

Introduction

Damien Nouvel



Plan

1. Généralités
2. Langues et langages
3. Le principe automatique
4. Les automates pour traiter le langage

Séances et modalités de contrôle

- ▶ **Un semestre**
 - Langages réguliers
 - Grammaires hors-contexte
- ▶ **Séances**
 - 12 séances
 - Chaque séance : **cours** et **exercices**
- ▶ **Modalités de contrôle** (semestre)
 - Un **contrôle** : 50%
 - Un **examen final** : 50%

Contenu du cours (deux semestres)

► **Progression**

- Introduction
- Théorie des langages
- Automates à états finis
- Langages hors-contexte
- Analyseurs syntaxiques

Plan

1. Généralités
2. Langues et langages
3. Le principe automatique
4. Les automates pour traiter le langage

Considérations linguistiques



Tablette d'Uruk (-3400)

- ▶ Moyen de **communication** (oral, écrit, signé, etc.)
- ▶ Classification : groupes, langues, dialectes

- ▶ Naturelles (biolangues) vs construites (conlangs, idéolangues)

⇒ <https://www.youtube.com/watch?v=rupVq4m8a8g>

- ▶ **Structure** : discours, propositions, mots, syllabes, lettres
- ▶ Double articulation (Martinet)
 - **Phonèmes** : unités minimales de prononciation
 - **Morphèmes** ou **monèmes** : unités minimales de sens

⇒ Comment **décomposer** le langage ?

Communiquer par le langage

- ▶ Entités en jeu
 - **Émetteur** (locuteur) du message (gestes, cordes vocale, stylo, clavier, braille)
 - ⇒ Acte de transmettre un concept abstrait
 - **Récepteur** du message (vision, ouïe, toucher, écrans)
 - ⇒ Soumis (activement ou passivement) au message
 - **Canal** (média) pour transporter le message
 - ⇒ Possibilité d'atténuation, de dégradation, de bruit
 - ⇒ Air, papier, sémaphore, vinyl, transistor, onde ...
- ▶ Le **message** est une **séquence de symboles**
 - ⇒ Nombre **fini** de symboles possibles
 - ⇒ Nombre **infini** de messages par combinaisons
 - ⇒ Linéarité de la représentation

Transmettre le langage

▶ Conditions

- L'émetteur **encode** (correctement) le message
- Le canal **transporte** le message sans dégradation
- Le récepteur **décode** (correctement) le message

▶ **Format** pour le message

- Implicite : usage, conventions sociales
- Explicite : règles, normes

⇒ Qu'est-ce qu'un message **acceptable** ?

⇒ Comment vérifier qu'un message est au bon format ?

La langue comme système de symboles

- ▶ **Système** : « totalité organisée, faite d'éléments solidaires ne pouvant être définis que les uns par rapport aux autres en fonction de leur place dans cette totalité » (Saussure, 1931)
- ▶ Élément (ou **symbole**) : unité minimale **signifiante**
 - Mot, syllabe, sigle, lettre, graphème
 - ⇒ Syllabe → mot : dans « avion », « av » a-t-il un sens ?
 - ⇒ Mot → expression : dans « cordon bleu », « cordon » a-t-il un sens ?
- ▶ Place (ou **combinaison**)
 - Séquentiellement : **concaténation**
 - ⇒ Le **sens** dépend des mots concaténés ...

Définitions du langage

- ▶ « Ensemble de signes utilisés par une communauté pour communiquer » (Saussure, 1913)
- ▶ « Système de signes identifiés permettant une communication entre une ou plusieurs entités » (Wikipedia, 2011)
- ▶ « Le langage est la capacité d'exprimer une pensée et de communiquer au moyen d'un système de signes [...] doté d'une sémantique, et le plus souvent d'une syntaxe [...] » (Wikipedia, 2019)
- ▶ Hiérarchie des langages formels (Chomsky, 1956)
 - Réguliers (automates)
 - Hors-contexte (grammaires)
 - Contextuel
 - Général

Plan

1. Généralités
2. Langues et langages
3. Le principe automatique
4. Les automates pour traiter le langage

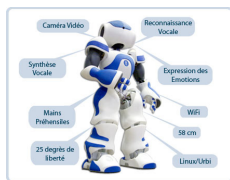
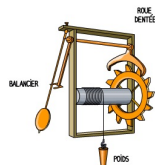
Automate, mécanique



Automate Jaquet-Droz (1770)

Qu'est-ce qu'un automate

- ▶ Du grec « automaton » : qui se meut par soi-même, qui imite les mouvements
- ▶ Pas nécessairement un « robot »



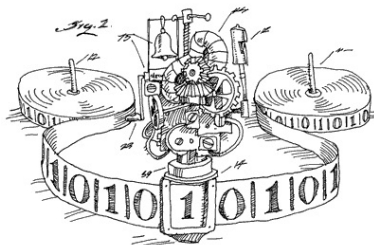
Pendule et robot Nao

- ▶ Horloges (XIII^{ème}) et pendules
- ▶ Pascaline (XVII^{ème})
- ▶ Boîtes à musique (XIX^{ème}) et orgues de barbarie
- ▶ Ordinateurs (XX^{ème})
- ▶ Nao, Pepper (2010)

⇒ Pas d'intervention humaine, mécanisme indépendant

Théorie d'un automate

- ▶ Éléments conceptuels fondamentaux
 - États
 - Transitions
- ⇒ Représentation graphique (diagramme)



Machine de Turing (1936)

Plan

1. Généralités
2. Langues et langages
3. Le principe automatique
4. Les automates pour traiter le langage

Théorie d'un automate

- ▶ « Actions » de l'automate
 - Chemin dans le diagramme
 - Séquence de transitions (symboles)

⇒ Langage « reconnu » par l'automate

- ▶ Langage de programmation
 - Emprunts aux langages naturels (sémantique)
 - Constructions artificielles (syntaxe)
 - De plus ou moins haut niveau (assembleur vs SPARQL)

Automatiser le langage

- ▶ Reconnaître un langage
 - Mots du langage (symboles élémentaires)
 - Arrangements possibles des mots (constructions)
- ⇒ Bon ajustement des contraintes (genre, flexions, etc.)
- ▶ État de l'automate : abstraction des matérialisations
- ▶ Bonnes propriétés des langages (informatiques)
 - **Efficacité** : rapidité des traitements
 - **Expressivité** : capacité à formuler des « choses »
 - **Précision** : degré de contrôle sur ce que fait la machine
- ▶ Principe **déterministe** : pas d'aléatoire

Quelques références

► Sites web

- Alexis Nasr <http://pageperso.lif.univ-mrs.fr/~alexis.nasr/Ens/THL/THL.html>
- François Yvon <https://perso.limsi.fr/yvon/classes/thl>