

Mémoire de fin d'études  
Pour l'obtention du diplôme d'Ingénieur d'État en Informatique  
Option : Systèmes Informatiques

---

Création d'un corpus de l'aphasie de Broca et  
développement d'un système Speech-to-speech de  
réhabilitation de la parole

---

*Réalisé par :*  
BELGOUMRI Mohammed  
Djameleddine  
[im\\_belgoumri@esi.dz](mailto:im_belgoumri@esi.dz)

*Encadré par :*  
M. SMAILI KAMEL  
[smaili@loria.fr](mailto:smaili@loria.fr)  
M. LANGLOIS DAVID  
[david.langlois@loria.fr](mailto:david.langlois@loria.fr)  
Mme. Zakaria  
Chahnez  
[c\\_zakaria@esi.dz](mailto:c_zakaria@esi.dz)

# Table des matières

<b>Page de garde</b>	<b>1</b>
<b>Table des matières</b>	<b>1</b>
<b>Table des figures</b>	<b>2</b>
 <b>I État de l’art</b>	 <b>4</b>
 <b>1 Traduction automatique</b>	 <b>5</b>
1.1 Classification des méthodes de traduction automatique . . . . .	5
1.2 Traduction automatique à base de règles . . . . .	6
1.3 Traduction automatique statistique . . . . .	7
1.4 Traduction automatique neuronale classique . . . . .	7
1.5 Traduction automatique neuronale profonde . . . . .	7
1.6 Traduction automatique à base d’attention . . . . .	7
 <b>Bibliographie</b>	 <b>8</b>

# Table des figures

1.1	Taxonomie des méthodes de traduction automatique. . . . .	6
1.2	Triangle de Vauquois . . . . .	7

# Abréviations

**IL** interlingue. 6

**LC** langage cible. 5, 6

**LS** langue source. 5, 6

**TA** traduction automatique. 4, 5

**TABR** traduction automatique à base de règle. 4–6

**TAL** traitement automatique du langage. 4

**TAN** traduction automatique neuronale. 4

**TAS** traduction automatique statistiques. 4

# Première partie

## État de l'art

# Chapitre 1

## Traduction automatique

La traduction automatique (TA) est une branche du traitement automatique du langage (TAL). Elle étudie l'utilisation des systèmes informatiques pour traduire le texte ou la parole d'une langue (appelée source) vers une autre (appelée cible) [1].

Dans ce chapitre, on introduit la traduction automatique du texte pour donner un point de références aux discussions des chapitres suivants. On y discute les plus répondues du domiane ainsi que les développements les plus récents qu'il a vécu.

### 1.1 Classification des méthodes de traduction automatique

La classification des méthodes de TA la plus citée à travers la littérature, repose sur les outils mathématiques de celles-ci. On distingue notamment trois familles de méthodes [3] :

1. Des méthodes basées sur des connaissances linguistiques (règles de traduction).
2. Des méthodes basées sur les statistiques.
3. Des méthodes basées sur les réseaux de neurones.

On les appelle respectivement Traduction automatique à base de règle (TABR), Traduction automatique statistiques (TAS), Traduction automatique neuronale (TAN). Les méthodes dans chacune de ces trois catégories peuvent être encore classifiées [3, 2], ce qui donne lieu à la hiérarchie représentée par la Figure 1.1.

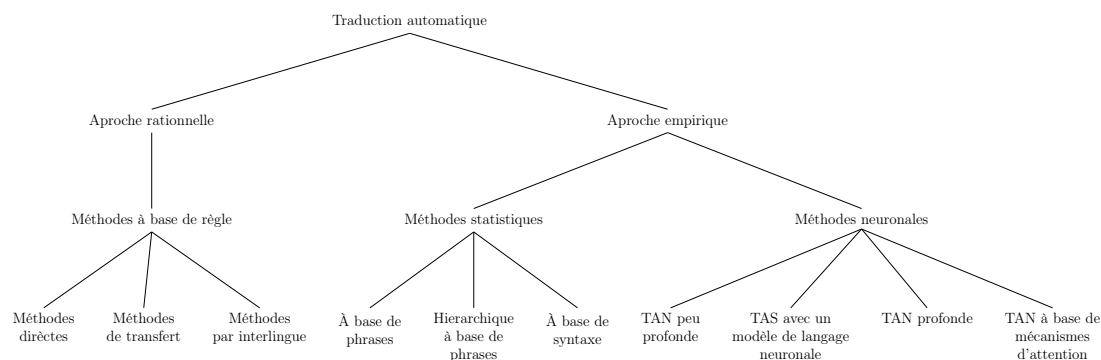


FIGURE 1.1 – Taxonomie des méthodes de traduction automatique.

Le reste de ce chapitre sera organisé selon la structure donnée par la Figure 1.1. Cela nous donne la structure suivante

- Traduction automatique à base de règles
- Traduction automatique statistique
- Traduction automatique neuronale classique
- Traduction automatique neuronale profonde
- Traduction automatique à base d’attention

## 1.2 Traduction automatique à base de règles

La traduction automatique à base de règle (TABR) est historiquement le premier paradigme de TA. Étant apparue pendant les années 1950s, elle resterait l’approche dominante de TA jusqu’aux 1980s [1].

Comme son nom l’indique, la TABR est basée sur des règles de traduction explicites, qui sont généralement créées manuellement à partir de connaissances linguistiques sur la langue source (LS) et la langue cible (LC). Les règles en question peuvent être d’ordre lexical (i.e des dictionnaires), syntaxique (i.e des grammaires) ou sémantique.

Toutes les méthodes de TABR passent par deux phases : l’analyse de l’entrée dans la LS et la génération (ou synthèse) de la sortie dans la LC. Cependant, les règles utilisées dans ces deux phases peuvent varier en profondeur et dans le types

de connaissances linguistiques employées. On distingue ainsi la TABR en trois sous-familles de méthodes comme indiqué sur le sous-arbre gauche de la Figure 1.1.

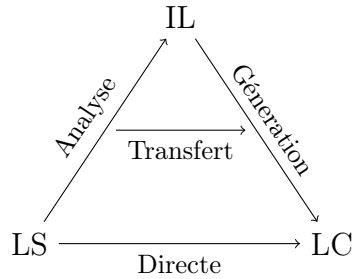


FIGURE 1.2 – Triangle de Vauquois

### 1.3 Traduction automatique statistique

### 1.4 Traduction automatique neuronale classique

Comme toutes les méthodes discutées dans ce chapitre, les méthodes neuronales passent par deux phases : celle d'analyse et celle de génération. Cela se fie naturellement au traitement par une architecture encodeur-décodeur. Il s'agit d'une architecture

### 1.5 Traduction automatique neuronale profonde

### 1.6 Traduction automatique à base d'attention



# Bibliographie

- [1] Sin-wai CHAN. *Routledge Encyclopedia of Translation Technology*. Routledge, Taylor & Francis Group, 2015.
- [2] Marta R. COSTA-JUSSÀ et al., éd. *Hybrid Approaches to Machine Translation*. en. Theory and Applications of Natural Language Processing. Cham : Springer International Publishing, 2016. ISBN : 9783319213101 9783319213118. DOI : [10.1007/978-3-319-21311-8](https://doi.org/10.1007/978-3-319-21311-8). URL : <http://link.springer.com/10.1007/978-3-319-21311-8> (visit  le 11/10/2022).
- [3] Shuoheng YANG, Yuxin WANG et Xiaowen CHU. « A survey of deep learning techniques for neural machine translation ». In : *arXiv preprint arXiv :2002.07526* (2020).