologique que linguistique. Venant du premier chercheur recruté dans le domaine, ce rapport a un fort impact négatif dans la communauté scientifique. Une de ses principales critiques des recherches en traduction automatique vise leur sous-estimation des connaissances contextuelles et encyclopédiques qui seraient nécessaire pour traduire correctement. Ainsi, pour traduire correctement le mot *pen* dans les deux phrases suivantes :

The box is in the **pen** (la boîte est dans l'enclos)

The **pen** is in the box (le stylo est dans la boîte)

il est nécessaire de disposer de connaissances générales sur le monde (par exemple, une boîte peut contenir un stylo mais un stylo ne peut pas contenir une boîte), dont Bar-Hillel affirme qu'elle sont totalement hors de portée des machines.

En 1964, l'administration américaine commande un rapport, le **rapport ALPAC** (*Automatic Language Processing Advisory Committee*), qui établit un constat d'échec sur les recherches en traduction automatique, et va conduire à l'arrêt des financements et à la disparition quasi totale des recherches dans le domaine.

Pendant la décennie qui suivit, seuls quelques groupes continuèrent les recherches : à Brigham Young University dans l'Utah (grâce à un financement de l'Eglise Mormone intéressée par la traduction de la Bible), au Canada (le groupe TAUM à Montréal), en France (le GETA à Grenoble), en Allemagne (le groupe SUSY à Saarbrücken) et en U.R.S.S. (les groupes dirigés par Mel'cuk et Apresian).

1.2.4. La renaissance

Le retour de balancier créé par le rapport ALPAC a peut-être été trop extrême et injuste. En effet, si les difficultés mentionnées étaient bien réelles, des progrès importants avaient été faits et trois systèmes étaient même opérationnels, même si leurs performances étaient très limitées (Aux Etats-Unis : à la base USAF *Wright Patterson* et au *Oak Ridge Laboratory* de

l'Atomic Energy Commission, et en Italie au centre EURATOM d'Ispra). Les quelques groupes qui ont continué les recherches (mentionnés plus haut) ont été à l'origine de systèmes commerciaux (WEIDNER et ALPS qui dérivent des travaux de l'Université Brigham Young) ou du système METEO (créé par le groupe TAUM de Montréal) en usage quotidien au Canada pour la traduction de bulletins météorologiques.

Ce n'est qu'en 1975, sous l'impulsion de la Communauté Européenne que les recherches en traduction automatique prendront un nouvel essor. Confrontée à l'explosion des besoins en traduction, la Communauté engage en effet le développement du système SYSTRAN, qui est toujours opérationnel et offre un service de traduction automatique gratuit sur le Web, que nous utiliserons dans le cadre des TD-TP⁹.

Curieusement, les critiques de Bar-Hillel sur la représentation des connaissances et la capacité de raisonnement des machines avaient commencé à recevoir des embryons de solution, tant dans le cadre de la traduction automatique elle-même (par exemple dans le groupe de Margaret Masterman à Cambridge qui avait implémenté la première base de connaissances en machine à partir du *Roget's thesaurus*, et avait montré son utilité dans la traduction des *Georgiques* de Virgile), ou dans le cadre d'une discipline née en même temps que la traduction automatique au début des années cinquante, l'Intelligence Artificielle (I.A.) et qui semblait à l'époque couronnée de succès. En fait, le traitement automatique des langues au cours des années soixante a été profondément marqué par ce courant, comme nous allons le voir ci-après.

1.3. L'intelligence artificielle

Bibliographie

- Dreyfus, H.L., Dreyfus, S.E. (1986). *Mind over Machine*. Free Press, New York.
- Hodges, Andrew (1991). *Alan Turing: the Enigma*, Vintage, Random House, London.
Traduction française (abrégée) par N. Zimmermann: *Alan Turing ou l'énigme de l'intelligence*, Payot, Paris.
- Turing, Alan M. (1950). Computing machinery and intelligence. *Mind*, 59, 236. Texte disponible sur le web : <http://www.sscf.ucsb.edu/~sung/comm115/writing-define-computing/Computing-machinery.html>. Traduction française et commentaire dans Pélissier, A., Tête, A. (1995). *Sciences cognitives : textes fondateurs (1943-1950)*. Presses Universitaires de France, Paris, 247-295.
- Dennett, Daniel (1998). *Brainchildren: Essays on Designing Minds*. MIT Press, 1998.
- Weizenbaum, Joseph (1966). ELIZA : a computer program for the study of natural language communication between man and machine. *Communications of the A.C.M.*, 9, 26-45.
- Weizenbaum, Joseph (1976). *Computer Power and Human Reason – From Judgment to Calculation*. W. H. Freeman and Company, New York.
- Winograd, Terry (1972). *Understanding Natural Language*. Academic Press, Edinburgh.

Webographie

Sur Alan Turing :

- <http://www.turing.gla.ac.uk/turing/turing.htm>
- <http://www.turing.org.uk/turing/>

Sur le Test de Turing :

- <http://www.wadham.ox.ac.uk/~ahodges/scrapetest.html>
- <http://www.sscf.ucsb.edu/~sung/comm115/writing-define-computing/Computing-machinery.html>

ELIZA et autres systèmes de dialogue :

- ELIZA : <http://www-ai.ijs.si/eliza/eliza.html>

- On-line psychiatrist : <http://www.parnasse.com/drwww.shtml>
- Le Prix Loebner : <http://info.acm.org/~loebner/loebner-prize.html>

Pages personnelles :

- McCarthy : http://soe.stanford.edu/compsci/faculty/McCarthy_John.html
- Minsky : <http://www.media.mit.edu/people/minsky/minsky.html>
- Winograd : <http://www.pcd.stanford.edu/~winograd/>
- Dreyfus : <http://socrates.berkeley.edu/~frege/dreyfus/>

1.3.1. L'ordinateur modèle du cerveau humain

Une sous-estimation des difficultés semble avoir été caractéristique des débuts de l'informatique. Divers chercheurs voyaient dans l'ordinateur un outil permettant de **modéliser le cerveau humain**¹⁰. Alan Turing (Figure 2), l'un des pères fondateurs de l'informatique pariait en 1950 que les machines pourraient **penser** en l'an 2000, ou du moins qu'on ne pourrait pas faire la différence entre une machine et un être humain dans une conversation.



Figure 2. Alan Turing

Cet argument a bien sûr influencé Clarke et Kubrick pour la conception de HAL. Arthur C. Clarke s'y réfère explicitement dans le roman associé au film (voir encadré).

Quant à savoir si [HAL] était réellement doué de pensée, la question avait été résolue dans les années 40 par le mathématicien britannique Alan Turing. Turing avait déclaré que si un homme était capable de converser longuement avec un ordinateur — peu importe que ce fût par l'intermédiaire d'un clavier ou d'un micro — sans distinguer de réelle différence entre celles que tout homme aurait pu donner, alors cet ordinateur *pensait* vraiment, selon l'exacte définition du terme. [HAL] eût passé facilement le test de Turing.

Arthur C. Clarke, 2001, *L'odyssée de l'espace*

Le test de Turing est basé sur le jeu de l'imitation, dans lequel un interrogateur humain doit deviner le sexe de deux personnes : un homme (A) et une femme (B) avec lesquels il dialogue par écrit sans les voir (par exemple par l'intermédiaire d'un télécscripteur). L'homme (A) peut mentir, mais pas la femme (B) (sinon le jeu est impossible). Turing imagine de remplacer A par

un ordinateur : si l'interrogateur ne se trompe pas plus souvent que dans la situation où le jeu comporte un homme et une femme, il ne peut donc faire la différence entre l'homme et la machine. Pour Turing, on peut conclure que la machine « pense ». Turing parie que cinquante ans plus tard, donc vers l'an 2000, les machines auront cette capacité.

Voici le fragment du texte décrivant le test :

The Imitation Game. I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, if the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the "imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be:

"My hair is shingled, and the longest strands are about nine inches long."

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.

We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"

[...]

I believe that in about fifty years' time it will be possible, to programme computers, with a storage capacity of about 10^9 , to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning.

Alan Turing, "Computing machinery and intelligence", *Mind* (1950)

La forme sous laquelle le test est décrit, faisant intervenir un homme et une femme, a été jugée par la plupart des commentateurs comme inutilement compliquée, et il est généralement présenté sous une forme simplifiée, dans laquelle un interrogateur dialogue par téléscripneur avec un interlocuteur unique, dont il ne sait pas s'il est un humain ou une machine.

Curieusement, Turing ne fait pas référence à **Descartes**. Or Descartes avait proposé le même test trois siècles auparavant dans le *Discours de la Méthode*, lui aussi faisant de la parole le critère de la pensée — avec une conclusion sans doute différente :

Ce qui ne semblera nullement étrange à ceux qui, sachant combien de divers *automates* ou machines mouvantes l'industrie des hommes peut faire sans y employer que fort peu de pièces, à comparaison de la grande multitude des os, des muscles, des nerfs, des artères, des veines, et de toutes les autres parties qui sont dans le corps de chaque animal, considérons ce corps comme une machine, qui, ayant été faite des mains de Dieu, est incomparablement mieux ordonnée et a en soi des mouvements plus admirables qu'aucune de celles qui peuvent être inventées par les hommes. Et je m'étais ici particulièrement arrêté à faire voir que s'il y avait de telles machines qui eussent les organes et la figure extérieure d'un singe ou de quelque autre animal sans raison, nous n'aurions aucun moyen pour reconnaître qu'elles ne seraient pas en tout de même nature que ces animaux; au lieu que s'il y en avait qui eussent la ressemblance de nos corps, et imitassent autant nos actions que moralement il serait possible, nous aurions toujours deux moyens très certains pour reconnaître qu'elles ne seraient point pour cela de vrais hommes : dont le premier est que jamais elles ne pourraient user de paroles ni d'autres signes en les composant, comme nous faisons pour déclarer aux autres nos pensées : car on peut bien concevoir qu'une machine soit tellement faite qu'elle profère des paroles, et même qu'elle en profère quelques unes à propos des actions corporelles qui causeront quelque changement en ses organes, comme, si on la touche en quelque endroit, qu'elle demande ce qu'on lui veut dire; si en un autre, qu'elle crie qu'on lui fait mal, et choses semblables; mais non pas qu'elle les arrange diversement pour répondre au sens de tout ce qui se dira en sa présence, ainsi que les hommes les plus hébétés peuvent faire. Et le second est que, bien qu'elles fissent plusieurs choses aussi bien ou peut-être mieux qu'aucun de nous, elles manqueraient infailliblement en quelques autres, par lesquelles on découvrirait qu'elles n'agiraient pas par connaissance, mais seulement par la disposition de leurs organes : car, au lieu que la raison est un instrument universel qui peut servir en toutes sortes de rencontres, ces organes ont besoin de quelque particulière disposition pour chaque action particulière; d'où vient qu'il est moralement impossible qu'il y en ait assez de divers en une machine pour la faire agir en toutes les occurrences de la vie de même façon que notre raison nous fait agir.

Réné Descartes, *Discours de la méthode*, Partie V (1636).

L'article de Turing a suscité de nombreux débats sur la nature de l'intelligence et de la pensée parmi les philosophes et spécialistes des sciences cognitives. Quoi qu'il en soit, la prédiction de Turing n'a pas été réalisée, malgré les **débuts prometteurs de l'Intelligence Artificielle**.

1.3.2. La naissance de l'Intelligence Artificielle

L'année où paraît l'article de Turing, **Claude Shannon** (fondateur de la théorie de l'information) publie un article intitulé *A Chess-Playing Machine* dans *Scientific American*¹¹, qu'il commence par la phrase « On peut concevoir des calculateurs électroniques qui jouent à un jeu très difficile, soulevant ainsi la question de savoir s'ils peuvent "penser" ». En 1953, Shannon embauche deux jeunes assistants aux *Bell Labs*, **John McCarthy** et **Marvin Minsky**¹², qui allaient devenir des figures-clés de l'I.A. Rapidement, McCarthy est recruté au Dartmouth College, Minsky au M.I.T., et en 1956 ils organisent la célèbre **conférence de Dartmouth**, qui consacra l'I.A. comme domaine de recherche autonome. C'est d'ailleurs à cette occasion que le terme *intelligence artificielle* est utilisé la première fois¹³. C'est aussi en 1956 que les premiers langages de programmation « symbolique » (IPL et LISP) apparaissent¹⁴. Des expériences d'I.A. fleurissent dans tous les domaines : simulations de jeux, manipulation d'images, démonstration logique. A la fin des années 50, un programme joue très honnêtement aux dames, d'autres démontrent des théorèmes de logique¹⁵ ou de géométrie plane. Les progrès continuent régulièrement au début des années 60.

Cette période euphorique de l'I.A. contraste curieusement avec la période dépressive de la traduction automatique au début des années 60, reflétée par les conclusions de Bar-Hillel et le rapport ALPAC (voir 1.2.3). Pourtant, **Bar-Hillel** et **Minsky** travaillaient dans la même

institution, le M.I.T. où travaillait également le jeune linguiste **Noam Chomsky**, qui publiait à l'époque des travaux sur les automates et le langage (sur lesquels nous reviendrons abondamment). Il est surprenant de constater que parmi les spécialistes pressentis pour participer à la conférences, ne figurent aucun des spécialistes de traduction automatique de l'époque. N'y figurent en particulier ni Bar-Hillel ni Noam Chomsky, en quelque sorte voisins de palier de Minsky¹⁶. Pourtant, l'un des thèmes proposés pour la conférence figurait celui-ci : « How can a computer be programmed to use a language ?¹⁷ ». Dans les publications de l'époque, il est presque impossible de trouver des références aux travaux de l'I.A. dans les publications de traduction automatique, et vice-versa. Les historiens auront à se pencher sur ce curieux cloisonnement.

En tous cas, en 1963, c'est-à-dire à peu près au moment où le rapport ALPAC allait marquer un coup d'arrêt des financements de l'administration américaine pour la T.A., le M.I.T. reçoit un financement de 2,2 millions de dollars de l'*Advanced Research Projects Agency (ARPA)* du Département de la Défense pour les recherches en intelligence artificielle, destinés à assurer que les Etats-Unis resteraient en tête de l'U.R.S.S. dans les technologies de pointe. De fait, la décennie allait être celle de l'I.A., et les recherches sur le traitement des langues se sont transférées sous cette nouvelle bannière.

1.3.3. **L'Intelligence Artificielle et le langage**

Les chercheurs du M.I.T. dirigés par Minsky s'attaquent à divers problèmes, mais le traitement du langage y a une place de choix. En particulier, Ross Quillian¹⁸ développe les réseaux sémantiques, qui vont permettre de représenter des connaissances simples et de faire des inférences. Comme ceux de Masterman et Richens, ces travaux donnent tort à Bar-Hillel : des systèmes de **représentation des connaissances** apparaissent, même s'ils sont pour l'instant sommaires.

De nombreux systèmes de dialogue ou de compréhension du langage fleurissent au cours des années 60, au M.I.T. et dans de nombreux autres centres de recherche. Il est impossible de les citer tous, et nous prendrons seulement deux des exemples les plus connus.

En 1966, Joseph Weizenbaum développe le programme **ELIZA**¹⁹, qui simule un dialogue avec un psychanalyste. ELIZA était basé sur une technologie extrêmement simple : le programme détectait des **mots-clés** dans la phrase tapée, et répondait selon des modèles pré-établis, en reprenant certaines parties de la phrase tapée en changeant les pronoms de façon adéquate :

Young lady: Men are all alike.

Computer: IN WHAT WAY

Young lady: They're always bugging us about something or other.

Computer: CAN YOU THINK OF A SPECIFIC EXAMPLE

Young lady: Well, my boyfriend made me come here.

Computer: YOUR BOYFRIEND MADE YOU COME HERE

Young lady: He says that I am depressed all the time.

Computer: I AM SORRY TO HEAR YOU ARE DEPRESSED

Young lady: It's true that I am unhappy

Computer: DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY

Young lady: I need help, that seems certain.

Computer: WHAT WOULD IT MEAN IF YOU GOT SOME HELP

Young lady: Perhaps I could learn to get along with my mother.

Computer: TELL ME MORE ABOUT YOUR FAMILY

Young lady: My mother takes care of me.

Computer: WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU

Young lady: My father.

Computer: YOUR FATHER

Computer: WHAT RESEMBLANCE DO YOU SEE
 Young lady: You are not very aggressive but I think you don't want me to notice that.
 Computer: WHAT MAKES YOU THINK THAT I AM NOT VERY AGGRESSIVE
 Young lady: You don't argue with me.
 Computer: WHY DON'T YOU THINK I ARGUE WITH YOU
 Young lady: You are afraid of me.
 Computer: DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU
 Young lady: My father is afraid of everybody.
 Computer: WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER
 Young lady: Bullies.
 Computer: DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE

Un exemple de dialogue avec ELIZA (Weizenbaum, 1966).

Le programme ne faisait évidemment aucune tentative pour « comprendre » en profondeur les phrases de l'interlocuteur humain, et il ne pouvait donner une illusion de réalisme que grâce à la situation particulière du dialogue, celui d'une conversation avec un psychanalyste, dont l'aspect décousu est acceptable. Malgré son principe simpliste, le caractère spectaculaire d'ELIZA a largement contribué au développement des recherches sur le traitement du langage.

En 1970, Terry Winograd développait un système appelé **SHRDLU**²⁰, qui constituait une étape importante dans les systèmes de dialogue, dans la mesure où la machine « comprenait » réellement les ordres qui lui étaient donnés, dans le cadre d'un **micro-monde** extrêmement restreint composé de blocs de diverses formes et diverses couleurs que la machine devait déplacer (Figure 3).

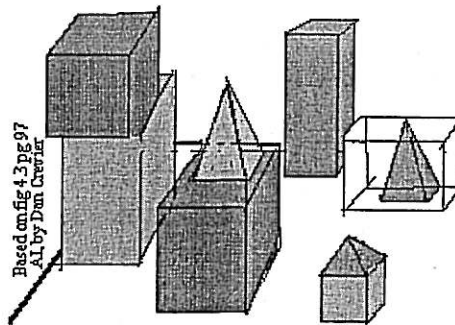


Figure 3. Le monde des blocs de SHRDLU

SHRDLU était capable de réaliser les différentes actions demandées et de répondre à des questions sur la situation spatiale à chaque instant :

Man : 1. pick up a red block
 Computer : OK
 Man : 2. find a block which is taller than the one you are holding and put it into the box.
 Computer : By "it," I assume that you mean the block which is taller than the one I am holding. OK.
 Man : 3. what does the box contain?

Man : 4. what is the pyramid supported by?
 Computer : the box
 Man : 5. how many blocks are not in the box?
 Computer : four of them
 Man : 6. is at least one of them narrowed than the one which I told you to pick up?
 Computer : yes, the red cube
 Man : 7. is it supported?
 Computer : yes, by the table
 Man : 8. can the table hold blocks?
 Computer : no
 Man : 9. the blue pyramid is mine
 Computer : I understand
 Man : 10. I own all blocks which are not red, but I don't own anything that supports a pyramid
 Computer : I understand
 Man : 11. do I own the box?
 Computer : no
 Man : 12. do I own anything in the box?
 Computer : Yes, two things: the blue box and the blue pyramid
 Man : 13. will you please stack up both of the red blocks and either a green cube or a pyramid?
 Computer : OK
 Man : 14. is there a large block behind a pyramid?
 Computer : yes, three of them: a large red one, a large green cube and the blue one.
 Man : 15. put a small one into the green cube which supports a pyramid
 Computer : OK
 Man : 16. put the littlest pyramid on top of it
 Computer : OK
 Man : 17. had you touched any pyramid before you put the green one on the little cube?
 Computer : Yes, the green one
 Man : 18. when did you pick it up?
 Computer : while I was stacking up the red cube, a large red block and a large green cube
 Man : 19. why?
 Computer : to get rid of it
 Man : 20. why did you do that?
 Computer : to clear off the red cube
 Man : 21. why did you clear off that cube?
 Computer : to put it on a large green cube
 Man : 22. why did you do that?
 Computer : to stack the red cube, a large block, and a large green cube
 Man : 23. why did you do that?
 Computer : because you asked me to

Un exemple de dialogue avec SHRLDU (Winograd, 1972).

1.3.4. Les critiques et le déclin

Malgré l'enthousiasme initial suscité par les succès parfois spectaculaires de l'I.A., des critiques ont commencé à s'élever à partir des années 70. Deux ouvrages, en particulier, par Weizenbaum (1976) lui-même, et Dreyfus et Dreyfus (1986), montrèrent les difficultés, et le chemin qui restait à parcourir entre les succès initiaux et les promesses parfois imprudentes. Ces publications eurent un effet analogue, bien que plus atténué, à celui du rapport ALPAC sur la T.A. et amorcèrent un déclin de l'intérêt pour l'I.A. à la fin des années 80.

Toutefois, les méthodes développées par l'I.A. ont eu une influence profonde sur l'évolution de l'informatique, et il est peut-être injuste de dire que l'I.A. a décliné. Elle s'est plutôt amalgamée dans l'informatique générale : bien des techniques nées des recherches en I.A. (exploration d'espaces de recherche, heuristiques, réseaux de neurones, etc.) font désormais partie des

algorithmes de base, et se retrouvent incorporées dans des logiciels qui ne font plus référence à la simulation de l'intelligence humaine, qui était le but de l'I.A. :

Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems.

Site Web *Whatis* (1998) <http://whatis.com/>

Artificial intelligence (AI): The capability of a device to perform functions that are normally associated with human intelligence, such as reasoning and optimization through experience.

Federal Standard 1037C, Glossary of Telecommunication Terms, General Services Administration, Washington, DC., 1996. <http://www.its.bldrdoc.gov/fs-1037/fs-1037c.htm>

2. La linguistique computationnelle

2.1. La phase d'expérimentation

Au cours des années 60, le traitement automatique des langues a commencé à s'ériger en discipline autonome, sous le nom de **linguistique computationnelle**, se démarquant de la traduction automatique. Un certain nombre d'événements marquent cette naissance :

- En 1962, création d'une **association**, l'*Association for Computational Linguistics*²¹, qui tient une conférence annuelle (37^{ème} édition en 1999²²).
- En mai 1965, à New York : première **conférence** internationale de linguistique computationnelle, qui a eu des éditions biannuelles depuis, sous le nom de **Coling** (18^{ème} édition en 2000²³).
- En 1975, création de la revue *Computational Linguistics*²⁴.
- Des **cours** de Linguistique Computationnelle se développent dans de nombreuses universités.

Des recherches spécifiques se développent dans diverses directions, morphologie, syntaxe, sémantique, et vers de multiples applications :

- interrogation de bases de données
- dialogue homme-machine
- recherche documentaire
- génération de textes
- correction orthographique
- traduction automatique (qui voit un renouveau à partir de la fin des années 70)
- traitement de la parole
- etc.

La synergie avec l'I.A. a été importante, pour tout ce qui concerne les aspects touchant au dialogue, à la compréhension, à la représentation des connaissances. Dans la suite du cours nous examinerons les principales techniques et méthodes qui ont émergé de cette période, que l'on pourrait qualifier de période d'expérimentation. Les applications concrètes, bien qu'existantes, sont restées relativement confidentielles jusqu'aux années 90, restreintes à une utilisation dans les centres de recherche ou sous forme de logiciels « sur mesure » dans l'industrie.

2.2. L'ère des applications effectives

Au cours des années 90, le traitement automatique des langues (T.A.L.) a été marqué par un tournant important. Si les problèmes sont loin d'être résolus (on ne dispose d'aucune machine ayant des capacités approchantes de celles de HAL) et si la recherche en linguistique computationnelle continue d'explorer tous les aspects du traitement des langues, de multiples applications ont franchi le pas des **logiciels grand public**.

- **correction orthographique** : tous les systèmes de traitement de texte ont des logiciels de correction orthographique intégrés, dont l'efficacité est devenue plus que satisfaisante ;
- **synthèse de la parole** : les systèmes de synthèse de la parole ont franchi un seuil de qualité et d'intelligibilité important, qui permet de les utiliser dans de nombreuses applications (systèmes téléphoniques, aide aux handicapés, etc.) ;

- **reconnaissance de la parole** : la reconnaissance de mots isolés est devenue d'usage courant (par exemple composition de numéros de téléphone à partir d'un nom prononcé) ; la reconnaissance de la parole continue a fait des progrès considérables, et l'on trouve désormais des logiciels grand public de dictée vocale donnant des résultats très satisfaisants pour un prix modique.

Cette transition marque une certaine maturité du domaine : pour passer des prototypes de laboratoire ou des applications industrielles confidentielles à des systèmes grand public, la technologie doit être robuste et éprouvée.

D'autres applications grand public sont en train d'émerger, en particulier sous l'influence de l'explosion du **World Wide Web**, qui pose des problèmes d'accès à l'information, et de multilinguisme :

- **traduction automatique** : on trouve divers systèmes de traduction sur ordinateur personnel, et même en libre accès sur le Web²⁵. La qualité de ces logiciels est loin d'être excellente, mais il semblerait qu'il offrent des traductions grossières qui peuvent aider certains utilisateurs ;
- **recherche documentaire** : divers moteurs de recherche (*Altavista*²⁶, etc.) permettent de localiser les pages d'information sur des sujets donnés sur le Web, et ils commencent à intégrer des capacités linguistiques minimales (compréhension des formes fléchies des mots dans les différentes langues, détection des expressions composées, thesaurus et réseaux sémantiques, etc.).

Etant donné le développement rapide du Web, il est probable que ces systèmes vont faire l'objet d'améliorations importantes dans les années à venir.

3. Notes

¹ Le terme **ordinateur** est un anachronisme. On parlait à l'époque de **calculateur** ou **calculatrice électronique**. Le mot « ordinateur » a été forgé par Jacques Perret, professeur à la Sorbonne à la demande de la compagnie I.B.M.-France, dans les années cinquante (voir le « Que Sais-Je ? » 832, *Les Ordinateurs Electroniques*, par Pierre Demarne et Max Rouquerol, de 1959 ; le mot Ordinateur porte d'ailleurs une majuscule dans tout l'ouvrage). Le mot **informatique**, quant à lui, a été créé en 1962 par Philippe Dreyfus à partir des mots « information » et « automatique », signifiant ainsi que les ordinateurs ne sont pas seulement des machines destinées au calcul, mais d'une façon générale au traitement automatique de l'information. On notera que ces termes n'existent pas en anglais, où l'on parle de **computer** (calculateur) et de « computer science » (science des calculateurs).

² Lors de la première présentation, en 1946, l'ENIAC calcula en 20 secondes devant des journalistes la trajectoire d'un projectile, que des êtres humains auraient mis trois jours à calculer.

³ A peu près à la même époque, le Britannique Alan Turing et l'Américain Claude Shannon jetaient les bases logiques et mathématiques de ce qui allait devenir le calcul électronique.

⁴ Pour être tout à fait exact, des travaux précurseurs ont été menés par le Russe Srimov-Trojanskij, qui a présenté une « machine à traduire » en 1933 à Moscou. Celle-ci n'a pas retenu l'intérêt des scientifiques russes à l'époque.

⁵ Wiener devint célèbre peu de temps après pour ses travaux sur la cybernétique. Weaver et Wiener s'étaient rencontrés pendant la guerre, le premier menant des recherches en balistique, le second sur les radars et la théorie de la prédiction. [Weaver, W. (1972) *Scene of Change: A Lifetime in American Science*, New York: Scribner]

⁶ Julian Huxley, premier directeur de l'UNESCO (frère de l'écrivain Aldous Huxley).

⁷ Cette association est toujours active, sous le nom d'*Association pour le Traitement Automatique des Langues*. voir <http://www.linguist.jussieu.fr/~atala/>

⁸ Devenue *Traitement Automatique des Langues (TAL)*. Voir <http://www.linguist.jussieu.fr/~atala/tal/tal.html>

⁹ <http://babelfish.altavista.com/>

¹⁰ Il ne s'agissait pas d'élucubrations de savants marginaux : les pères fondateurs mêmes de l'informatique, von Neumann, Shannon, Turing et Wieve, poursuivaient explicitement le but de construire des machines qui simulent le cerveau humain. L'architecture de von Neuman, base des ordinateurs jusqu'à nos jours, a été guidée par son désir de construire un « cerveau artificiel » (voir von Neuman, J., 1958, *The Computer and the Brain*, Yale University Press, New Haven), et constitue en quelque sorte une mixture de la « machine de Turing » (modèle abstrait de calculateur universel proposé par Turing en 1936) et des travaux de Warren McCulloch et Pitts sur les neurones artificiels. Dans un article intitulé *le premier ordinateur copiait le cerveau humain* (1996), Philippe Breton écrit qu'on pourrait presque dire « que l'informatique découle de l'intelligence artificielle plutôt que l'inverse ».

¹¹ vol. CLXXXII, p. 48-51. Traduction française et commentaire dans Pélissier, A., Tête, A. (1995). *Sciences cognitives : textes fondateurs (1943-1950)*. Presses Universitaires de France, Paris, 231-245.

¹² Minsky, rappelons-le, fut conseiller scientifique pour le film *2001*.

¹³ D'après John McCarthy, dans le document proposant l'organisation de la conférence : McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*. Source : site Web de John McCarthy : <http://www-formal.stanford.edu/jmc/history/dartmouth.html>

¹⁴ Il est remarquable que les langages de base de l'intelligence artificielle aient été parmi les premiers jamais inventés, immédiatement après FORTRAN.

¹⁵ Le programme de démonstration logique d'Allen Newell et Herbert Simon, *Logic Theory Machine*, est souvent considéré comme le premier programme d'I.A. Il était capable de démontrer des théorèmes de logique du premier ordre, et trouva même une preuve particulièrement élégante du théorème 2.85 des célèbres *Principia Mathematica* de Whitehead et Russell, là où les auteurs n'avaient pu élaborer qu'une démonstration compliquée. Lord Russell fut impressionné et écrivit à Simon en Novembre 1956 : « I am delighted to know that *Principia Mathematica* can now be done by machinery... I am quite willing to believe that everything in deductive logic can be done by machinery ».

¹⁶ Également étrange : Norbert Wiener, avec qui Weaver semble avoir discuté pour la première fois de traduction automatique était également au M.I.T, et Weaver lui-même dirigeait une division de la fondation Rockefeller qui finançait la conférence de Dartmouth.

¹⁷ Extrait :

"2. How Can a Computer be Programmed to Use a Language. It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new word and some rules whereby sentences containing it imply and are implied by others. This idea has never been very precisely formulated nor have examples been worked out."

¹⁸ Quillian, M. Ross (1968). Semantic memory. In Minsky, M. (Ed.), *Semantic Information Processing*, MIT Press, 227-270.

¹⁹ On notera le clin d'œil à la pièce de George Bernard Shaw.

²⁰ Le nom SHRDLU est emprunté au magazine MAD, et est composé de la septième à la douzième lettre par ordre de fréquence en anglais.

²¹ Site: <http://www.aclweb.org/>

²² Site: <http://www.aclweb.org/acl99>

²³ Site : <http://www.coling.org>

²⁴ Site: <http://mitpress.mit.edu/journal-home.tcl?issn=08912017>

²⁵ voir par exemple: <http://babelfish.altavista.com/>

²⁶ Voir par exemple: <http://www.altavista.com/>

VOUZZAVEDIBISAR

L'électeur Gaullicoque

Edmond Bianco

Loi de 2004

La Gaullicoquie est un pays qui ronronne. Il y court beaucoup d'idées de gauche, beaucoup plus, en tous cas que dans la plupart des pays industrialisés, mais c'est une politique de droite qui y règne traditionnellement. Allez comprendre des choses pareilles. L'électeur Gaullicoque fonctionne sur contraste, et cela donne des choses savoureuses. Lorsque Savate Premier s'est trouvé en concurrence avec l'Homme de l'extrême droite, ce sacré électeur en a fait un défenseur des idées de gauche. Très drôle ! Depuis, notre Savate National n'arrête plus d'arroser ce succès en ripailles continuelles. On le comprend, car il s'est retrouvé avec un score digne d'un Tsar de toutes les Russies. Tout ça pour dire que finalement, la Gaullicoquie est un pays extrêmement conservateur.

