

**Communication Langagière**  
Ingénierie des langues et de la parole



ANDON TCHECHMEDJIEV [ANDON.TCHECHMEDJIEV@IMAG.FR](mailto:ANDON.TCHECHMEDJIEV@IMAG.FR)  
(DIAPOS. DE LAURENT BESACIER)

**COURS. MARDI, 10H15 -11H45**  
**TPS. MERCREDI, 8H -11H15**

---

---

---

---

---

---

---

---

**Communication Langagière**  
Ingénierie des langues et de la parole



**1. Introduction générale**

**2. Ingénierie des langues**

- ☞ 2.1. Représentation et codage des textes
- ☞ 2.2. Théorie de l'information et probabilités
- ☞ 2.3. Quelques applications du traitement automatique du langage naturel
  - ☞ 2.3.1 Dictionnaires et étiquetage de surface
  - ☞ 2.3.2 Traduction automatique statistique

**3. Ingénierie de la parole**

- ☞ 3.1. Rappels de traitement numérique du signal
- ☞ 3.2. Le signal de parole : analyse, unités pertinentes et variabilité
- ☞ 3.3. Codage et compression de la parole

---

---

---

---

---

---

---

---

**Support de cours**



**Avertissement**

Ce support de cours **ne se substitue pas** aux notes que vous aurez prises pendant les séances !

Résumé des principales notions qui ont été abordées !



[HTTP://ANDON.TCHECHMEDJIEV.EU/RICM4-MCL/](http://andon.tchechmedjiev.eu/ricm4-mcl/)

---

---

---

---

---

---

---

---

## Les Travaux Pratiques



- ☞ Sujets & Fichiers/Outils mis à disposition au fur et à mesure
  
- ☞ 3h en salle de TP
  - ☞ Compte rendu pour la séance d'après
  
- ☞ Vous pouvez continuer d'y travailler entre temps

---

---

---

---

---

---

---

---

## Communication Langagière : Ingénierie des langues et de la parole



- ☞ **Objectifs du cours:**
  - ☞ Connaître les principales applications en ingénierie des langues et de la parole
  
  - ☞ Être sensibilisé aux problèmes et aux besoins du domaine
  
  - ☞ Avoir un aperçu de quelques techniques et méthodes (nombreuses !) mises en œuvre pour le traitement automatique du langage écrit et parlé
  
  - ☞ Manipuler quelques-uns des outils développés

---

---

---

---

---

---

---

---

## Sources



- ☞ Ancien cours G. Sérasset & J. Goulian
  
- ☞ Cours de P. Koehn (Univ. Edimbourg)
  
- ☞ Cours « automates » de A. Lecomte
  
- ☞ "Speech and Language Engineering",  
*Editor: Martin Rajman, EPFL Press,  
distributed by CRC Press, 2007*

---

---

---

---

---

---

---

---

## Ingénierie Linguistique



- ☞ Acquisition, gestion, structuration, analyse, interprétation, modélisation de l'information ... langagière (orale ou textuelle)
- ☞ Finalités :
  - ☞ Mise en place d'applications concrètes
    - ☞ Résumé de textes, Indexation et recherche d'information,
    - ☞ Traduction Automatique, Dialogue Homme Machine
    - ☞ Vérification ou correction orthographique,
    - ☞ Dictée & synthèse vocale...
  - ☞ Confronter les modèles de la linguistique aux exigences de la modélisation informatique
  - ☞ Domaine privilégié de l'IA et de l'Informatique

---

---

---

---

---

---

---

---

## Bref Historique



- ☞ **Pionniers** – Turing, Von Neumann , ... , Descartes, ...
- ☞ **1950** – Premiers espoirs ... premières déceptions :
  - ☞ Traduction Automatique, processus cryptographique
- ☞ **60-70s** – Avènement de l'Intelligence artificielle :
  - ☞ Strong A.I. & Artificial Human (HAL)
- ☞ **2000** – Les technologies vocales atteignent des performances honorables
  - ☞ Synthèse, reconnaissance, compréhension
  - ☞ Seulement quelques langues traitées!...

---

---

---

---

---

---

---

---

## Bref panorama des applications



- |   |                               |
|---|-------------------------------|
| ☞ Traduction Automatique                            | ☞ Communication Homme-Machine |
| ☞ Compréhension ou condensation de textes           | ☞ Dictée vocale               |
| ☞ Génération de textes                              | ☞ Synthèse vocale             |
| ☞ Détection ou correction d'erreurs, édition        |                               |
| ☞ Documentation, indexation et Moteurs de recherche |                               |

**Ingénierie linguistique :**  
*Domaine pluridisciplinaire vaste et composé de sous-domaines bien identifiés ayant leurs problèmes spécifiques.*

---

---

---

---

---

---

---

---

## Principes généraux

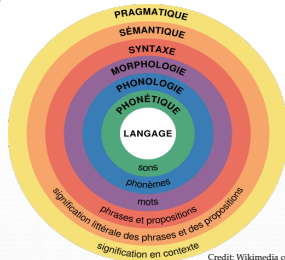
### Analyse

Construire une représentation formelle du message d'entrée  
 à manipuler et dépourvue d'ambiguïtés

VS

### Génération

Générer un message à partir d'une représentation interne



- ☞ Composés de suites de traitements complexes
- ☞ Plus ou moins liées

Credit: Wikimedia commons

---

---

---

---

---

---

---

---

## Problèmes Spécifiques

☞ Principal: Ambiguïté à tous les niveaux!

*J'ai vendu la voiture à Jules  
 Je dessine un homme avec un stylo  
 L'école de cuisine grenobloise  
 Cet avocat est encore vert  
 Il a frappé sa femme, et son fils aussi*

- ☞ Constitution des connaissances nécessaires
  - ☞ Ressources linguistiques :
    - ☞ Lexiques + Systèmes de règles
    - ☞ Connaissances du monde (Ontologie)

---

---

---

---

---

---

---

---

## Informatiser une langue

	Classes de Services	Criticités
A	Traitement de textes, services de base (saisie, affichage, impression, recherche, sélection, tri)	Importante
B	Traitement de textes, services avancé (correcteurs d'orthographe, de grammaire, de style)	Moyenne
C	Synthèse et reconnaissance de la parole	Moyenne
D	Traduction automatisée	Assez importante
E	Reconnaissance optique de caractères	Assez importante
F	Ressources, Dictionnaires bilingues et d'usage	Importante

D'après [Berment 2004]

---

---

---

---

---

---

---

---

**Exemple du Khmer**  
« Niveau d'informatisation de la langue »

Services / Ressources	Criticité (/10)	Note (/20)	Note pondérée (Criticité x Note)
<b>Traitement du texte</b>			
Saisie simple	10	16	160
Visualisation / impression	10	14	140
Recherche et remplacement	8	12	48
Sélection du texte	6	12	72
Tri lexicographique	5	0	0
Correction orthographique	2	0	0
Correction grammaticale	0	0	0
Correction stylistique	0	0	0
<b>Traitement de l'oral</b>			
Synthèse vocale	5	0	0
Reconnaissance de la parole	5	0	0
<b>Traduction</b>			
Traduction automatisée	8	4	32
<b>ROC</b>			
Reconnaissance optique de caractères	9	0	0
<b>Ressources</b>			
Dictionnaire bilingue	10	4	40
Dictionnaire d'usage	10	0	0
<b>Total</b>			540 / 1760
<b>Moyenne</b>			6,2 / 20

[Berment 2004]  
Évaluation par un chercheur de l'INALCO)

---

---

---

---

---

---

---

---

---

---

---

---

**Communication Langagière**  
Ingénierie des langues et de la parole

— ∞ —

**1. Introduction générale**

**2. Ingénierie des langues**

- ☞ 2.1. Représentation et codage des textes
- ☞ 2.2. Théorie de l'information et probabilités
- ☞ 2.3. Quelques applications du traitement automatique du langage naturel
  - ☞ 2.3.1 Dictionnaires et étiquetage de surface
  - ☞ 2.3.2 Traduction automatique statistique

**3. Ingénierie de la parole**

- ☞ 3.1 Rappels de traitement numérique du signal
- ☞ 3.2 Le signal de parole : analyse, unités pertinentes et variabilité
- ☞ 3.3 Codage et compression de la parole

---

---

---

---

---

---

---

---

---

---

---

---