

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université A.MIRA-BEJAIA



Faculté des lettres et langues
Département de français

Cours magistral et travaux dirigés en vue de l'obtention de l'Habilitation à Diriger des Recherches

Intitulé du cours : Traitement Automatique des Langues

Élaboré par :

Dr. BEDDAR Mohand

Maître de conférences classe B

**Public visé : Master 1 Didactique des Langues Etrangères
Année universitaire : 2020/2021**

Avant propos

L'usage de la langue au quotidien n'est plus l'apanage des humains car de plus en plus de machines arrivent à communiquer dans différents domaines et dans diverses situations. En effet, grâce au Traitement Automatique des Langues, des ordinateurs sont en mesure aujourd'hui de manipuler la plupart des langues et sont capables d'accomplir toutes sortes de tâches linguistiques réservées jusqu'au-là aux êtres humains. De la synthèse vocale à la traduction des textes multilingues, en passant par l'extraction d'informations, les applications en ingénierie linguistique rivalisent de créativité technologique en repoussant chaque jour les limites de l'intelligence artificielle.

Le pari est presque réussi puisque l'objectif théorique du TAL (Traitement Automatique des Langues) est pleinement atteint. Des modèles linguistiques, établis sur au moins une partie de la langue, implémentables sur un ordinateur simulent notre capacité à parler et à comprendre les "langues naturelles" comme le français, l'anglais ou l'allemand.

Le TAL s'impose de nos jours, somme toute, comme un domaine clé dans l'avenir de la recherche et comme une spécialité d'avenir pour nos étudiants inscrits en français. Son enseignement dans notre département ne fait plus débat car en adéquation totale avec les contenus de formation préexistants notamment ceux portant sur l'analyse de la langue.

OBJECTIF PRINCIPAUX DU COURS :

- Présenter aux étudiants les besoins d'un domaine en pleine expansion : industries de la langue ; préciser les compétences nécessaires pour y accéder, présenter les débouchés professionnels ;
- Donner aux étudiants des bases solides leur permettant de s'orienter vers les métiers de l'ingénierie linguistique,
- Les familiariser avec quelques applications du TAL ;
- Les initier aux problématiques spécifiques à ce domaine en abordant la question de la complexité des langues naturelles.
- Initier les étudiants aux méthodes de formalisation pour la conception d'applications en TAL (traduction automatique)

STRUCTURE DU COURS :

Cours magistral 1 : TAL et quelques notions linguistiques

Cours Magistral 2 : Complexité du langage naturel et TAL

- Chapitre 1 : Ambiguïtés des langues naturelles et TAL
- Chapitre 2 : Classification des ambiguïtés d'un point de vue TAL

Cours magistral 3 : Applications en Traitement Automatique des Langues : Exemple de la traduction automatique

- Chapitre 1 : Introduction à la traduction automatique
- Chapitre 2 : Construire un traducteur automatique

CONTENU DU COURS :

- Un panorama du domaine du Traitement Automatique des Langues (TAL) : définition, domaines d'application, etc. ;
- Outils et méthodes : analyseurs (morphologique, syntaxique), traduction automatique, reconnaissance vocale, extraction d'information, etc. ;
- Initiation aux problèmes linguistiques majeurs : ambiguïté, anaphore, figement etc. et aux problèmes de représentation et de description des données ;
- Présentation de la traduction automatique comme application phare du TAL (définition et conception)

PUBLIC VISE : Master 1 Didactique des Langues Etrangères

CONTROLE DES CONNAISSANCES : Examens sur table, devoirs à la maison, projets en groupe

VOLUME HORAIRE HEBDOMADAIRE : 1H30 (12 semaines au semestre 2)

PRE-REQUIS ET CONNAISSANCES PREALABLES RECOMMANDES :

- Connaissances en linguistique appliquée et contrastive
- Connaissances élémentaires en informatique

CM 1

TAL et quelques notions linguistiques

Descriptif du cours

Le cours intitulé « TAL et quelques notions linguistiques » est une introduction à un domaine nouveau pour les étudiants en M1 Didactique des Langues Etrangères, celui du Traitement Automatique des Langues. A travers un texte ponctué régulièrement de questions, l'objectif de ce cours est double, vulgariser des connaissances souvent considérées comme trop techniques ou complexes et les faire assimiler facilement par des apprenants novices dans ce domaine.

Ce premier cours prend en compte les contenus pédagogiques de la formation français ainsi que les acquis linguistiques des étudiants inscrits dans cette formation. L'accent est davantage mis sur le rapport entre Traitement Automatique des Langues et connaissances linguistiques. C'est d'ailleurs l'objectif de la deuxième partie qui passe en revue les différents niveaux de langues. Le but étant de démontrer que le TAL est avant tout un domaine d'application de la linguistique loin des idées reçues qui le placent exclusivement dans celui de l'informatique. Ainsi, les acquis en linguistique des étudiants seront renforcés et leurs mécanismes de raisonnement et d'inférence aiguisés grâce aux modèles de formalisation linguistique mis en exergue dans ce cours.

Objectifs du cours

- Définir le domaine du TAL et les notions clés qui l'encadrent.
- Identifier le rapport entre traitement automatique des langues et langues naturelles.
- Décrire le fonctionnement d'un système de traitement automatique des langues.
- Approfondir les connaissances linguistiques des étudiants pour une meilleure compréhension des besoins en TAL.
- Apprendre les mécanismes de formalisation du langage naturel.

Contenu du cours :

1	Qu'est-ce que le TAL ?	3
2	Qu'est-ce que le langage naturel ?.....	3
2.1	Langues artificielles/Langues naturelles.....	4
3	Qu'est-ce que l'adjectif « automatique » ?	4
4	Quelles sont les connaissances nécessaires pour faire du TAL ?	5
5	Quelles sont les sciences qui font des recherches en TAL ?	5
6	Schéma général d'un système de traitement automatique des langues	5
7	Pourquoi avons-nous besoin du TAL ?	6
7.1	Enjeux économiques	6
7.2	Enjeux politiques, géopolitiques et culturels	6
8	Quelles sont les principales applications du TAL ?	7
9	Les connaissances sur la langue	7
9.1	Phonétique.....	7
9.2	Phonologie	7
9.3	Morphologie.....	8
9.3.1	Les sous-domaines de la morphologie	8
9.3.1.1	Morphologie flexionnelle	8
9.3.1.2	Morphologie dérivationnelle	9
9.3.1.3	Morphologie compositionnelle.....	10
9.4	Syntaxe.....	11
9.5	Morphosyntaxe	14
9.6	Sémantique.....	15
9.7	Pragmatique	15
10	Exemple d'application.....	17
11	Bibliographie.....	19
12	Exercices	20

1 Qu'est-ce que le TAL ?

L'objectif du Traitement Automatique des Langues est la conception de logiciels capables de traiter de façon automatique des données linguistiques, c'est-à-dire des données exprimées dans une langue (dite naturelle).

Les données linguistiques peuvent être de différents types :

- il peut s'agir de **textes écrits** ou **oraux**, de différentes formes (des mots isolés, des textes, des dialogues, etc.).

Qui dit traitement dit manipulation d'un **objet d'entrée**, aboutissant à la modification de cet objet en un nouvel objet (**objet de sortie**).

Selon la nature de l'application, les traitements peuvent soit agir sur **un texte préexistant** (par exemple pour le corriger, ou en extraire l'information, ou le résumer, ou le traduire), nous parlerons dans ce cas, **d'analyse d'un texte**, soit **créer un texte** à partir d'informations données (on parlera dans ce cas de **génération** d'un texte). Pour ce faire, il est nécessaire non seulement de **connaître les manipulations informatiques**, mais aussi de maîtriser les principes de **la construction interne** du texte à manipuler. C'est là qu'intervient la linguistique.

Conclusion : pour faire du TAL, il faut des connaissances dans au moins deux domaines : linguistique et informatique.

2 Qu'est-ce que le langage naturel ?

Le langage est la faculté qu'ont les humains à communiquer entre eux à l'aide d'un système de signes. Cette capacité est partagée par les Français, les Chinois, les Grecs etc. Le langage est une notion très générale.

Les langues telles que : le français, le chinois, le grec sont les réalisations concrètes de la capacité que l'on a à communiquer.

On dit **langues naturelles** pour distinguer les langues comme le français, le chinois etc. (les langues qui se sont créées au fil du temps) des langues créées de façon complètement artificielle (comme l'espéranto par exemple). C'est également pour les distinguer des langages de programmation (qu'on appelle des fois les langues de programmation), qui sont des codes artificiels créés de toute pièce par les humains. Par conséquent, les langues naturelles et les langues artificielles n'ont ni la même structure, ni les mêmes caractéristiques.

2.1 Langues artificielles/Langues naturelles

LANGUES ARTIFICIELLES	LANGUES NATURELLES
Fixe : Une fois que l'on a fixé le code, il ne change plus	souple et évolutive : Elles évoluent au cours de l'histoire ; il suffit de comparer le français des différentes époques
Explicites : On connaît toutes les règles (puisque c'est nous qui les avons créées) et la valeur de ces règles	Implicites : Il y a beaucoup de choses qui nous échappent ; bien qu'on comprenne la plupart du temps, on n'arrive pas toujours à expliquer comment l'information nous parvient
Non-ambiguës : Chaque chose possède une valeur précise.	Ambiguës : La même chose peut être interprétée de différentes façons.

3 Qu'est-ce que l'adjectif « automatique » ?

Un **traitement automatique** est un traitement qui s'opère par des moyens « **mécaniques** », par opposition à un traitement effectué par **un humain**. La machine qu'on utilise est l'ordinateur, c'est à dire une machine qui est capable de faire des calculs. Cela présuppose que l'on soit capable de représenter **les données linguistiques** sous forme d'objets et les soumettre à des calculs. Une suite de calculs effectuée dans un certain ordre s'appelle **un programme**.

L'automatisation du traitement peut être **totale** ou **partielle** :

Il peut y avoir des traitements entièrement automatiques, c'est-à-dire effectués en totalité par un ordinateur, sans intervention d'un humain ; dans ce cas on parle par exemple de traduction automatique.

Il peut y avoir des traitements **assistés par un ordinateur**, c'est-à-dire que l'humain intervient pour préparer ou pour terminer la tâche effectuée par une machine. Il peut également intervenir pendant le traitement ; dans ce cas on parle par exemple de **traduction assistée par ordinateur** ou de publication assistée par ordinateur ou d'enseignement assisté par ordinateur.

La notion de traitement automatique soulève tout de même un certain nombre de **controverses et de questions**, comme d'ailleurs toute nouvelle innovation visant à remplacer un humain dans certaines tâches :

Est-ce qu'une machine peut vraiment comprendre une langue ? Est-ce que ce n'est pas une compétence réservée aux humains ? Est-ce qu'il n'y a pas dans la langue une bonne part de psychologie, d'intelligence, de subjectivité que l'on ne pourra jamais transmettre à une machine ? Et si c'est le cas, si on apprend à des machines l'intelligence, la psychologie etc., Est-ce qu'elles ne vont pas nous dépasser ?

Bien plus concret : Est-ce que les ordinateurs ne vont pas mettre au chômage des milliers de personnes comme les traducteurs humains par exemple ?

IL FAUT ETRE REALISTE :

L'ordinateur ne sait pas faire que ce qu'on lui dit de faire, sa seule supériorité réside dans la **rapidité des calculs** et (éventuellement) dans la **taille de sa mémoire**.

4 Quelles sont les connaissances nécessaires pour faire du TAL ?

- **connaissances en linguistiques** : il faut savoir décrire les mécanismes de fonctionnement des langues ;
- **connaissances en formalisation** : la description d'une langue en vue d'un traitement automatique doit être très précise, puisqu'elle est destinée à être implémentée sur une machine qui ne sait pas faire des choix par elle-même ; on a donc recours à des formalismes qui expriment les faits de la langue d'une façon acceptable par un ordinateur ;
- **connaissances en informatique** : il faut plus au moins avoir des connaissances dans les langages de programmation afin de pouvoir transformer les formalismes en un programme.

Résumé :

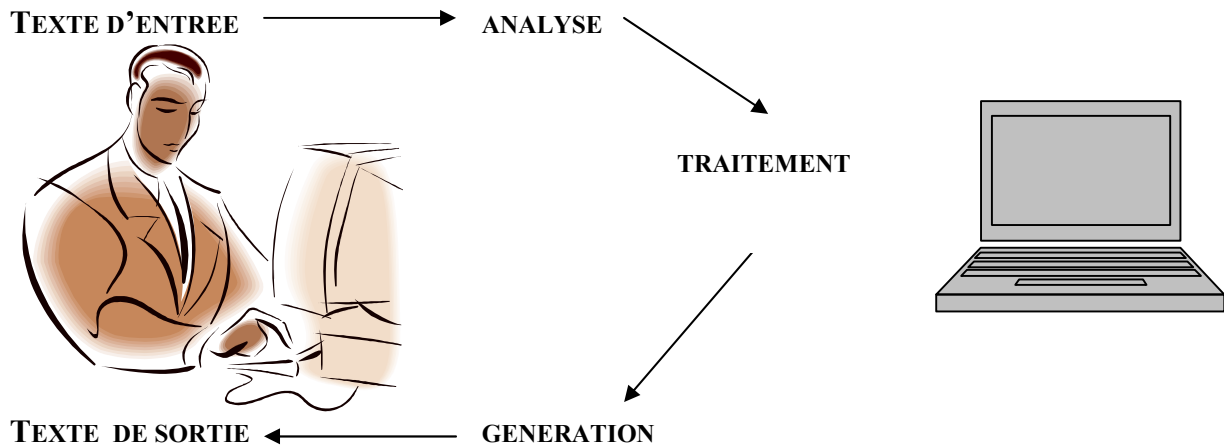
Le traitement automatique des langues a comme matériaux **des données linguistiques** exprimées dans **une langue naturelle**, et pour pouvoir traiter ces données, il faut être capable d'explicitier **les règles de la langue**, de les représenter **dans un formalisme** et de les implémenter sur une machine sous la forme **d'un programme**.

5 Quelles sont les sciences qui font des recherches en TAL ?

- **La linguistique informatique** et **l'informatique linguistique**, qui développent des programmes en TAL et créent, à cet effet, des langages informatiques, propres au TAL ;
- **La linguistique**, qui fournit des théories explicites sur le langage ;
- **L'informatique** qui permet d'optimiser les algorithmes et les programmes de traitement, mais aussi de développer des techniques formelles (de démonstration ou de résolution des problèmes, par exemple).
- **Les mathématiques** qui étudient les propriétés formelles des outils de traitement et des théories ;
- **L'intelligence artificielle** qui s'occupe de la représentation des connaissances et de leur utilisation.

6 Schéma général d'un système de traitement automatique des langues

Le schéma général d'un système de traitement automatique des langues peut être illustré de la façon suivante :



Ce schéma constitué principalement de 3 phases (analyse, traitement et génération) sous-tend diverses applications telles que : le dialogue homme-machine, la traduction automatique, la dictée vocale, le résumé automatique du texte, etc.

Chacune de ces phases se compose d'une suite de traitements complexes. Lorsque l'on traite d'une façon automatique les langues naturelles, toutes les connaissances sur la langue doivent être mobilisées.

7 Pourquoi avons-nous besoin du TAL ?

7.1 Enjeux économiques

Nous vivons dans une société d'information où d'avantage d'informations sont véhiculées en langues naturelles : Nous avons besoin, par conséquent, de méthodes efficaces et rapides pour classer, archiver, analyser, interroger et traduire des documents ;

Exemples :

- les catalogues informatisés de la bibliothèque
- documentation scientifique et technique : document d'entretien d'une centrale nucléaire, d'un avion, ... (Des milliers de pages).
- Informations sur internet

7.2 Enjeux politiques, géopolitiques et culturels

Au début de l'informatique, la plupart des applications étaient en anglais et à travers la langue anglaise, les utilisateurs se sont retrouvés sous l'influence de la culture américaine. Il y a donc dans le développement des applications du TAL une volonté de sortir d'une dépendance certaine aux géants du numérique.

Face à cette hégémonie américaine, les pays européens se sont mis à développer ce que nous appelons communément aujourd'hui les **industries de la langue**. C'est-à-dire un ensemble très diversifié de travaux qui contribuent au développement ou à l'amélioration des logiciels de traitement des données linguistiques.

Conséquence : De nouveaux métiers ont vu le jour dans le domaine du TAL (Ex : ingénieur linguiste qui travaille dans ce qu'on appelle « l'ingénierie linguistique », c'est-à-dire un ensemble d'activités ayant comme objectif la création et/ou l'amélioration des applications en TAL)

8 Quelles sont les principales applications du TAL ?

- Les systèmes de dialogue homme-machine et l'interrogation des bases de données
- La recherche documentaire ; la constitution et l'interrogation de systèmes documentaires ; l'extraction d'information
- La reconnaissance optique de caractères typographiés (OCR)
- La vérification et correction orthographique et stylistique
- La traduction automatique :
- Les moteurs de recherche
- Le résumé automatique
- La synthèse vocale
- Etc.

9 Les connaissances sur la langue

Pour traiter automatiquement les langues naturelles, le programme idéal devrait inclure différentes connaissances relatives à la langue : quels sont les différents mots ? Comment se prononcent-ils, que signifient-ils, comment se combinent-ils pour former une phrase et comment le sens des différents mots contribue au sens de la phrase. De plus, il devrait aussi pouvoir utiliser des connaissances générales sur le monde et les contextes d'utilisation des textes.

Ces différentes connaissances sont généralement classées et étiquetées sous les domaines suivants :

9.1 Phonétique

Premier découpage : on découpe la phrase en sons ; comment fait-on pour la prononcer ? Comment produit-on les sons ? Quelles sont ses caractéristiques ?

La phonétique est l'étude scientifique des sons de la langue. Elle exclut les autres sons, produits par les êtres humains, même s'ils servent parfois à communiquer (les toux, les raclements de gorge, ...). L'étude phonétique d'une langue peut se faire sans faire appel au sens.

Comment est produit le son /p/, /s/, /r/ ?

9.2 Phonologie

Deuxième découpage : on découpe la phrase en sons significatifs ; comment le son produit le sens ? Quelle est la différence (au niveau du son) entre « bougent » et « rouge » ?

La **phonologie** s'occupe de la fonction des sons dans la transmission d'un message. Il faut donc comprendre une langue pour faire de la phonologie. En d'autres termes, la phonologie recherche les différences de prononciation qui correspondent à des différences de sens, c'est

ce qu'on appelle des **oppositions distinctives**. Quelle est la différence entre « rouge » et « bouge » ?

La phonétique et la phonologie sont à la base de tout ce qui concerne le **traitement de la parole**, c'est-à-dire, la reconnaissance et la synthèse de la parole.

9.3 Morphologie

C'est le troisième découpage de la phrase, elle est segmentée en mots et on regarde attentivement ce qui se passe à l'intérieur des mots ; on s'aperçoit rapidement qu'il existe des unités de sens qui sont plus petites que le mot !

La morphologie s'occupe des plus petites unités de forme et de sens, qu'on appelle les **morphèmes**.

Les morphèmes peuvent être des mots entiers (*pomme*), ou des parties de mots (im –possible), que l'on appelle des affixes. Un mot est généralement construit à partir de plusieurs morphèmes qui se combinent entre eux.

mot	préfixe	racine	suffixe
penser		pens-	er
pensable		pens-	able
impensable	im	pens-	able

9.3.1 LES SOUS-DOMAINES DE LA MORPHOLOGIE

9.3.1.1 Morphologie flexionnelle

Elle désigne le processus par lequel les mots reçoivent des affixes qui expriment des informations, sans changer leur catégorie grammaticale (la conjugaison pour les verbes, le pluriel pour les noms et les adjectifs etc.)

Remarque :

Chaque morphe représente au moins un morphème ; mais il peut représenter plusieurs morphèmes

Exemples :

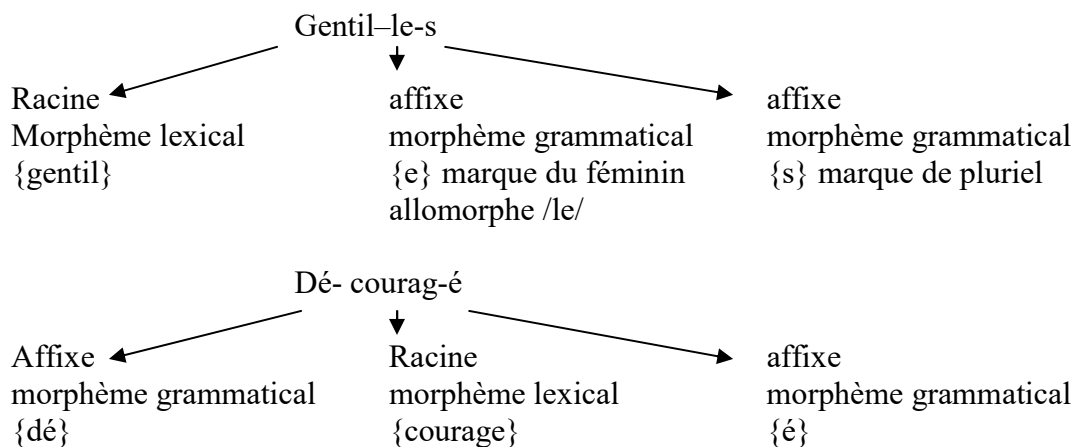
Femmes

/Femme/	/s/
morphème lexical	morphème grammatical
{femme}	{s} marque de pluriel

Chantais

/chant/	/ais/
morphème lexical	morphème grammatical
{chanter}	{imparfait}, {singulier}, {1 ou 2 personne}

L'affixation est le processus qui combine une racine (base, radical) correspondant à un morphème lexical et différents affixes (préfixes et suffixes).



La racine est un morphème lexical qui constitue la base d'un mot (il n'y a pas de mot sans racine !)

Les affixes ce sont de morphèmes grammaticaux (ils ne sont donc pas autonomes), qui peuvent soit précéder (préfixe) soit suivre la racine (suffixe).

Remarque :

Dans certaines langues, il existe aussi des infixes : des affixes qui s'insèrent entre les deux morphèmes lexicaux, d'un mot composé.

9.3.1.2 Morphologie dérivationnelle

Elle est le processus de formation des mots, soit par changement de leur catégorie grammaticale (*penser V : pensable Adj, triste Adj : tristesse N*), soit par changement de leur sens (*possible Adj : impossible Adj ; coiffer V : décoiffer V*)

Les différences avec la flexion :

Les affixes dérivationnels changent souvent la catégorie du mot

Exemple :

Base	Suffixe de flexion	Suffixe de dérivation
triste Adj	triste-s Adj	triste –ment Adverbe
embrass- Base verbale	embrass-ait Verbe	embrass-ade Nom
Aliment Nom	aliment-s Nom	aliment-aire Adj

Remarque 1 :

Ce sont surtout les suffixes qui changent la catégorie du mot ; les préfixes le font très rarement.

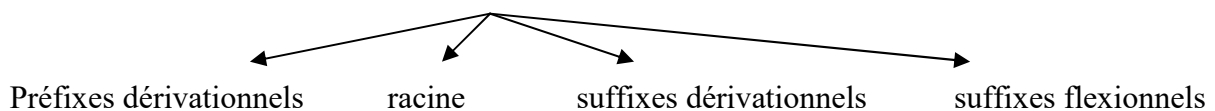
Remarque 2 :

Si changement de catégorie il y a → alors dérivation.

Les affixes dérivationnels modifient le contenu sémantique du mot :

Base	Suffixe de flexion	Affixe de dérivation
Mortel Adj	mortel-s Adj	im-mortel Adj
Marche Nom	Marche-s Nom	Dé-marche Nom
Gland-er Base verbale	gland-e (il) Verbe	gland-ouille (il) verbe

Structure générale d'un mot:



9.3.1.3 Morphologie compositionnelle

Elle désigne le processus par lequel on forme les nouveaux mots à partir des mots existants : *porte-serviette, machine à coudre, aigre-doux etc.*

Exemple :

- Un chou-fleur*
- Une machine à écrire*
- Un tableau noir*
- Un cordon bleu*
- Un fait divers*

Les composés ressemblent à des syntagmes normaux, donc, pour les distinguer, plusieurs tests ont été mis en place :

Tests sémantiques :

Opacité sémantique

Exemple :

- Une boîte bleu = une boîte qui est bleu*
- Un cordon bleu = *un cordon qui est bleu = un excellent cuisinier*
- Un tableau noir = *un tableau qui est noir (il peut être vert) = un certain type de tableau*

Règle :

Le sens d'un composé n'est pas une simple addition des sens des lexèmes qui le composent.
On appelle ce phénomène l'opacité sémantique.

Sauf que ce n'est pas toujours vrai (*un fait divers*)

Tests syntaxiques :

- insertion d'autres éléments à l'intérieur de mots composés

Exemple :

Un fait évident

Un fait totalement évident

Un fait particulièrement évident

Un fait divers

* *Un fait totalement divers*

* *Un fait particulièrement divers*

Les pommes de mon verger

Les pommes pourris de mon verger

Les pommes de terre

* *Les pommes pourris de terre*

Les pommes de terre pourris

Remplacement par les synonymes :

Exemples :

Un acte manuscrit

Un document manuscrit

Un acte de naissance

* *Un document de naissance*

J'ai cassé le vase.

J'ai brisé le vase.

J'ai un cœur brisé.

* *J'ai un cœur cassé.*

9.4 Syntaxe

D'un point de vue TAL, on s'intéresse à la façon dont les mots sont liés entre eux, Pourquoi certaines phrases sont grammaticalement correctes et d'autres ne le sont pas par exemple?

La syntaxe est l'étude des règles qui régissent la combinaison des mots en phrases. Dans toutes les langues, certaines combinaisons des mots sont possibles, d'autres sont interdites.

Pour décrire la syntaxe, on utilise la notion de **catégorie grammaticale** et de **fonction syntaxique** des mots.

Les catégories grammaticales du français sont : les noms (substantif), les déterminants, les verbes, les adverbes, les adjectifs, les prépositions, les pronoms, les conjonctions, les interjections.

Les fonctions syntaxiques sont: sujet, objet direct ou indirect, compléments.

Comment les catégories grammaticales se combinent en phrases ?

Les mots ne se combinent pas de manière anarchique pour former une phrase. Ils le font dans un ordre précis. En français, l'ordre de mots dans une phrase est essentiel pour la construction du sens.

Exemple:

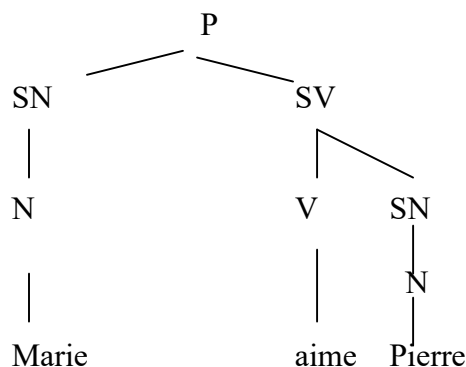
Marie aime Pierre.

Pierre aime Marie. (changement de sens)

**Aime Marie Pierre. (grammaticalement incorrecte)*

La troisième phrase est agrammaticale, tandis que la signification dans les deux premières est complètement renversée.

On représente les relations entre les mots sous forme d'un arbre syntaxique. Un arbre syntaxique est une structure ordonnée qui hiérarchise les relations entre les différents mots dans une phrase :



De même que les mots ont une catégorie syntaxique (Nom, verbe), les éléments regroupant plusieurs mots sont appelés les syntagmes (syntagmes nominaux, verbaux etc.). Les éléments de la phrase sont appelés les constituants.

Cet arbre nous renseigne sur :

- La nature des constituants de la phrase (SN, V, N, etc.)
- La structure hiérarchique de ces constituants (SN (Marie) dépend de P, N (Pierre) dépend de SN)
- Leurs fonctions syntaxiques (*Marie* est le sujet de la phrase, *Pierre* est l'objet direct)

L'étude de la syntaxe comporte trois parties:

- Les catégories syntaxiques
- Les règles syntaxiques et la construction des phrases
- La transformation des phrases.

On appelle **syntagme** un groupe de mots qui remplit la même fonction qu'un mot d'une catégorie syntaxique. Un syntagme est une unité de décomposition de la phrase.

Types de syntagmes :

Nom (syntagme nominal SN): objet, personne pouvant être sujet : main, chien, Jeanne, ...

Adjectif (syntagme adjectival SA): qualifie un syntagme nominal : long, petite, illuminé, ...

Verbe (syntagme verbal SV): action ou état : être, manger, chercher, ...

Adverbe (syntagme adverbial SAdv): modifie un syntagme verbal, adjectival ou adverbial : vraiment, parfois, ...

Préposition (syntagme prépositionnel SP): introduit un complément : de, avec, dans, ...

Analyse de la phrase en syntagmes :

(Le petit chien de mon voisin)_{SN}(court très vite)_{SV}.

((Le petit chien)_{SN}(de mon voisin)_{SP})_{SN}((court)_V(très vite)_{SADV})_{SV}.

((Le petit chien)_{SN}(de (mon voisin)_{SN})_{SP})_{SN}((court)_V(très vite)_{SADV})_{SV}.

((Le_{DET} petit_{ADJ} chien_N)_{SN}(de_{PREP}(mon_{DET} voisin_N)_{SN})_{SP})_{SN}((court)_V(très_{ADV} vite_{ADV})_{SADV})_{SV}.

Remarque :

Un syntagme peut inclure un ou plusieurs autres syntagmes.

Pour décrire la phrase, il faut définir comment les syntagmes peuvent se combiner entre eux, c'est ce qu'on appelle : **règles syntaxiques**.

Est-ce que les mots imposent des contraintes sur l'ordre de la phrase ? (Syntaxe et lexique)

Analysons les exemples suivants :

- Michel est grand.
- Michel est fier.
- Michel est fier de son frère.
- *Michel est grand de son frère.

On voit que l'adjectif *grand* n'accepte pas de complément prépositionnel alors que l'adjectif *fier* accepte un tel complément. Ou encore:

- Sylvie dort.
- Sylvie prend le train.
- *Sylvie dort le train.
- *Sylvie prend. (*dans un contexte particulier uniquement*)

Ici, le verbe *dormir* ne prend pas de complément d'objet direct, tandis que *prendre* en exige un. Donc, de par leurs natures mêmes, les mots d'une langue impliquent un ensemble de combinaisons possibles.

C'est pour cela d'ailleurs que les verbes sont regroupés dans plusieurs sous-catégories :

- verbe intransitif : qui n'accepte pas de complément d'objet direct (COD) (dormir, marcher)
- verbe transitif ; qui accepte le complément d'objet direct (manger, boire)
- verbe bi-transitif : qui accepte un complément d'objet direct COD et un complément d'objet indirect COI (*donner qqch. à qqn*)

Le COD, le COI et le Sujet sont appelés aussi arguments du verbe. Ainsi, on a des verbes :

- à un argument (Sujet) = verbe intransitif (*Marie dort*)
- à deux arguments (Sujet, COD) = verbe transitif (*Marie mange une pomme*)
- à trois arguments (Sujet, COD, COI) = verbe bi-transitif (*Marie donne une pomme à Simon*)

Remarque :

Le même verbe peut appartenir à plusieurs catégories à la fois :

Marie mange. (verbe à 1 argument)

Marie mange une pomme. (verbe à 2 arguments)

Conclusion :

Les mots peuvent imposer un ordre syntaxique à la phrase.

9.5 Morphosyntaxe

Ce qui détermine l'appartenance d'un mot à une catégorie (nom, verbe, adverbe etc.), ce n'est pas tellement sa structure interne (qui peut être très diverse), mais le rôle qu'il joue dans la phrase. Ce qui veut dire que l'on est à la limite de la morphologie (qui étudie la structure interne des mots) et de la syntaxe (qui étudie les relations entre les mots dans une phrase).

Exemple :

Maison

Terminologie traditionnelle : Nom

Une belle maison

Mais :

Une tarte maison

Fonction : Adjectif, puisque *maison* modifie un autre nom : *tarte* ; de plus, il n'est pas précédé d'un déterminant, et n'est pas mis au pluriel (ne prend pas de « s ») :

Des tartes maison

Selon sa position dans la phrase, le même mot peut se comporter comme un nom ou comme un adjectif. Cela est l'objet d'étude de la **morphosyntaxe**.

La morphosyntaxe étudie les catégories grammaticales de mots selon la fonction qu'ils occupent dans une phrase.

9.6 Sémantique

D'un point de vue TAL, on s'intéresse au sens ; Quel est le sens d'un mot ? Quel est le sens d'une suite de mots (d'une phrase) ? Comment les sens de mots se combinent pour donner une signification à la phrase ?

La sémantique est l'étude scientifique de la signification.

Exemple :

Une phrase peut être syntaxiquement bien formée, et pourtant incompréhensible :
Les idées violettes sont à la mode.

La signification d'une phrase n'est pas facile à étudier. De multiples approches ont été proposées pour étudier la façon dont se construit le sens d'une phrase. L'une des plus simples consiste en une traduction des énoncés en **logique des prédicats**, dite aussi **logique du premier ordre** ou **logique classique**

Ainsi les énoncés :

Socrate est grec.
Les grecs sont des menteurs

Pourront être traduits par les formules suivantes :

$\text{grec}(\text{Socrate})$
 $\forall x \text{ grec}(x) \implies \text{menteur}(x)$

(\forall Quantificateur universel signifiant : quel que soit.)

A partir d'une telle représentation, l'ordinateur peut déduire de nouvelles propositions telles que : « menteur(Socrate) » et répondre donc correctement à la question « est-ce que Socrate est un menteur ? ». Ce type de logique est cependant très limité, car il ne porte que sur les valeurs de vérité des énoncés. On s'est très vite aperçu que des phrases comme : « *La bataille aura lieu demain* » ne peuvent être qualifiées ni de vraies ni de fausses au moment où elles sont prononcées. En effet, de nombreux phénomènes échappent à la logique classique tels que : le rapport au temps, les modalités (je peux, je dois etc.), les croyances (je crois que), les ordres (fais la vaisselle), les interrogations (est-ce qu'il fait beau ?). La logique classique ne peut pas, par conséquent, être l'unique solution pour représenter le sens en Traitement Automatique des Langues.

9.7 Pragmatique

La **pragmatique** étudie tout ce qui est au-delà du langage, qui appartient donc aux domaines des connaissances encyclopédiques ou résulte de la situation de communication.

Exemple (situation de communication):

Si on dit :

Pouvez-vous ouvrir la fenêtre ?

On ne veut pas généralement que la personne à qui on pose cette question nous réponde : *oui, je peux ouvrir* ou *non, je ne peux pas ouvrir*. On attend plutôt que l'interlocuteur ouvre la fenêtre. La question est donc un ordre.

Cependant, dans une situation particulière, cette question peut-être une vraie question, par exemple lorsque l'on s'adresse à une personne qui a le bras dans le plâtre.

On voit bien que l'interprétation de cette question dépend de la situation concrète dans laquelle elle a été prononcée.

Exemple (connaissances encyclopédiques):

Jean a évité le tigre, le singe aussi

Jean a évité le singe, le tigre aussi.

(Notre connaissance des rapports de force entre un humain, un singe et un tigre, nous permet de comprendre ces énoncés)

Exemple (connaissances générales) :

La pragmatique recouvre aussi tout ce qui est implicite (non-dit et pourtant évident) dans une situation. Quelles sont les connaissances qui nous permettent de comprendre un texte comme :

Pierre est allé au restaurant. Il a commandé un steak, mais il était froid. La serveuse ne s'est même pas excusée lorsqu'il a réclamé.

Connaissance 1 : *Pierre est allé au restaurant.*

→ Il y est entré, il s'est assis, il a probablement consulté le menu

Connaissance 2 : *Il a commandé un steak*

→ Quelqu'un est venu pour prendre la commande, quelqu'un lui a apporté un steak, il a commencé à manger

Connaissance 3 : *Mais il était froid*

→ Un steak est normalement chaud

Connaissance 4 : *La serveuse ne s'est même pas excusée*

→ Le fait que le steak soit froid avait déplu à Pierre

→ La coutume veut que l'on s'excuse quand quelque chose déplaît au client

Etc.

Une personne qui ne connaît rien aux restaurants, ne pourrait pas comprendre ce texte. Pourtant la plupart de ces connaissances sont indispensables pour la compréhension de cet énoncé. L'ordinateur, à la base, ne connaît rien aux restaurants. Comment faire alors pour qu'il interprète correctement un texte comme ce dernier ?

Résumé :

On peut imaginer le traitement des données linguistiques comme une suite de modules d'analyse opérant les uns après les autres. Il se base sur une **architecture séquentielle** ou **stratifiée** du traitement automatique des langues.

Les modules sont les suivants :

Si les données d'entrée sont orales, nous avons tout d'abord besoin d'un **module phonétique** qui retranscrit les sons en mots ;

Nous avons ensuite besoin d'un **traitement morphologique** capable d'associer des informations grammaticales à chaque mot considéré isolément ;

L'étape suivante consiste en un **traitement syntaxique** pour extraire les relations grammaticales des mots et des groupes des mots entre eux ;

Nous avons également besoin d'un **traitement sémantique** pour déterminer le sens des phrases ;

Nous avons recours enfin à un **traitement pragmatique** pour interpréter la phrase en fonction des connaissances générales sur le monde et de la situation de communication.

10 Exemple d'application

Analysons la phrase ci-dessous en passant par les différents niveaux sus-cités :

Les enfants mangent du bœuf.

Morphologie :

Découpage en catégories grammaticales :

Les enfants mangent du bœuf.

Det N V Det N

Découpage en morphèmes

Le (racine) + s (marque du pluriel)

Enfant (racine) + s (marque du pluriel)

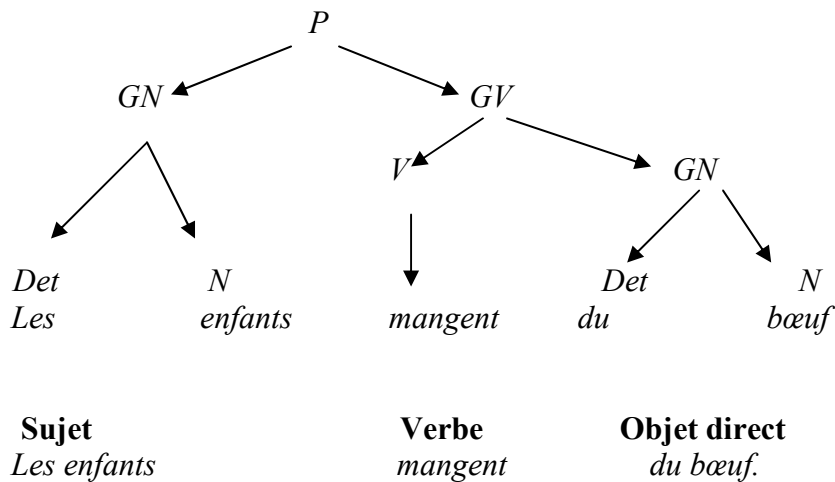
Mang (racine) + ent (marque du pluriel + présent)

Du (déterminant partitif)

Bœuf (racine)

Syntaxe :

Une représentation syntaxique par un arbre. Quelles sont les relations entre les différentes parties de discours ?



Sémantique :

On peut représenter la sémantique d'une phrase de diverses façons, mais pour l'instant, nous nous contenterons du sens des mots, d'une façon très simplifiée :

Les : /le+ pl./

Enfants : /enfant+ pl./ (être humain non-adulte)

Mangent : / manger + 3 pers.pl + présent/ (se nourrir)

Du : /du/ article partitif

Bœuf :

Soit --> /bœuf 1/ animal (Anglais = ox)

Soit --> /bœuf 2/ la viande de cet animal (Anglais = beef)

Solution :

La solution provient de la syntaxe ; la construction syntaxique (X (Article partitif) + Y(animal) impose que Y = viande de l'animal.

Quelle sont les autres informations qui nous permettent de comprendre cette phrase ?

* *Les fleurs mangent du bœuf.*

Seuls les humains et les animaux peuvent effectuer l'action de manger, donc un enfant est un humain.

* *Les enfants mangent le ciel.*

On ne peut manger que ce qui est comestible, donc le bœuf est comestible.

* *Les enfants mangent du jus.*

Ce que nous mangeons ne doit pas être seulement comestible, mais doit être aussi d'une consistance solide.

Conclusion :

Nous ne nous rendons pas compte du nombre important d'informations que nous possédons sur un seul mot car nous utilisons la langue d'une manière complètement spontanée. Pour

qu'une machine puisse déchiffrer la langue de façon correcte, il faut lui donner toutes ces informations.

Remarque :

En TAL, on ne sait plus très bien ce qui appartient à la sémantique d'un mot (sa signification) et ce qui appartient aux connaissances pragmatiques, au savoir qu'on a sur le monde.

D'une part, il y a la sémantique des mots, hors phrases, hors contexte :

Avocat : /fruit/ ou /un homme de loi/. D'autre part, il y a la sémantique de la phrase, qui restreint les possibilités d'interprétation d'un mot :

Mon avocat n'est pas venu à notre rendez-vous. /l'homme de loi/

J'ai mangé un avocat (Un fruit)

Cependant, nous avons aussi des phrases telles que :

- *Ton avocat est pourri ! (fruit ou l'homme)*
- *Mon avocat est merveilleux ! (fruit ou l'homme)*
- *J'arrête les avocats (fruit ou l'homme)*

11 Bibliographie

- Choi-Jonin Injoo, Delhay Corinne (1998), « Introduction à la méthodologie en linguistique », Presses Universitaires de Strasbourg
- Gary-Prieur Marie-Noëlle (1985), « De la grammaire à la linguistique », Armand Colin, Paris
- Guelpa Patrick (1997), « Introduction à l'analyse linguistique », Armand Colin, Paris
- Moeschler Jacques, Auchlin Antoine (1997), « Introduction à la linguistique contemporaine », Armand Colin, Paris
- Petiot Geneviève (2000), « Grammaire et linguistique », Armand Colin, Paris
- Siouffi G., Van Raemdonck (1999), « 100 fiches pour comprendre la linguistique », Bréal, Rosny
- Grevisse, Gosse (1995), « Nouvelle grammaire française », De Boeck Duculot, Louvain-la-Neuve
- Grevisse, Gosse (1995), « Nouvelle grammaire française. Applications », De Boeck Duculot, Louvain-la-Neuve
- Grevisse, Gosse (1995), « Nouvelle grammaire française. Corrigé des applications », De Boeck Duculot, Louvain-la-Neuve

12 Exercices

Exercice 1 : Parties du discours

Identifiez la partie du discours à laquelle appartient chacun des mots des phrases suivantes (nom, pronom, déterminant, adjectif, verbe, adverbe, préposition, conjonction, interjection).

J'arrivai dans un lieu très étroit et obscur, mais parfumé d'une odeur triste de violettes étouffées.

Ô mon aimé, dit-elle, je serais malhabile à t'enseigner comme autrefois.

Exercice 2 : Noms

Quelles problématiques pose le genre des noms en italique ?

Indice : Regardez le genre et l'accord.

1. Jean-Michel est une excellente *recrue* pour notre équipe. — 2. Ma voisine est une *pimbêche*. — 3. Paris avait environ 100 000 *habitants* à la fin du XIII^e siècle. — 4. La poule est un *gallinacé*. — 5. Ma mère a été *témoin* de l'accident. — 6. Dans la pièce, Isabelle jouait le rôle d'un *page*. — 7. S'il n'y a pas d'enfant, les biens du ménage reviendront à l'*époux* survivant. — 8. Le capitaine avait à son service une *ordonnance* toute dévouée.

1. Elle était à la fois le *souffleur* de la troupe et l'habilleuse de ses dames. (C. Lemonnier.) — 2. Cette transformation printanière, qui a fait de la *linotte* mâle un des plus beaux oiseaux de nos climats, ne touche guère sa femelle. (J. Delamain.) — 3. Marcelle Chaumont est l'un de nos *couturiers* dont la personnalité se renouvelle dans un style qu'on reconnaîtrait entre mille dans une présentation anonyme. (Dans les *Nouvelles littéraires*.) — 4. J'aurai le préfet, j'aurai les deux maréchaux, j'aurai le bâtonnier, j'aurai l'*ambassadrice* de Belgique, et peut-être l'ambassadeur ! (Hériat.) — 5. Née *romancier*, je fais des romans, c'est-à-dire que je cherche, par les voies d'un certain art, à provoquer l'émotion. (G. Sand.) — 6. Ah ! Suzon, vous êtes une bath copine, mon *petit*. (G. Duhamel.)

Exercice 3 : Adjectifs

Relever les adjectifs et indiquer leur fonction (épithète ou attribut) ?

1. L'homme sérieux est dangereux ; il est naturel qu'il se fasse tyran. (S. de Beauvoir). — 2. Rien ne nous rend si grands qu'une grande douleur. (Musset.) — 3. Le silence suspect du paysage était rendu plus sensible par les arrêts brusques et les reprises hésitantes de la pluie. (J. Gracq.) — 4. Très craintive, la jument levait la patte dès que Meaulnes voulait la toucher et grattait le sol de son sabot lourd et maladroit. (Alain Fournier.) — 5. À côté de mon assiette, je trouvai un œillet (...). L'usage, pourtant aussi nouveau pour moi, me parut plus intelligible quand je vis tous les convives masculins s'emparer d'un œillet semblable qui accompagnait leur couvert et l'introduire dans la boutonnière de leur redingote. (Proust.)

Exercice 4 : Verbes

Relever les verbes intransitifs et les verbes transitifs, en indiquant les compléments de ceux-ci.

1. Ce mur penche. — 2. Nous ne manquerons pas à nos promesses. — 3. Mon travail avance. — 4. Cette porte ouvre sur la rue. — 5. Ouvrons notre cœur à la pitié. — 6. Baisse la tête, fier

Sicambre. — 7. J'ai avancé ma besogne. 8. Il manquera son train. 9. Le baromètre baisse. — 10. Consultez un avocat. — 11. Vous racontez vos voyages. — 12. Le feu prend mal. — 13. Les médecins consultent sur sa maladie. — 14. Tu racontes avec agrément. 15. Penchez le corps en avant. — 16. Le soir tombe.

1. Pierre qui roule n'amasse pas mousse. — 2. Que voulez-vous pour votre dîner ? — 3. Les chiens aboient, la caravane passe. — 4. Je me rappelle que vous êtes déjà venu chez nous. — 5. Corrigez les fautes que vous avez faites. — 6. Le Parthénon mesure environ soixante-dix mètres de longueur. 7. Le géomètre mesure le jardin avec des instruments de précision. — 8. Les assaillants ont pris la ville sans coup férir.

Exercice 5 : Prépositions et articles

Distinguer *de* préposition et *de* article ou partie d'article.

1. La pipe *de* l'oncle Jules fait *de* la fumée comme une locomotive à vapeur. — 2. Il n'y a pas *de* fumée sans feu ! — 3. Celui qui n'a pas *de* patience essuiera *de* nombreux mécomptes. — 4. Yachts : bateaux *de* plaisance possédés par *de* riches amis qui parlent souvent *de* vous inviter à bord. Le mauvais temps peut permettre *d'*échapper à ce genre *d'*invitation. Mais le plus souvent le mauvais temps ne survient qu'une fois la haute mer atteinte. (Daninos.)

Exercice 6 : Morphologie

Identifiez les morphèmes dans les listes de mots suivantes. Dégagez la racine et les affixes.
acheteur, fleur, danseur, acteur
cuisinette, cigarette, toilette, fillette

Exercice 7 : Morpho- sémantique

Quels sont les morphèmes en français qui désignent « celle qui fait l'action indiquée par le verbe » ? Trouvez en au moins deux. Donnez des exemples.

Exercice 8 : Lexicologie et morphologie

Déterminez lesquels des syntagmes en italique sont des mots composés. Justifiez votre réponse par de différents tests :

Nous ignorons l'existence de cette *boîte noire*.

Toutes mes affaires sont dans cette *boîte verte*.

Je *donne ma langue au chat*.

Je *donne des croquettes au chat*.

C'est *simple comme bonjour*.

Exercice 9 : Morphosyntaxe

a) Analysez les mots en italique. A quelle partie du discours appartiennent-ils ? Justifiez votre réponse.

On peut apporter son *manger* tous les jours.

Diable, est-ce que tu vas *manger* tout ça ?

Quel *diable*, ce *petit* !

Tu m'agaces avec tes *mais* continuels !

J'ai envie d'y aller, *mais* j'hésite à *cause de* Paul.

Ce mec est la *cause* de tous tes ennuis.
Ce piano joue complètement *faux* dans les *aigus*.
Vous avez fait un *faux* numéro.
Il a été condamné pour utilisation de *faux*.
Les enfants marchaient *devant*.
Il s'est arrêté *devant* la porte du garage.
Pierre est un *timide*.
Tu dormiras *en haut* et moi en bas.
Elle a enlevé *le haut* de son maillot de bain.
Peux-tu poser le livre *en haut de* l'étagère ?
C'est un *m'as-tu vu*.

b) Peut-on cantonner des mots dans une seule catégorie grammaticale ? Est-ce qu'il existe une catégorie plus naturelle qu'une autre pour un mot ?

Exercice 10 : Syntaxe et sémantique

Expliquez le sens des adjectifs épithètes.

1. Un homme seul; un seul homme. — 2. De l'eau pure; une pure calomnie. — 3. Une nouvelle vraie; du vrai marbre. 4. Un repas maigre; un maigre repas. — 5. Son enfant propre; son propre enfant. — 6. Une pomme verte ; une verte vieillesse. — 7. Un visage triste ; un triste personnage.

CM 2

Complexité du langage naturel et TAL

Descriptif du cours

Ce cours vise à donner aux étudiants de Master 1 en Didactique des Langues Etrangères des outils nécessaires à la formalisation de la langue en vue de construire des applications en TAL. La complexité du langage naturel n'est certainement pas une évidence pour tout locuteur natif. Lorsqu'un individu manipule une langue dans le but de communiquer une information, il n'est souvent pas conscient de toutes les connaissances linguistiques qu'il mobilise en accomplissant son acte du langage. L'acte devient ainsi spontané, mécanique pour ne pas dire banal ; les compétences linguistiques dont il dispose dans sa langue maternelle lui permettent de tenir des conversations courantes avec une facilité déconcertante. Bien plus, sa maîtrise de la langue est telle qu'il est capable, à travers des éléments du langage, de faire intervenir les mécanismes de l'inférence et de raisonnement dans l'interprétation des énoncés. Qu'il soit en position d'émetteur ou de récepteur du message, tout usager de la langue arrive facilement à transcoder et à décoder un message sans pour autant être en mesure d'expliquer ces deux processus d'encodage et de décodage. A cet effet, Le traitement automatique des langues ouvre une perspective nouvelle sur les langues naturelles et porte un regard formel sur leur mode de fonctionnement. L'objectif premier est d'arriver à sortir de la complexité du langage naturel :

- En traquant les phénomènes langagiers récurrents d'une langue ou communs à plusieurs langues
- En modélisant ces phénomènes sous forme de données linguistiques compatibles au langage informatique

En effet, la complexité infinie du langage naturel n'apparaît que lorsqu'on s'intéresse à son mode de fonctionnement et à toutes les composantes qui le constituent. Parmi ces composantes et qui n'est pas des moindres d'ailleurs, la sémantique pierre angulaire de toute communication est considérée comme une des pierres d'achoppement du traitement automatique des langues. Ce cours se propose donc de faire un zoom sur les problèmes que pose une composante particulière, celle de la sémantique.

Objectifs généraux du cours

Le cours réparti sur deux chapitres. Le premier chapitre, intitulé « L'ambiguïté des langues naturelles et TAL » a pour objectif de cerner la notion d'ambiguïté dans le langage naturel sous le prisme du traitement automatique des langues. En effet, quoi de plus pédagogique pour comprendre l'opacité du langage naturel que l'application des méthodes de

formalisation, épine dorsale du traitement automatique des langues. L'étudiant sera en capacité de comprendre et de prendre conscience des problèmes que pose le sens et la signification dans le domaine du TAL. Le deuxième chapitre sous le titre « Classification des ambiguïté d'un point de vue TAL » mettra au jour plusieurs types d'ambiguïtés aussi diverses que complexes, auxquelles se heurte le traitement automatique des langues.

Objectifs spécifiques du cours :

- Définir d'une manière précise ce qu'est l'ambiguïté en langue naturelle,
- Etre capable de faire le départ entre l'ambiguïté en langue naturelle et les autres phénomènes relevant de l'opacité sémantique,
- Pouvoir classier les différentes ambiguïtés liées aux différents niveaux de la langue (phonologique, lexical,...),
- Comprendre dans quelle proportion les ambiguïtés naturelles posent problème au TAL et pouvoir mettre en place des formalismes capables de les surmonter

Chapitre 1

L'ambiguïté des langues naturelles et TAL

Contenu du cours

1	Qu'est-ce qu'une ambiguïté ?.....	4
2	Ambiguïté et autres phénomènes linguistiques	5
2.1	La sous-détermination du sens :	5
2.1.1	Le non-dit.....	5
2.1.2	La généralité du sens.....	6
2.1.3	Le sens flou	6
2.2	La surdétermination du sens	7
2.2.1	Le sens implicite	7
2.2.2	Les présupposés	7
2.2.3	Les sous-entendus	8
2.2.4	Les jeux de mots	8
3	Types d'ambiguïtés.....	9
3.1	Réelle, effective	9
3.2	Accidentelle, virtuelle.....	9
3.3	L'équivoque réelle	11
4	Les ambiguïtés artefacts	12
5	Le rôle du contexte	12
6	Le rôle des facteurs extralinguistiques :	13
7	La production volontaire des ambiguïtés.....	13
8	Exercices.....	15

1 Qu'est-ce qu'une ambiguïté ?

On parle de l'ambiguïté dès qu'un élément peut recevoir au moins deux interprétations différentes :

Exemple 1 :

- A l'unique forme phonique, correspondent plusieurs formes graphiques :
/ so /

- Sot (« bête »)
- Saut (« un bond »)
- Sceau (« cachet, empreint »)
- Seau (« récipient »)

Exemple 2 :

- A l'unique forme graphique correspondent plusieurs significations :
– Bière :

- « boisson alcoolique »
- « cercueil »

Exemple 3 :

- A une seule phrase correspondent plusieurs structurations syntaxiques sous-jacentes :
J'ai fait porter les fleurs à Marie.

- ✓ N1 faire V-infinitif N2 à N3 (ambigu)
 - N1 faire que X V N2 à N3 (« J'ai fait que quelqu'un porte les fleurs à Marie », « J'ai fait porter les fleurs à Marie par qqn »)
 - N1 faire que N3 V N2 (« J'ai fait que Marie porte les fleurs », « J'ai obligé Marie à porter les fleurs »)

Pour qu'il y ait ambiguïté linguistique, il faut que deux conditions soient remplies :

- Il faut d'une part que les différentes significations en jeu soient prédictibles dans la langue, c'est-à-dire que l'analyse linguistique doit pouvoir en rendre compte :
 - Les deux significations de 'bière' : 'boisson' et 'cercueil' doivent être consignées dans un dictionnaire de français.

- Il faut aussi que les différentes significations soient disjointes ou distinctes du point de vue de la dénotation (c'est-à-dire qu'il faut qu'elles renvoient à des objets différents dans la réalité).
 - 'Bière – boisson' ne renvoie pas du tout au même référent que 'bière – cercueil'.
 - 'Aiguille' d'une montre ne renvoie pas du tout au même référent que 'aiguille' à tricoter.

En résumé :

Les significations associées à une forme ambiguë doivent être prédictibles et mutuellement exclusives : Ceci entraîne l'idée du choix : nous sommes obligés de choisir l'une ou l'autre des significations.

2 Ambiguïté et autres phénomènes linguistiques

Il faut distinguer l'ambiguïté des autres phénomènes avec lesquels elle peut être confondue :

- les phénomènes de sous-détermination du sens (le non-dit, la généralité du sens, le sens flou).
- les phénomènes de surdétermination du sens (les présupposés, les sous-entendus et les jeux de mots).

Dans l'usage courant, « l'ambigu » possède le sens de 'vague', 'incertain', 'obscur' ; on dit par exemple de quelqu'un qu'il s'est contenté d'une « réponse ambiguë » pour signifier que sa réponse n'est pas assez précise.

L'ambiguïté linguistique n'est pas un phénomène d'imprécision, mais un phénomène qui implique le choix.

2.1 La sous-détermination du sens :

Une expression linguistique a le sens indéterminable lorsqu'il est impossible d'associer à sa forme (ou à l'ensemble des formes qui la composent) une ou plusieurs significations déterminées.

Exemple :

- Le silence vertébral indispose la voile licite.

2.1.1 Le non-dit

le non-dit est l'ensemble des informations qui ne sont ni signifiées en langue par une expression linguistique, ni inférables à partir de celle-ci, mais sur lesquelles un récepteur peut s'interroger lorsqu'il décode l'expression.

Exemple 1 :

- Il est bien tôt.

'Bien tôt' pour faire quoi ? Le choix ne correspond pas à un choix linguistique (entre un sens et un autre, par exemple), mais à un choix 'dans la réalité' :

- Il est trop tôt pour se lever / pour travailler/ pour partir/ pour manger, etc.

Exemple 2 :

-Pour faire de la place, j'ai donné tous mes livres.

A qui? Le verbe "donner" présuppose l'existence d'un bénéficiaire (on donne qqch à qqn), donc on peut s'interroger sur l'identité du bénéficiaire.

En résumé :

Le non-dit résulte du fait que pour décrire une situation, on ne peut être totalement exhaustif : on choisit les éléments que l'on veut communiquer et il reste toujours des éléments qu'on n'évoque pas. Ces éléments ne sont pas ambigus : ils restent simplement ouverts à d'éventuelles précisions référentielles.

2.1.2 La généralité du sens

Il s'agit des phrases dont la dénotation (la dénotation = ce qu'elles décrivent dans la réalité) est plus ou moins large ou étroite.

Exemple :

- J'ai été mordu par un chien.

'Un chien' est un mot à la dénotation large, très générale, mais pas ambiguë. Bien évidemment, dans cette phrase je ne précise pas la race du chien ni ses particularités (Sa couleur ? Sa taille ? Son âge ? etc.), parce que je choisis de ne pas le faire.

2.1.3 Le sens flou

Une expression linguistique a un sens flou lorsque la signification qui lui est associée couvre une amplitude variable.

Exemple :

- Pierre est grand.

Cette expression possède un sens univoque ('Pierre a une taille qui dépasse la moyenne'), mais ne permet pas de savoir qu'elle est la taille exacte de Pierre (1m40 (mais peut-être plus ?) si c'est un garçon de 8 ans ; 1m 90 si c'est un adulte etc.).

Résumé:

Tous les phénomènes qui viennent d'être évoqués partagent un point commun : le sens associé à chaque expression est univoque, mais demeure sous-déterminé d'une manière ou d'une autre : l'interprétation reste ouverte, dans la mesure où le sens peut toujours être précisé, voire rectifié.

2.2 La surdétermination du sens

Elle renvoie d'une part au sens implicite composé des présupposés et des sous-entendus et d'autre part aux jeux de mots.

2.2.1 Le sens implicite

Le sens implicite se présente comme un sens second, plus ou moins caché, que le récepteur peut (ou doit) inférer à partir d'une expression linguistique, en plus du sens explicitement exprimé. Contrairement au non-dit, qui est absent du message, l'implicite est présent dans le message, mais sous des modalités très particulières. On distingue classiquement deux types d'implicites : les présupposés et les sous-entendus

2.2.2 Les présupposés

Le présupposé est partie intégrante du sens des énoncés. Il est inféré indépendamment du contexte contrairement aux sous-entendus qui en sont tributaires.

Exemple :

- Jean a arrêté de fumer -- > présupposé : Jean fumait avant.
- Le fils de Pierre a grandi -- > présupposé : Pierre a un fils
-

Ces phrases n'ont rien d'ambigu : elles ne possèdent pas deux significations explicites disjointes et exclusives, bien au contraire, elles conjoignent plusieurs contenus sémantiques : le premier, explicite et formulé et le deuxième, implicite, présenté comme connu et allant de soi.

2.2.3 Les sous-entendus

Un sous-entendu est un contenu implicite qui n'est pas entraîné automatiquement par la présence d'une expression linguistique, mais qui peut être inféré à partir de sa forme ou de sa signification dans certaines conditions.

Exemple :

- La poubelle est pleine ! -- > Peut vouloir dire : « s'il te plait, descend la poubelle ! » (Le constat se réinterprète comme une demande)
- Cette robe te va à ravir, c'est fou comme elle t'amincit !--> Peut vouloir dire « si je te dis que tu es belle quand tu parais plus mince que tu ne l'es, c'est pour suggérer combien en réalité tu es grosse » (le compliment se réinterprète comme une critique)

2.2.4 Les jeux de mots

Les jeux de mots cumulent volontairement différentes significations (ou structures) d'un mot, ils les superposent pour obtenir un certain effet.

Exemple 1 :

- Ces crêpes, je vais t'expliquer comment les fourrer et surtout à qui...
 - Fourrer les crêpes = les garnir de qqch (de chocolat, de confiture, etc.)
 - Fourrer les crêpes à qqn. = les donner (de force) à qqn

Exemple 2 :

- Votre corps est un précieux capital, placez-le sur un matelas Lité !
 - L'énoncé joue sur le double sens du verbe 'placer ' et la présence du mot 'capital':
 - ✓ Placer un objet quelque part = l'y mettre
 - ✓ Placer son argent = l'investir pour le faire fructifier

Conclusion:

- Les expressions sous-déterminées sont univoques (possèdent un seul sens) bien que leurs significations restent ouvertes.
- Les expressions surdéterminées sont plurivoques (possèdent plusieurs sens), mais leurs différentes significations se cumulent.
- Les expressions ambiguës sont plurivoques aussi, mais leurs différents sens s'excluent mutuellement.
- Les phénomènes de sous-détermination et de surdétermination appartiennent plutôt au domaine du discours, c'est-à-dire de l'usage de la langue par des sujets parlants. Le phénomène d'ambiguïté, quant à lui, relève au contraire du système de la langue elle-même (et non de son utilisation).

3 Types d'ambiguïtés

Pour un récepteur humain, le problème de l'ambiguïté ne se pose pas souvent, dans la mesure où le contexte linguistique sélectionne souvent une des interprétations :

- « La soif les poussa à commander deux bières bien fraîches au comptoir du bar. » (boisson)
- « Les croque-morts déposèrent la bière dans le corbillard » (cercueil)

Il existe cependant des cas où le contexte linguistique autorise plusieurs interprétations :

- « Comme il faisait chaud le jour de l'enterrement, on sortit la bière » (boisson ou cercueil ?)

3.1 Réelle, effective

Lorsqu'un mot, structure etc. sont vraiment ambigus pour un humain, c'est-à-dire qu'un récepteur humain ne peut choisir entre deux interprétations (souvent hors contexte).

Exemples :

- « orange » (couleur ou fruit ?)
- « Comme il faisait chaud le jour de l'enterrement, on sortit la bière » (Boisson ou cercueil ?)
- « Pour que la cérémonie soit réussie, j'ai fait porter les fleurs à Marie » (« pour Marie (Marie = bénéficiaire) »? ou « par Marie (Marie = agent) ?»)

Normalement, un contexte linguistique plus large permet de lever cette ambiguïté aussi :

- « Pour que la cérémonie soit bien réussie, j'ai fait porter les fleurs à Marie. »
 - Elle était très contente de les recevoir. (Marie = bénéficiaire)
 - Pour ne pas trop la fatiguer, j'ai donné de gros paquets à Antoine. (Marie = agent)

3.2 Accidentelle, virtuelle

Lorsque l'ambiguïté existe potentiellement, mais elle est effectivement levée par le contexte linguistique ; les formes linguistiques sont presque toutes potentiellement ambiguës.

Exemple 1 :

- Potentiellement, le mot 'orange' possède deux interprétations « fruit » et « couleur » alors que dans « une robe orange » 'orange' ne peut signifier que « couleur ».

Les récepteurs humains lèvent très facilement les ambiguïtés accidentelles dès lors qu'ils ont accès au contexte, ce qui n'est pas le cas des ordinateurs : un ordinateur n'a pas accès au contexte de la même manière que les humains. Si on ne lui donne pas les éléments extralinguistiques, il ne saura pas quelle interprétation choisir. De ce fait, la levée des ambiguïtés s'avère très difficile pour les systèmes automatiques.

De plus, la levée d'ambiguïté dépend souvent de la taille du contexte auquel on peut accéder :

Exemple 2 :

- Critique -- > Adjectif, Nom ou Verbe?
- Un rapport critique -- > Det + Nom + Adj ou Det + Nom + Verbe
- Un rapport critique le plan de la banque -- > Det + Nom + Verbe

Le contexte le plus souvent utilisé pour lever les ambiguïtés est le contexte phrastique (de la phrase).

Toutefois, il arrive que la phrase ne soit pas suffisante comme contexte :

Exemple 3:

- « Quand elle revient, Marie était de mauvaise humeur. »
 - Marie sortit faire des courses. Les magasins étaient bondés, et les courses durèrent longtemps. Quand elle revient, Marie était de mauvaise humeur -- > (elle = Marie)
 - Jeanne sortit faire des courses. Les magasins étaient bondés, et les courses durèrent longtemps. Quand elle revient, Marie était de mauvaise humeur. -- > (elle = Jeanne)

Remarque 1

Bien que cela ne pose pas des problèmes à des humains, les ordinateurs gèrent mal les contextes qui dépassent les contextes phrastiques. Ainsi, une ambiguïté virtuelle, qui n'est qu'une configuration provisoire pour un récepteur humain, construite à un moment donné de l'analyse et levée par la suite, peut être très difficile à résoudre pour un système automatique.

Remarque 2 :

Certaines ambiguïtés linguistiquement effectives sont levées en discours :

Exemple 4:

- « C'est en principe le 1^{er} Juin que devrait être discuté au Parlement le projet de loi sur la famille du ministre des affaires sociales, Simon Veil » (le Monde)
 - ✓ (« le projet de loi du ministre » ou « la famille du ministre » ?)
- « Les députés doivent débattre cet après-midi du plan gouvernemental de relance de la croissance d'Alain Juppé. » (France-Inter)
 - ✓ (« plan d'Alain Juppé » ou « croissance d'Alain Juppé » ?)

Dans d'autres cas, si nous n'avons pas suffisamment d'informations, nous pouvons être incapable de lever l'ambiguïté en discours.

Exemple 5 :

- J'ai rencontré un professeur de droit anglais.
 - ✓ Nous pouvons être incapable de décider à quoi est rattaché l'adjectif « anglais » (est-ce le professeur ou le droit) ; dans ce cas précis, on parle de « **l'équivoque réelle** ».

3.3 L'équivoque réelle

L'équivoque réelle renvoie à une situation dans laquelle une ambiguïté linguistique effective ne peut pas être levée en discours, faute d'informations suffisantes.

La distinction entre l'ambiguïté réelle, ambiguïté virtuelle et l'équivoque réelle, nous permet de distinguer différents types de récepteurs :

Un récepteur humain : un récepteur humain n'est arrêté dans son analyse que par les équivoques réelles ; les ambiguïtés virtuelles et même les ambiguïtés réelles levées en discours ne lui posent pas de problèmes (il ne les remarque même pas) ;

Le récepteur machine : le programme automatique achoppe déjà sur les ambiguïtés virtuelles, et même sur les **ambiguïtés artefacts** (qui ne sont même pas des ambiguïtés virtuelles).

4 Les ambiguïtés artefacts

Les ambiguïté artefacts sont des ambiguïté créés par le système automatique lui-même et dues aux règles linguistiques qui lui ont été implémentées. Imaginons qu'un concepteur d'analyseurs morphosyntaxiques introduit une règle dans son programme pour décrire les noms composés de type : N+N (jupe-culotte, visite éclair, etc.). Si on lui donne ensuite à analyser la phrase : « le valet dessert la terrasse », l'analyseur morphosyntaxique la trouvera ambiguë puisqu' il lui associera deux interprétations différentes :

- La première interprétation, correcte : Det (le) N (valet) V (dessert) Det (la) N (terrasse)
- La deuxième interprétation, erronée, basée sur la règle N+N (ambiguïté artefacts) : Det (le) N (valet) N (dessert) Pron (la) V (terrasse)

5 Le rôle du contexte

Dans la majorité des cas, le contexte aide à lever les ambiguïtés dans la phrase.

Exemple :

« sol » : « plancher » ou « note de musique »

- (1) Mégots et papiers gras jonchaient le *sol*.
- (2) C'est un *sol* dièse qu'il faut jouer.

Le contexte peut-être antérieur ou postérieur :

Antérieur : il est situé à gauche de l'expression ambiguë. Ce contexte nous permet d'anticiper sur l'interprétation de la phrase. Comme dans l'exemple 1, la suite « Mégots et papiers gras jonchaient le » situé à gauche du mot « sol » permet de trancher sur le sens à donner à ce mot qui est celui de « plancher ».

Postérieur : il est situé à droite de l'expression ambiguë. Dans l'exemple 2, lorsque l'expression ambiguë arrive, aucune interprétation n'est présélectionnée, et on peut très bien imaginer une suite autre que celle de « ...dièse qu'il faut jouer » comme par exemple : « ..dallé qu'il faut poser dans la cuisine ».

6 Le rôle des facteurs extralinguistiques :

La connaissance de la situation de production du message ainsi que de son émetteur, nous permet de le désambiguïser.

Exemple 1 (entendu à la radio) :

- Le problème c'est que la population aide la police. (aide = subjonctif ou indicatif ?)
 - ✓ Si c'est un préfet commentant les émeutes : Le problème c'est qu'il faut que la population aide la police. (subjonctif)
 - ✓ Si c'est un jeune d'une banlieue défavorisée : Le problème c'est que la population a l'habitude d'aider la police. (Indicatif)

Des fois, il n'est pas nécessaire d'éliminer une des interprétations, puisque dans les faits, elles se neutralisent :

Exemple 2 :

- Bonnet de coton blanc -- > Peu importe si c'est le bonnet ou le coton qui soit blanc, le résultat est le même.

Il arrive que les facteurs extralinguistiques démultiplient les interprétations possibles d'un message, ceci ne résulte pas d'une ambiguïté liée à la langue, mais d'une multitude de référents possibles.

Exemples :

- Il habite Valence. (en France ou en Espagne ?)
- J'ai parlé à Marie. (laquelle de Marie, si j'en connais plusieurs ?)
- Pourquoi n'irais-tu pas chez les voisins du dessus? (ceux de gauche sur le palier ou ceux de droite ?)

Les exemples suscités ne sont pas considérés comme des ambiguïtés linguistiques. Afin de trouver la bonne interprétation, nous ne faisons pas appel à nos connaissances linguistiques mais plutôt à des connaissances liées à une situation réelle : Combien de Marie ai-je le plaisir de connaître ? Combien ai-je de voisins etc. ?

7 La production volontaire d'ambiguïtés

Dans certaines situations, on peut chercher à créer l'ambiguïté volontairement. Ceci est particulièrement vrai lorsqu'il s'agit de jeux de langage (jeux de mots, devinette, blagues etc.).

Exemples :

- L'homme descend du singe, et le singe descend de l'arbre -- > (double sens du verbe 'descendre')
- Comment appelle-t-on un chien sans pattes ? On ne l'appelle pas, on va le chercher. -- > (double sens du verbe 'appeler')

8 Exercices

Exercice 1

Quels phénomènes linguistiques sont mis en évidence par les dialogues suivants ?

a)

A : Marie a adapté un enfant.

B : Une fille ou un garçon ?

b)

A : Marie a adapté un enfant.

B : Quand ?

c)

A : Pierre a menti.

B : A qui ?

d)

A : Aujourd'hui, j'ai d'abord fait les courses, puis j'ai bu un coup avec une amie.

B : Avec qui ?

e)

A : L'eau était trop froide pour qu'on s'y baigne.

B : Elle était à combien ?

Exercice 2

Quels présupposés et/ ou sous-entendus peut-on déduire des phrases/ dialogues suivantes ?

a) *A : Voulez-vous du café ?*

B : Non, merci, je commence tôt demain.

b) *Jean continue d'embêter sa sœur.*

c) *Jean ne déteste pas le vin.*

d) *Je démissionne dès que j'obtiens la confirmation de mon nouvel employeur.*

e) *Jacques L. aimerait bien que son fils cesse de lui emprunter sa voiture. Doit-il changer de fils ? (pub pour Renault)*

Exercice 3

Sur quelles significations ou structures sont construits les jeux de mots suivants ?

a) « *Le béton est armé, pourquoi pas vous?* » (un tag)

b) « *Je me casse le cul pour pas qu'il tombe.* » (Claude Sarraute sur sa lutte contre le vieillissement)

c) « *Un écologiste mal léché* » (France Inter)

d) « *Agfa, le grain de beauté* » (pub pour une marque de pellicules)

e) « *Cité des Francs-Moisin : le restaurant des femmes du monde* » (Le Monde sur un restaurant ouvert par des femmes immigrées)

Exercice 4

Combien de sens possède le mot 'salade' ? Comment peut-on distinguer ces différents sens ? Appuyez votre raisonnement sur les phrases suivantes :

- a) *Luc sème de la salade.*
- b) *Luc repique de la salade.*
- c) *Luc récolte de la salade.*
- d) *Luc mange de la salade.*

Exercice 5

Est-ce que les phrases suivantes sont réellement ambiguës ? Donnez deux lectures à chaque phrase réellement ambiguë.

- a) *Elle porte une petite robe noire.*
- b) *La lampe est allumée.*
- c) *Les invités entendaient le bruit de la fenêtrée.*
- d) *Je l'ai quitté joyeux.*
- e) *Laure amuse Alain.*
- f) *Nadine couvre la corbeille de fleurs.*
- g) *Richard et Sylvie sont mariés.*
- h) *La petite brise la glace.*

Exercice 6

Même si les phrases suivantes ne sont pas réellement ambiguës, quels problèmes peuvent poser les mots en italique à un récepteur machine ? Quels sont les éléments qui permettent à un récepteur humain de les désambigüiser rapidement ?

- a) *Les invités* ont entendu un bruit.
- b) *La glace fond* rapidement.
- c) J'ai *un* problème avec ce texte.
- d) Sa *femme* vient de partir.

Chapitre 2

Classification des ambiguïtés d'un point de vue TAL

Contenu du cours :

1	Ambiguïtés phonologiques	2
2	Problèmes de segmentation du texte	2
2.1	Ambiguïtés des signes de ponctuation.....	2
2.2	Ambiguïtés morphologiques.....	2
2.2.1	Les mots polycatégoriels.....	3
3	Ambiguïtés lexicales.....	5
4	Ambiguïtés syntaxiques (ou structurales).....	6
4.1	Ambiguïtés liée à la segmentation de la phrase.....	6
4.2	Ambiguïtés syntaxiques : reconnaître les fonctions des syntagmes	10
5	Ambiguïtés sémantiques : le calcul des relations	10
6	Ambiguïtés pragmatiques	11
6.1	La connaissance du monde, le bon sens	11
6.2	Difficulté de calcul des valeurs référentielles.....	12
7	Conclusion :	13
8	Bibliographie	13
9	Exercices.....	14

1 Ambiguïtés phonologiques

un mot ou une phrase peuvent avoir plusieurs interprétations à l'oral :

Exemple 1 :

- pluriel ou singulier ?
- ✓ leur (s) livre(s) (pluriel ou singulier ?)
 - m'a plu
 - m'ont plu

Le traitement ne peut pas être linéaire car les informations qui désambiguisent le syntagme « leur (s) livre (s) » sont placés à la suite de ce syntagme (contexte postérieur).

Exemple 2 :

les homonymes : à l'unique forme phonique correspondent plusieurs formes graphiques

- /mɛtR / : maître, mètre, mettre
- /ku/ : cou, coup, coût

Hors contexte, il est évidemment impossible de choisir parmi les homonymes. Dans une phrase, cela devient possible pour les récepteurs humains et difficile pour un ordinateur.

Exemple 2 :

- Les mûres sont mûres le long des murs.

2 Problèmes de segmentation du texte

Pour comprendre une phrase, il faut tout d'abord comprendre les unités qui la composent. Or, ce n'est pas toujours facile de réaliser une segmentation fiable.

2.1 Ambiguïtés des signes de ponctuation

Le « . » (point) est (potentiellement) ambigu en français :

- Il peut signifier une fin de phrase : Maurice Gary est l'auteur de ce roman.
- Il peut indiquer une abréviation : M. Gary est l'auteur de ce roman.

2.2 Ambiguïtés morphologiques

L'analyse morphologique doit nous fournir une décomposition des mots en morphèmes et en catégories morphosyntaxiques. Cependant, ceci n'est pas toujours facile puisque un mot peut avoir plusieurs catégories morphosyntaxiques.

(Rappel des catégories morphosyntaxiques : nom, prénom, déterminant, verbe, adjectif, adverbe, conjonction, interjection).

Si nous prenons un mot isolé, il peut souvent recevoir plusieurs interprétations morphologiques :

– portes = port (er), base verbale + es (2ème pers. sing. prés. ind.)

– portes = porte (N) + s (plur)

En définitif, le mot « porte » peut-être soit nom ou verbe.

2.2.1 Les mots polycatégoriels

On appelle polycatégoriels les termes qui peuvent prendre plusieurs catégories morphosyntaxiques.

Ceci peut être encore plus compliqué comme dans l'exemple 1 ci-dessous:

souris = souri (re) V + s (1 ^{ère} pers. Sing. prés. ind.)	<i>Je souris à la dame.</i>
souris = souri (re) V + s (2 ^{ème} pers. Sing. prés. ind.)	<i>Tu souris à la dame.</i>
souris = souri (re) V + s (1 ^{ère} pers. Sing. passé simple.)	<i>Je souris à la dame</i>
souris = souri (re) V + s (2 ^{ème} pers. Sing. passé simple.)	<i>Tu souris à la dame</i>
souris = souri (re) V + s (2 ^{ème} pers. Sing. impératif)	<i>Souris, on te prend en photo !</i>
souris = souris (N) + 0 (sing)	<i>La souris mange du fromage.</i>
souris = souris (N) + 0 (pl)	<i>Les souris mangent du fromage.</i>

Evidemment, hors contexte phrastique, il est impossible de donner une interprétation correcte au mot « souris ». On parle ici d'une ambiguïté réelle

Dans le contexte, l'ambiguïté morphologique peut-être résolue :

Exemple 2 :

Si nous considérons séparément les mots « le » (Déterminant ou Nom) et « bois » (Nom ou Verbe), Nous obtenons 4 possibilités d'analyse :

Det N Pron V Det V Pron N

Si nous considérons maintenant le syntagme nominal « le bois », nous éliminons deux combinaisons impossibles en langue française.

Det N Pron V ~~Det V~~ ~~Pron N~~

Si en plus , nous considérons le syntagme « le bois » dans un environnement phrastique précis, nous arrivons à une seule solution (soit l'une, soit l'autre)

- Le bois était sec (Det + N)
- Je le bois (Pron+V)

Les ambiguïtés morphologiques sont souvent levées au niveau syntaxique (par la structure phrastique).

Exemple 3 :

– L'ordinateur ne *marche* plus.

- Verbe
- Nom (une marche militaire, j'aime la marche)

Comment l'ordinateur peut-il savoir qu'il ne s'agit pas d'un nom ? Comment les humains le savent-ils ?

→ grâce à la négation

– Je la comprends très bien.

- pronom (la = quelqu'un, une fille)
- déterminant (article défini) (la femme, la lettre)

Comment l'ordinateur peut-il savoir qu'il ne s'agit pas d'un déterminant ? Comment les humains le savent-ils ?

→ Parce qu'un article est toujours suivi d'un nom
Cependant, le calcul n'est pas toujours évident.

Exemple 4 :

La petite glisse sur la glace.

Det	nom	verbe	prep	det	nom
Pron	adj	nom	adj	Pron	verbe
Nom				nom	

3 2 2 2 3 2 = 144 possibilités au niveau de la forme !

Sans parler du fait que « glace » en tant que verbe peut être à la :

1 pers sing du présent

3 pers sing présent

1 pers sing du subjonctif

3 pers sing du subjonctif

Idem pour « glisse ».

3 2 6 2 3 6 = 900 possibilité

Et nous n'avons pas encore parlé du fait que chaque mot peut posséder plusieurs sens !

La plupart des ambiguïtés morphologiques peuvent être levées au niveau syntaxique (en analysant la structure de la phrase).

3 Ambiguïtés lexicales

Il s'agit dans la plupart des cas, d'identifier le sens d'un mot.

Exemple 1 :

➤ Identification du sens d'un lexème (d'un mot plein ex. verbe, nom, adjectif)

« Bar », Nom:

- poisson
- unité de pression
- débit de boisson

« Couler », Verbe :

- Attention, l'eau coule du robinet.
- Le navire a coulé au large

Ces ambiguïtés sont souvent levées en contexte, il peut cependant y avoir une ambiguïté effective.

Exemple 2 :

- Il m'a fallu des quantités de scotch pour faire cette thèse. (whisky/ ruban adhésif).
- Je suis un imbécile. (J'm'identifie à un imbécile (le verbe être)/ je marche derrière un imbécile (le verbe suivre).

- Identification du sens d'un grammème (un mot grammatical : préposition, conjonction etc.)

Par nature, tous les grammèmes sont polysémiques. Comme il est (presque) impossible de construire des phrases sans les mots grammaticaux, l'ambiguïté virtuelle est présente donc systématiquement dans toutes les phrases.

Exemple 3 :

- Conjonction :
Donnez-moi du citron **et** de l'eau. (du citron dans (mélangé à) de l'eau/ du citron d'une part, de l'eau de l'autre)

- Prépositions :
Veux-tu voter **pour** moi ? (en faveur de ma personne/ à ma place)

- Articles :
C'est **le** même problème. (c'est un nouveau problème identique au précédent/ C'est toujours le problème précédent)

- Auxiliaires modaux :
Vous **devez** le savoir. (Vous êtes obligé de le savoir/ Il est vraisemblable que vous le savez)

4 Ambiguïtés syntaxiques (ou structurales)

4.1 Ambiguïtés liée à la segmentation de la phrase

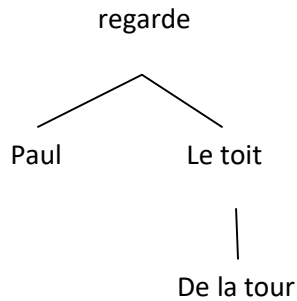
On a à faire à ce type d'ambiguïtés lorsque l'on peut rattacher un mot ou un syntagme à plusieurs (au moins deux) autres mots ou syntagmes. Selon l'attachement d'un mot (ou d'un syntagme) à un autre syntagme, la phrase peut changer d'interprétation.

Exemple 1 :

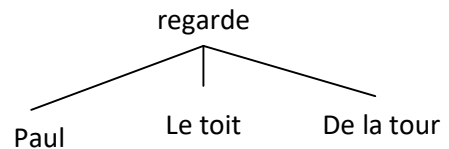
Attachement d'un groupe prépositionnel :

GN V GN Gprép

- Paul regarde le toit de la tour



(V(GN Gprép))



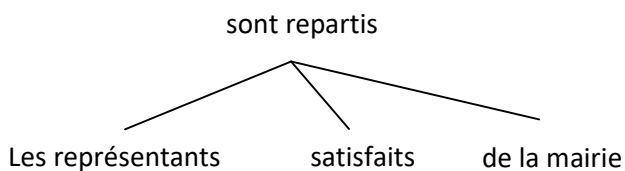
(V GN (Gprép))

Le toit de la tour, Paul le regarde / De (= depuis) la tour, Paul regarde le toit.

Exemple 2 :

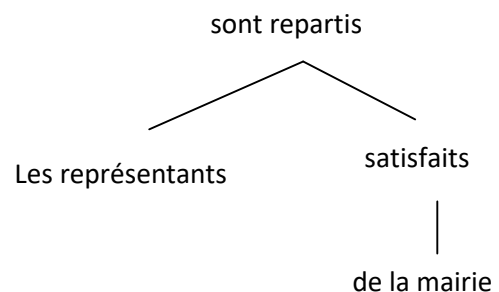
V A Grép

Les représentants de l'association sont repartis satisfaits de la mairie.



(V A (Gprép))

repartis satisfaits (de la mairie)



(V (A Gprép))

repartis (satisfaits de la mairie)

Exemple 3 :

- Elle épouse un professeur de droit anglais
- Structure ambiguë : N de N Adj (un professeur de droit anglais)

- ✓ C'est le professeur qui est de nationalité anglais -- > (N de N) Adj (un professeur de droit) anglais
- ✓ C'est le droit anglais que ce professeur enseigne -- > N de (N Adj) : un professeur de (droit anglais)

Remarque 1 :

L'ambiguïté de cette structure n'est pas absolue ; elle dépend des unités lexicales qui la composent :

- Elle épouse un professeur de droit commercial -- > Seule structure possible : N de (N Adj) un professeur de (droit commercial)
- Elle épouse un professeur de judo anglais-- > Seule structure possible : (N de N) Adj (un professeur de judo) anglais

Remarque 2 :

Les ambiguïtés structurelles peuvent être levées au niveau lexical même si ce n'est pas toujours le cas.

Exemple 4 :

V V Gprép

- Il a dit qu'il donnerait son avis par fax -- > (ce qu'il a dit, c'est qu'il donnerait son avis par fax / ce qu'il a dit par fax, c'est qu'il donnerait son avis)

((V(V Gprép)) soit (V V (Gprép))

Le groupe prépositionnel peut-être rattaché soit directement au verbe qui le précède (donner) soit aux deux verbes (dire et donner)

Ambiguïté potentielle basée sur cet exemple :

Exemple 5 :

- Il a dit qu'il donnerait son avis sur la formation.

Même si cette phrase n'est pas réellement ambiguë, elle l'est potentiellement, puisqu'elle a la même structure que la précédente : V V Gprép.

Exemple 6

V et V Gprép

- Il a parlé et il a menti à Marie -- > D'une part il a parlé (à qqn), et d'autre part, il a menti à Marie / A Marie, il a à la fois parlé et menti.

V et (V Gprép) soit ((V et V) Gprép)

Exemple 7 :

Rattachement des adjectifs

- J'aime boire mon café chaud.
✓ J'aime boire mon café qui est chaud --> Qu'est-ce que tu aimes boire ? Mon café chaud. ((V (N A))
✓ J'aime boire mon café quand il est chaud --> Comment aimes-tu boire ton café ? Chaud. ((V N) A)

Exemple 8 :

Rattachement des relatives

- Connais-tu le neveu du ministre qui a été élu président ? Qui a été élu président, le ministre ou son neveu ?

N (prep (N Rel)) / ((N (prep. N)) Rel)

Exemple 9 :

Rattachement des groupes nominaux

Le coiffeur, le banquier et sa femme se sont rencontrés.

X Y Z = 3

Le président de la République, M. Hollande, et le premier ministre se sont rencontrés.

X Y Z = 2 (en 2016)
 = 3 (en 2020)

Combien étaient-ils ?

Tous les cas d'ambiguïté que nous venons de voir présentent le problème de **segmentation syntaxique**. Ils résultent d'une possibilité de plusieurs structures syntaxiques d'une même phrase. En somme, nous pouvons lui attribuer (au moins) deux représentations arborescentes (ou parenthétiques) différentes. Les différentes représentations sont mutuellement exclusives.

4.2 Ambiguïtés syntaxiques : reconnaître les fonctions des syntagmes

Ces types d'ambiguïtés résultent du fait que, d'une part, une même marque grammaticale peut correspondre à plusieurs fonctions distinctes, et d'autre part, la même place dans une phrase peut être occupée par des termes ayant une fonction différente.

Exemple 1 :

- La circulation a été déviée par la gendarmerie.
 - ✓ Complément d'agent : par les gendarmes (ce sont les gendarmes qui ont dévié la circulation).
 - ✓ Complément circonstanciel de lieu : par le bâtiment de la gendarmerie (la circulation passe par ...)

Exemple 2 :

- Il attend la nuit.
 - ✓ Objet direct : il attend que la nuit tombe.
 - ✓ Complément circonstanciel de temps : il attend (quelqu'un ou quelque chose) pendant la nuit.

5 Ambiguïtés sémantiques : le calcul des relations

Le terme d'ambiguïté sémantique peut prêter à confusion : toutes les ambiguïtés ne sont-elles pas sémantique ? Dans la mesure où chaque fois qu'on a à faire à une ambiguïté (effective), on change d'interprétation. Effectivement, toute ambiguïté est un phénomène sémantique. Cependant, le classement des ambiguïtés selon les niveaux qu'elles affectent a pour objectif d'indiquer le niveau où il est possible de les décrire, c'est-à-dire de leur associer plusieurs représentations correspondant aux différentes significations. Par conséquent, sont appelées ambiguïtés sémantiques, les ambiguïtés dont la représentation relève du niveau d'analyse sémantique (et non morphologique, syntaxique etc.). Il s'agit plus particulièrement de la sémantique des relations entre les éléments de la phrase.

Exemple 1 :

Sur quoi porte la négation ?

- Je ne vois rien.
 - ✓ (1) Il n'y a rien dans mon champ de vision (pas de voitures arrivant de gauche par exemple)
 - ✓ (2) Il n'est pas vrai que je vois quoi que ce soit (il y a un camion qui me bouche la vue par exemple).

En logique de prédicats :

- (1) Il n'existe pas de y tel que $p(x, y)$ --> 'il n'existe pas de y tel que je voir y '
- (2) Quel que soit y , non $p(x, y)$ --> 'quel que soit y , il est faux que je voir y '

Exemple 2 :

Sur quoi porte la restriction ?

Il n'a connu qu'une série d'échecs dans sa vie.

- (1) Il a connu une seule série d'échecs
- (2) Il n'a rien connu d'autre que des échecs en série.
- ✓ Dans (1) la restriction « ne que » porte sur le seul opérateur de quantification (un) : « il existe un et un seul y , tel que y = série d'échecs et tel que il connaître y ».
- ✓ Dans (2) la restriction porte sur le prédicat : « quel que soit y , il connaître y implique y = série d'échecs »

Exemple 3 :

Sur quoi porte le procès ?

- l'alcool brûle.
- ✓ L'alcool provoque des brûlures (procès orienté vers l'objet, c'est-à-dire interprété comme affectant l'objet, ici indéterminé).
- ✓ L'alcool est enflammée (ou inflammable) (procès orienté vers le sujet, c'est-dire compris comme une pure activité du sujet).

6 Ambiguïtés pragmatiques

L'ambiguïté pragmatique est liée d'une part à la connaissance du monde, ou à ce qu'on appelle le bon sens dont un ordinateur ne dispose pas, et d'autre part à une difficulté de calcul des valeurs référentielles.

6.1 La connaissance du monde, le bon sens

Exemple 1 :

Connaissane du monde

- La fermière a vendu la vache, parce qu'elle ne donnait pas de lait
- Elle = (la fermière ou la vache)

Comment l'ordinateur peut-il savoir laquelle de ces deux interprétations choisir ? Comment les humains le savent-ils ?

-> Parce qu'on sait tous que ce sont les vaches qui donnent du lait et non pas les fermières

Exemple 2 :

Pronoms anaphoriques

- Essayez bien les **poissons**, coupez les **têtes**, huilez-**les** très légèrement, mettez-**les** sur les grilles -- > (poissons ou têtes)
- Le premier ministre a rencontré à sa demande le président de la République -- > à la demande de qui ?

Exemple 3 :

Cumul des ambiguïtés :

- Pierre a dit que Paul a répondu qu'il ne viendra pas, et que ça lui est complètement égal -- > Qui est 'il' et 'lui' ? Pierre, Paul ou quelqu'un d'autre ?

6.2 Difficulté de calcul des valeurs référentielles

Elle se déclenche lorsque le récepteur tente de reconstruire les valeurs référentielles du procès et celles des actants, il peut se trouver confronté à diverses ambiguïtés.

Exemple 1 :

Interprétation de temps verbaux :

- Paul va à l'école
 - ✓ Il marche en ce moment en direction de l'école (situation actuelle)
 - ✓ Il est écolier (propriété actuelle).
- Sans vous, je m'ennuyais.
 - ✓ Comme vous n'étiez pas là, je me suis effectivement ennuyé (réalisé)
 - ✓ Si vous n'aviez pas été là, je me serais ennuyé (non-réalisé).

Interprétation des déterminants

- Le chien aboie.
 - ✓ L'animal appelé chien a pour cri l'aboiement » (le chien = générique : désigne le représentant de la classe des chiens)
 - ✓ Le chien en question est en train d'aboyer (le chien = non-générique : renvoie à un chien particulier).

Exemple 2 :

Interprétation des pronoms

- Je cherche quelqu'un.

- ✓ J'étais en compagnie de quelqu'un que j'ai perdu de vue : je le cherche, je voudrais le retrouver (quelqu'un = spécifique)
- ✓ Pour ne pas être seul, ou pour m'aider à faire quelque chose, j'ai besoin de quelqu'un, il faudrait que je trouve quelqu'un (quelqu'un = non spécifique).

7 Conclusion :

Il ne faut pas se fier aux apparences ! Le fait que nous ne discernons pas d'ambiguïtés dans la langue vient de notre capacité à désambigüiser sur le champs à partir de plusieurs types d'informations en contexte. Si on n'a pas l'accès à ces informations (ce qui est le cas des ordinateurs), tout devient difficile dans une langue.

8 Bibliographie

- Bouillon P. (eds) (1998), Traitement automatique des langues naturelles, Aupelf Uref et Editions Duculot, Paris
- Carré R. et all. (1991), Langage humain et la machine, Presses du CNRS, Paris
- Fuchs C. et all (1993), Linguistique et traitement automatique des langues, Hachette, Paris
- Fuchs C. (1996), Les ambiguïtés du français, Ophrys, Paris
- Gross , M. (2002), « Les ambiguïtés » in : Lingvisticae Investigationes 24 : 1, Benjamins, Amsterdam.

9 Exercices

Exercice 1

Donnez deux découpages morphologiques possibles des phrases qui suivent :

1. 1 Le petit garde la montre.
2. 2 J'ai acheté pour le goûter un rosé de Provence.
3. 3 Devant cette somme, il hésite.

Exercice 2

De quoi résulte l'ambiguïté effective des phrases suivantes (quelle est la forme qui provoque l'ambiguïté ?)

1. 1 Il cherche une maison qui donne sur la mer.
2. 2 C'est ainsi que finit cette lamentable histoire.
3. 3 Il a rencontré un vieux maniaque.

Exercice 3

Quel mot polysémique provoque l'ambiguïté effective des phrases suivantes ? Imaginez les suites qui les désambiguïseraient.

1. 1 Dans le milieu des conservateurs, cette mesure a été rejetée.
2. 2 C'est un véritable artiste Il a réussi l'opération à merveille.
3. 3 il fut extrêmement cordial avec son hôte.
4. 4 Paul recule.
5. 5 On a relevé quarante blessés.
6. 6 Cette année, le gouvernement a changé.
7. 7 Votre pantalon, voulez-vous le repasser ?
8. 8 Tu es comme mon fils.

Exercice 4

Pourquoi les phrases suivantes sont-elles ambiguës ? Décrivez les deux structures syntaxiques qui permettent de les désambigüiser (écriture parenthétique)

1. Jean a abordé la fille avec les fleurs.
2. Les lecteurs rapporteront les livres empruntés avant le 1er avril.
3. Il est venu et il est reparti avec tristesse.
4. Moules marinières et frites à volonté.
5. J'ai visité la maison du père de Paul et de Marie.
6. J'ai acheté du chocolat au lait suisse.
7. Il nous a montré une pierre de taille impressionnante.

Exercice 5

Quels types d'ambiguïtés trouve-t-on dans les phrases suivantes ? Décrivez-les brièvement.

1. Je ne serai pas le premier président à perdre la guerre.
2. Vous pouvez tous gagner une semaine de sports d'hiver.
3. Les oiseaux volent Paul a confiance en lui.

CM3

Applications en Traitement Automatique des Langues : Exemple de la traduction automatique

Descriptif du cours

Les applications du traitement automatique sont multiples et variées car le langage, sous ses formes écrites ou orales, est présent partout. En effet, les outils que le TAL peut offrir s'inscrivent tous dans la même perspective, manipuler le langage par un ordinateur pour offrir pouvoir communiquer rapidement et surtout avec la maximum d'interlocuteurs. D'autant plus que l'avènement de l'internet n'a fait qu'accentuer la nécessité de faire appel aux ordinateurs pour faire face à un flux d'informations important et mondial.

La traduction automatique est aujourd'hui l'une des technologies du TAL les plus utilisées que ce soit par les professionnels ou les particuliers. Les logiciels de traduction se sont démocratisés et leur usage est devenu nécessaire dans des contextes souvent multilingues où plusieurs langues se côtoient. On notera que la traduction automatique s'est considérablement améliorée grâce à une recherche de plus en plus pointue, qui a tendance à l'intégrer dans les différents outils du quotidien (Facebook, Google Chrome, Trip Advisor, Skype, etc.). Les systèmes de traduction varient selon l'approche qui les sous-tend et leur conception tient compte de plusieurs paramètres. Ce cours se propose à travers le premier chapitre intitulé « Introduction à la traduction automatique » de passer en revue les différents systèmes de traduction après un bref historique retraçant les étapes marquantes de l'évolution de la traduction automatique (TA). Le deuxième chapitre intitulé « Comment construire un traducteur automatique » exposera les différentes étapes nécessaires à la construction d'un système de traduction automatique à base de règles linguistiques.

Objectifs du cours :

- Définir la traduction automatique et identifier ses différents domaines d'application.
- Retracer l'histoire de la traduction automatique et identifier les faits les plus marquants de son évolution.
- Présenter les différents types de systèmes de traduction automatique en fonction de plusieurs paramètres.
- Présenter les étapes nécessaires à la construction d'un système de TA à base de règles linguistiques.

Chapitre 1

Introduction à la traduction automatique

Contenu du cours :

1	Introduction	4
2	Qu'est ce que la TA ?	4
3	Aperçu historique de la TA	4
4	Type de systèmes de TA.....	6
4.1	Degré de généralité.....	6
4.1.1	Système de Traduction Entièrement Automatique de Haute Qualité (TAEHQ).....	6
4.1.2	Système de Traduction Assistée par Ordinateur, hyperonyme de TAAH (Traduction Automatique Assistée par l'Homme)	6
4.2	Nombre de langues traitées et sens de traduction.....	6
4.2.1	Systèmes bilingues	6
4.2.2	Systèmes multilingues	6
4.2.3	Systèmes unidirectionnels	6
4.2.4	Systèmes bidirectionnels	6
4.3	Type d'informations requis	7
4.3.1	Systèmes à règles linguistiques	7
4.3.2	Systèmes à informations non linguistiques	7
5	Approches de TA.....	7
5.1	Approche directe	7
5.2	Approche indirect	7
5.2.1	Système à interlangue (pivot).....	7
5.2.1.1	Inconvénient des systèmes à interlangue.....	7
5.3	Systèmes à transfert.....	8

Traitement Automatique des Langues

Niveau : Master 1 Didactique des langues étrangères

CM 3: Applications en Traitement Automatiques des Langues : Exemple de la traduction automatique

Chapitre 1 : Comment construire un traducteur automatique

5.3.1	Avantages d'un système de transfert.....	9
5.3.1.1	Extensibilité.....	9
5.3.1.2	Saisie des dépendances structurales	9
5.3.1.3	Lien bilingue.....	9
6	Quelles sont les conditions permettant de construire un système de traduction automatique robuste ?	9
7	Mode d'intervention humaine sur les systèmes de TA.....	10
7.1	Prédiction.....	10
7.2	Traduction interactive.....	10
7.2.1	Inconvénients de la traduction interactive	11
7.3	Postédition.....	11

1 Introduction

L'intérêt porté à la traduction automatique est en nette progression, et le nombre croissant de publications ainsi que les manifestations scientifiques qui lui sont consacrées attestent de cet intérêt. Le domaine de la TA est un domaine stratégique pour les industriels et est en parfaite accord avec l'économie du marché actuelle. Elle est aussi un espoir pour ceux qui sont confrontés au multilinguisme et à des quantités considérables de documents qu'il faut lire et transmettre dans différentes langues étrangères. La traduction est un domaine de recherche appliquée et interdisciplinaire, elle touche à la linguistique, l'informatique, la logique, l'intelligence artificielle, etc.

2 Qu'est-ce que la TA ?

La traduction automatique est l'application de l'informatique à la traduction des textes d'une langue naturelle (A) appelée langue de départ ou langue source vers une langue (B) appelée langue d'arrivée ou langue cible.

3 Aperçu historique de la TA

La traduction automatique a connu des moments de gloire, elle les connaît encore d'ailleurs, mais elle a connu aussi des moments de marasme. Après la seconde guerre mondiale, on assiste au développement des ordinateurs et avec, des nouveaux projets s'esquissent. Des chercheurs veulent s'inspirer des méthodes appliquées avec succès pour le décryptage de messages. La traduction de textes dans une autre langue en contexte de guerre intéresse en premier les militaires et les services d'espionnage. La recherche commence dans l'euphorie. Les langues traitées se limitent à l'anglais et le russe. Un compte rendu de 1960, sur les progrès de la TA se montre plutôt pessimiste, BAR-HILLEL chargé de remettre le compte-rendu explique qu'il est impossible d'imaginer une traduction de haute qualité sans prendre en compte le sens, un problème qui semble insoluble à cette époque-là. Ce désaveu et l'échec de la TA trouvent de nombreuses explications :

1- L'informatique était à ses débuts (mémoire limitée).

2- Les langages de programmation étaient primitifs.

3- La linguistique était peu développée. (Le structuralisme basé sur les énoncés réalisés et écartant la question du sens ne correspond pas aux systèmes de TA généralistes qu'ils voulaient développer à cette époque-là).

4- les chercheurs ne se rendaient pas compte des problèmes que pose la traduction automatique.

En 1966, le comité ALPAC (Automatic Language Processing Advisory Committee) chargé par les autorités américaines ayant investi des sommes considérables dans la TA, d'évaluer le progrès atteint dans le domaine de TA, rend un rapport alarmant et les conséquences de ce rapport furent catastrophiques. Le comité conclut que les recherches dans leur état actuel ne sont pas rentables pour les Etats-Unis. Les subventions sont donc coupées immédiatement. Face à ce manque de moyens, les chercheurs se tournent vers deux orientations différentes :

1- Recherches théoriques en syntaxe : à la fin des années 60, les théories de Chomsky sur les grammaires formelles et la grammaire générative transformationnelle eurent un double impact.

1. La structure du langage humain est assez régulière pour être formalisée et décrite.
2. Le traitement automatique peut être limité à un traitement syntaxique pur : la syntaxe est indispensable pour extraire le sens de la phrase et l'analyse syntaxique peut se faire indépendamment de l'analyse du sens.

2- la théorie chomskyenne suscite beaucoup d'interrogations quant à sa validité. Est-il vraiment possible de faire une analyse syntaxique non ambiguë et générale d'une phrase sans avoir recours à l'analyse sémantique et la situation d'énonciation ? L'analyse syntaxique est-elle vraiment indispensable pour l'extraction du sens ?

De ces interrogations, naît une réflexion autour de la représentation du sens. En parallèle, se développent des recherches sur la représentation des connaissances en IA (intelligence artificielle). Ces recherches aboutissent à des résultats concrets dans le TAL et à des programmes qui montrent la possibilité de traiter le langage naturel automatiquement sans mettre en œuvre une théorie linguistique.

Parallèlement à ce souffle de recherche, des équipes travaillent en Europe et au Canada notamment avec des fonds privés dans le développement de systèmes de TA et se convertissent en entreprises privées spécialistes du domaine. Le Russe cède sa place progressivement à d'autres langues.

A partir de 1975, face au nombre croissant de traductions dans différentes langues, la communauté européenne envisage de se tourner vers la traduction automatique. Elle engage de ce fait l'année suivante, en 1976 des projets qui visent la coordination des recherches sur le multilinguisme et la TA. Elle annonce la même année, la commercialisation de SYSTRAN, un système de traduction commercial, et la traduction automatique devient à cette occasion une application connue du grand public et suscite à nouveau l'intérêt des firmes privées. La TA est désormais une application très rentable aux yeux des industriels et les systèmes de traduction automatique se multiplient. Cet attrait économique est le même au États-Unis et au Japon qui, de leur côté, essaient de se tailler des parts du marché.

4 Type de systèmes de TA

Les systèmes peuvent être classés selon le :

4.1 Degré de généralité

4.1.1 Système de Traduction Entièrement Automatique de Haute Qualité (TAEHQ)

La traduction entièrement automatique de haute qualité appliquée à la langue générale semble ne pas atteindre ses objectifs et reste éloignée de la qualité attendue. Elle nécessite généralement l'intervention humaine soit avant le processus de traduction (préédition), soit au moment de la traduction (interactive) soit après la traduction (postédition). Seuls les systèmes qui traitent d'un sous-langage bien délimité peuvent produire des traductions correctes sans intervention humaine.

4.1.2 Système de Traduction Assistée par Ordinateur, hyperonyme de TAAH (Traduction Automatique Assistée par l'Homme)

La TA n'inclut donc pas les outils informatiques d'aide à la traduction qui permettent aux traducteurs humains d'accéder aux multiples dictionnaires ou à des bases de données terminologiques pour l'édition des textes ou la gestion des glossaires.

NB : il est indispensable de faire la différence entre la traduction humaine assistée par machine (THAM) et TA ou TAO. En THAM, c'est l'utilisateur qui traduit en s'aidant de dictionnaires bilingues de bases terminologiques, accessibles depuis un traitement de texte. Le tout, forme un poste de travail pour le traducteur, réalisé sur micro-ordinateur ou sur une station de travail.

4.2 Nombre de langues traitées et sens de traduction

4.2.1 Systèmes bilingues

Ces systèmes traitent une seule paire de langue.

4.2.2 Systèmes multilingues

Ils traitent plusieurs langues à la fois.

4.2.3 Systèmes unidirectionnels

Ils comportent une seule direction de traduction.

4.2.4 Systèmes bidirectionnels

Ils comportent les deux directions de traduction.

4.3 Type d'informations requises

4.3.1 Systèmes à règles linguistiques

Ce sont des systèmes qui n'utilisent que des informations linguistiques (morphologiques, syntaxiques, etc.).

4.3.2 Systèmes à informations non linguistiques

Ce sont des systèmes qui prônent, le recours à des informations non linguistiques comme les corpus alignés, les connaissances générales sur le monde, les données statistiques, etc.

5 Approches de TA

5.1 Approche directe

Cette approche concerne les premiers systèmes de traduction automatique mis en place. Ils sont conçus pour une seule paire de langues et sont unidirectionnels. Ils se basent sur des équivalences terme à terme et se passent de toute analyse approfondie du texte source pour la traduction. En effet, après une analyse morphologique, le système donne en consultant le dictionnaire bilingue des correspondances de traduction mot à mot. En d'autres termes, les systèmes n'utilisent pas de représentations abstraites mais passent directement de la chaîne de la langue source à la chaîne de la langue cible.

5.2 Approche indirecte

Cette approche englobe deux types de systèmes, appelés communément systèmes de deuxième génération.

5.2.1 Système à interlangue (pivot)

Ces systèmes ne comportent que deux modules : un module d'analyse et un module de génération. Ils sont largement orientés vers la recherche du sens, et la compréhension du domaine de connaissances. Il s'agit de représenter le sens de la langue source en une représentation interlangue pour passer ensuite à la langue cible. Ils recourent aux primitifs sémantiques et aux représentations du monde, souvent sous forme de hiérarchies arborescentes d'entités. Cette arborescence sert de généralisateur et évite les redondances. Le lien établi entre les deux langues source et cible se réfère aux objets du monde.

5.2.1.1 Inconvénients des systèmes à interlangue

Il est difficile de définir l'ensemble des primitifs sémantiques indépendamment d'une langue donnée et d'assurer la complétude de cette représentation du sens. Il est aussi difficile d'extraire le sens d'un texte, et lorsque ce dernier est extrait, il est rendu généralement par des paraphrases en langue cible.

5.3 Systèmes à transfert

Ces systèmes comportent trois modules : un module d'analyse, un module de transfert et un module de génération.

Le module d'analyse traite la chaîne de la langue source morphologiquement, pour en produire une représentation syntaxique de surface. Cette représentation syntaxique est transformée en une représentation abstraite en faisant abstraction de tous les phénomènes sans incidence sur la traduction. Le module de transfert convertit cette représentation abstraite en une représentation abstraite de la langue cible que le module de génération utilisera pour produire la chaîne en langue d'arrivée.

La représentation abstraite peut prendre en compte les dépendances syntaxiques, ainsi on peut imposer des contraintes sémantiques sur les structures sources.

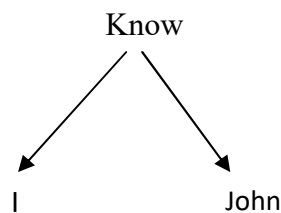
Exemple 1

I know John (deux équivalents potentiels au verbe Know en français : connaître et savoir)

- a) Je connais John
- b) Je sais John

Structure permettant la reconnaissance du verbe know dans le contexte de la phrase :

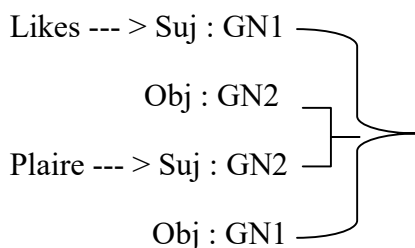
Relation abstraite « Sujet + Verbe + Obj [Nom Propre] - - > Verbe = [connaître]



Exemple 2

John likes Marie

Marie plait à john



Les systèmes de transfert établissent les équivalences traductives grâce à des liens lexicaux en contexte ; ce qui peut amener à des changements structuraux.

De nos jours, Les systèmes de traduction de transfert sont loin d'être des systèmes basés uniquement sur l'analyse syntaxique ou fonctionnelle, et ne s'opposent plus aux systèmes à interlangue en termes de profondeur. L'aspect sémantique a fait son entrée dans sa conception et des systèmes de transfert utilisent maintenant des informations sémantiques au moment de l'analyse et durant le transfert. En effet, s'il reste vrai qu'un système interlangue ne peut se passer de représentations du sens et des connaissances du monde, il n'est pas exclu qu'un système de transfert y recourt. La différence entre systèmes de transfert et systèmes à interlangue se résume dans le rôle qu'assume le lexique dans les équivalences traductives. Si le système relie la langue source à langue cible via le lexique (ce qui ne le réduit pas au mot à mot), on dira que son architecture est une architecture de transfert. Si le lien renvoie aux objets du monde et à leurs relations, alors le système traduit par interlangue.

5.3.1 Avantages d'un système de transfert

5.3.1.1 Extensibilité

En cas d'ajout de paires de langues, dans un système de transfert, il suffit d'ajouter une composante de transfert, d'analyse et de génération des nouvelles langues sans refaire ce qui a déjà été fait sur les langues déjà existantes. Dans un système direct il faut créer un système entièrement nouveau. Dans un système à interlangue, seules les composantes monolingues doivent être créées.

5.3.1.2 Saisie des dépendances structurales

Les systèmes de transfert permettent de saisir les dépendances structurales et n'analysent pas la langue en termes de chaîne de caractères. C'est la profondeur de l'analyse que prône l'approche indirecte dans les systèmes de transfert, qui permet de comprendre mieux le fonctionnement des structures et leurs variations générées par le lexique.

5.3.1.3 Lien bilingue

Les systèmes de transfert permettent plus facilement de mettre en contact deux langues en mettant en exergue leurs convergences et divergences. L'information linguistique qui découle de l'analyse profonde est intelligemment redistribuée sur les règles de transfert. Autrement dit, il y a plus d'adaptabilité de la microsystemique dans les systèmes de transfert que dans les systèmes directs.

6 Quelles sont les conditions permettant de construire un système de traduction automatique robuste ?

- **Disposer d'un modèle syntaxique** pour représenter la syntaxe d'une langue, on peut associer lexique et grammaire ou faire la distinction entre les deux. Ce modèle doit

permettre une mise en application informatique, rapide en temps d'exécution, efficace en reconnaissance des ambiguïtés.

- **Disposer d'un modèle sémantique** qui permet une représentation fine de la sémantique de la langue traitée (structure profonde).
- **Faire le lien entre modèle syntaxique et modèle sémantique** qui sont à priori indépendant l'un de l'autre. La qualité et la finesse du système de traduction automatique dépendent de la nature de l'association qu'on établit entre modèle syntaxique et modèle sémantique. Ce lien s'esquisse au moment de la définition des représentations abstraites où syntaxe et sémantique peuvent être intégrées en même temps.
- **Faire le lien entre le modèle sémantique intégré dans la syntaxe de la langue source et le modèle sémantique de la langue cible intégré lui aussi dans la syntaxe.** Ce lien doit prendre en considération les particularités sémantiques de chaque langue.
- **Travailler sur une langue de spécialité bien définie.** les langues de spécialité sont bien définies syntaxiquement, sémantiquement, et stylistiquement. Cependant elles continuent à puiser leurs ressources de la langue générale.

7 Mode d'intervention humaine sur les systèmes de TA

L'intervention humaine peut se situer à divers moments du processus de traduction : avant, pendant ou après. Cela correspond soit à une préparation du texte, soit à une désambiguïsation du texte, soit à une révision du texte.

7.1 Prédiction

L'utilisateur à ce stade du processus adapte soit la forme du texte (isoler les phrases ou les paragraphes) soit le contenu de manière à le rendre plus explicite et conforme aux exigences du système de traduction.

7.2 Traduction interactive

Plusieurs systèmes recourent à l'expertise humaine pour résoudre des cas d'ambiguïtés qui restent la principale difficulté du traitement automatique des langues. Le système tente par ce procédé d'exploiter l'intuition du locuteur natif.

- Présenter un choix de rattachement syntaxique :
 - Le professeur de judo français
 - Le professeur français de judo
- Présenter à l'utilisateur un choix de sens pour un mot donné : « avocat : le fruit ou la profession »
- Présenter un cas de choix de référents pour un mot donné : « le pronom personnel 'il' dans une reprise anaphorique »

7.2.1 Inconvénients de la traduction interactive

- Ce type de système repose sur la connaissance linguistique de l'utilisateur, or cette connaissance peut être limitée ou insuffisante.
- Les ambiguïtés générées par le système de TA ne sont pas toujours compréhensibles par l'utilisateur, il faut connaître dans certains cas l'architecture du système pour pouvoir localiser le problème et le comprendre.

7.3 Postédition

Il s'agit d'une relecture humaine avec des outils informatiques similaires à la prédiction tels que : les dictionnaires bilingues, de synonymes, etc.

Chapitre 2

Construire un traducteur automatique

Contenu du cours

1	Délimitation du mot	3
2	Délimitation de la phrase	4
3	Formalisation par règles linguistiques	6
3.1	Le dictionnaire	6
3.2	Dictionnaire et délimitation du mot.....	7
3.3	Mots complexes source d'analyses erronées	8
4	Étiquetage	8
4.1	Dictionnaire avec catégorisation	8
4.2	Variation flexionnelle	10
4.3	Dictionnaire de formes (en extension).....	10
4.4	Dictionnaire de formes canoniques	11
5	De l'analyse du mot à celle de la phrase	11
5.1	L'analyse syntaxique	11
5.2	La phrase, séquence linéaire de mots	12
6	La phrase, unité linguistique structurée hiérarchiquement	13
6.1	Ordre des analyses syntaxiques	15
6.2	Correspondance entre les règles de réécriture et les modes de catégorisation	16
6.3	Résolution des ambiguïtés catégorielles.....	16
7	Analyse syntaxique et traduction	17
8	Autres représentations de la phrase.....	17
9	Limites des grammaires formelles en TAL.....	18
10	Prise en compte du sens	18
10.1	Informations sémantiques sur les mots.....	18
10.2	Marquage des classes sémantiques par des traits	19

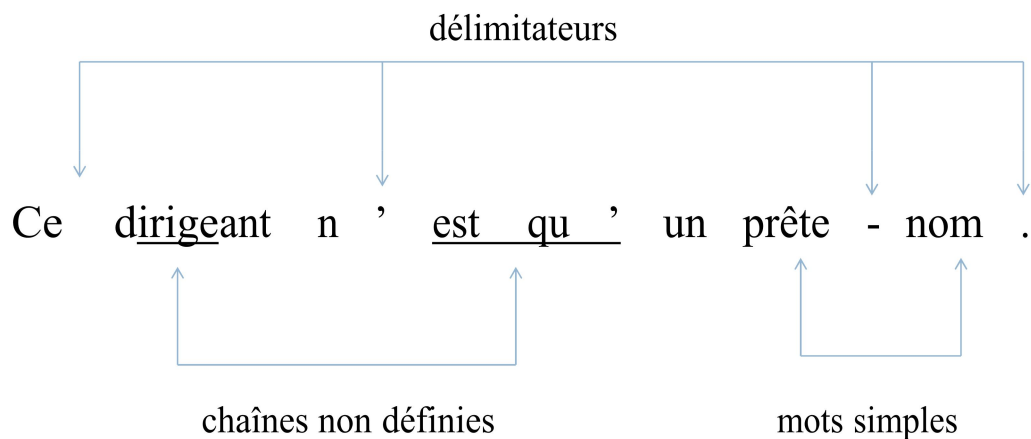
10.3	Vérifications des compatibilités sémantiques entre les mots	20
10.4	Résolution des ambiguïtés sémantiques	21
11	Conclusion.....	22
12	Bibliographie.....	23
13	Exercices	24

1 Délimitation du mot

Le traitement automatique doit, afin de délimiter les mots, distinguer les différents caractères qui entrent dans la formation de ces mots de ceux qui n'en font pas partie. Les principaux repères permettant la délimitation des mots sont les espaces. A ces premières balises doivent être ajoutés les différents signes de ponctuation (point, virgule, ...) et les autres caractères non alphabétiques (apostrophe, trait d'union). L'ordinateur va, par exemple, considérer que « nom » est un mot dans les combinaisons suivantes : « son nom de famille », « un porte-nom », « notez votre nom, prénom et adresse », ...Le découpage d'un texte en mots, lorsqu'il est réalisé de manière automatique porte le nom de **segmentation**.

Dorénavant, nous appellerons ces chaînes de caractères délimitées au moyen de ces indicateurs graphiques : **mots simples**. Quant aux chaînes non délimitées, nous les nommerons : **chaînes non définies**. Les deux types de suites de caractères que nous venons de voir sont regroupés sous le terme de **chaîne de caractères**.

Exemple de délimitation de mots



Si la délimitation graphique convient à la majorité des mots, on constate que parfois cette délimitation échoue. La délimitation ne permet pas, par exemple, de considérer « prête-nom » comme un mot unique. A cette étape, le programme informatique qui s'appuie exclusivement sur les limites graphiques, tel que nous les avons définies ci-avant, ne peut déceler les mots qui ne coïncident pas avec ces limites. Le problème posé par exemple par « prête-nom » peut se résoudre si on analyse le trait d'union comme une partie intégrante du nom et non pas comme un délimiteur. Cependant, et peu importe la stratégie adoptée, le traitement effectué reste toujours systématique à cette étape. Ainsi, les signes de ponctuation et autres caractères non alphabétiques, qui dans certains cas font partie intégrante des mots et dans d'autres cas

jouent vraiment un rôle de délimitateurs, seront incontestablement des obstacles à la reconnaissance des mots compte tenu de leur ambiguïté. Le traitement automatique peut cependant venir à bout de ce type de problématique dès lors qu'il pourra consulter un **dictionnaire**.

2 Délimitation de la phrase

La délimitation de la phrase s'appuie également sur des indices graphiques. Les marqueurs utilisés sont la majuscule (pour signaler le début d'une phrase) et un signe de ponctuation au choix (le point, le point d'exclamation, le point d'interrogation ou les points de suspension) suivi d'un espace puis d'une nouvelle majuscule. A ces premiers repères, on ajoute des **éléments structurels** appelés également des formatages. Les éléments structurels peuvent être par exemple une tabulation ou le retour à la ligne. Ces derniers indices servent à fragmenter un paragraphe en phrases même s'il n'y a qu'une majuscule et un seul point, comme par exemple :

Pour activer la fonction, effectuer les opérations suivantes :

Cliquer sur la barre de menu ;

Faire dérouler le menu édition ;

Cliquer sur l'option coller.

N°1	Pour activer la fonction, effectuer les opérations suivantes :
N°2	Cliquer sur la barre de menu ;
N°3	Faire dérouler le menu édition ;
N°4	Cliquer sur l'option coller.

Tous ces indicateurs permettent dans un grand nombre de cas de découper convenablement la plupart des phrases. Cependant, une délimitation fondée uniquement sur des repères graphiques n'est pas entièrement satisfaisante. Nous allons voir, à travers quelques exemples, pourquoi cette délimitation n'est pas entièrement satisfaisante. Des parties de phrases sont parfois disposées sur des lignes différentes, comme par exemple des titres et des éléments qui figurent dans des énumérations :

Si le blessé présente une plaie hémorragique :

- Essayez d'arrêter l'hémorragie en comprimant la plaie avec un gros pansement (serviette par ex) ou avec votre poing ;
- Vous pouvez fixer le pansement avec une bande serrée fortement mais sans excès ;
- Ne placez un garrot qu'en cas de nécessité absolue sans oublier d'indiquer l'heure où vous l'avez posé.

Le traitement automatique qui s'appuie exclusivement sur des repères graphiques découpera toutes les suites de mots figurant sur des lignes différentes comme des phrases.

N°1	Si le blessé présente une plaie hémorragique:
N°2	Essayez d'arrêter l'hémorragie en comprimant la plaie avec un gros pansement
N°3	(serviette par ex) ou avec votre poing.
N°4	Vous pouvez fixer le pansement avec une bande serrée fortement mais sans excès.
N°5	Ne placez un garrot qu'en cas de nécessité absolue sans oublier d'indiquer l'heure où
N°6	vous l'avez posé.

Remarque : pour éviter toute ambiguïté, le système idéal fait une analyse syntaxique pour s'assurer que la séquence analysée est bien une phrase. Mais peu de systèmes (notamment commercialisés) vont aussi loin. Un nouveau problème peut survenir lorsque les caractères définis comme séparateurs de phrases ont d'autres finalités. Ainsi, le point peut être par exemple placé à la fin d'une abréviation : « M. le doyen » et ne signale donc pas nécessairement une fin de phrase. L'ambiguïté est levée si on examine la chaîne de caractères qui succède au point :

- commence par une majuscule → phrase
- commence par une minuscule → partie d'une phrase

De même, certains mots débutant par une majuscule comme par exemple les noms propres peuvent succéder à une abréviation sans signaler une nouvelle phrase : « M. Dupont, doyen de la faculté des lettres ». Une solution à ce problème consiste à répertorier les abréviations les plus courantes telles qu'elles sont dans le texte (avec leur point) dans un dictionnaire pour rendre le programme apte à différencier les points de fin de phrases des abréviatifs : M., Dr., ...

D'autres signes de ponctuation, comme les parenthèses, les crochets, les guillemets ou encore les tirets, posent des problèmes lors du découpage d'un texte en phrases. Les traitements de ces signes varient selon le type d'applications. On peut par exemple envisager le contenu d'une parenthèse comme une phrase entière. Cependant, une approche systématique de ce domaine est rarement satisfaisante. Prenons comme exemple la parenthèse, elle peut contenir quelques mots, une proposition voire une phrase entière :

- Les enfants (Jean et Marie) ont pu sauver leur mère.
 - nom
- Les enfants (dont l'école proposait une initiation aux premiers secours) ont pu sauver leur mère.
 - proposition
- Les enfants ont pu sauver leur mère (soulignons que Jean et Marie avaient suivi une initiation aux premiers secours dans leur école).
 - phrase entière

3 Formalisation par règles linguistiques

Les techniques de reconnaissances que nous venons de voir, ne constituent qu'une amorce nécessaire au traitement automatique de la langue. Elles ne peuvent cependant mener à une véritable représentation : elles ignorent la fonction ou encore la signification des mots par exemple. Le programme de traitement automatique doit se doter d'un ensemble de connaissances linguistiques qui le rendront « apte à comprendre » les unités qui se succèdent dans un texte. Les techniques qui s'appuient sur cette stratégie constituent les **règles linguistiques**.

Pour consigner ces connaissances, le linguiste puise dans les modèles que lui propose la linguistique. Le traducteur automatique utilise les descriptions des linguistes qui donnent les formes des unités linguistiques ou des éléments sur leurs comportements. Il a également recours aux règles qui permettent une généralisation dans des ensembles, de prime abord, hétérogènes. Le linguiste s'appuie sur les principes d'une **linguistique dite formelle**, dont l'objet est de décrire l'ensemble des phénomènes linguistiques de façon explicite (contrairement à l'humain et aux éléments implicites).

L'information donnée dans un dictionnaire ne sera que partielle, car les utilisateurs connaissent certaines choses avant de chercher dans cet ouvrage. Une machine ne décode pas tous ces renseignements sauf s'ils sont inscrits clairement dans le programme. L'explicitation entreprise par la linguistique formelle est réalisée au moyen d'un langage sans équivoque. Chaque unité linguistique et chacune des propriétés de cette unité sont associées à un code unique et distinct des autres. Le linguiste manipule une langue naturelle chargée d'ambiguïtés au moyen d'un langage qui doit en être totalement dépourvu. En plus de décrire les unités linguistiques, le linguiste établit des procédures pour les manipuler. Le traitement automatique interprète les comportements des différentes unités linguistiques (comme par ex, l'agencement des mots dans la phrase). Les manipulations pratiquées sur les mots et les phrases sont réalisées via un langage mathématique.

3.1 Le dictionnaire

Les règles linguistiques doivent reconnaître les mots et les associer à certaines propriétés. Pour ce faire, elles passent par un **dictionnaire**. Ce dictionnaire est, en fait, une liste qui contient « idéalement » tous les mots d'une langue. Le dictionnaire constitue une composante essentielle dans la traduction automatique. Il constitue le premier répertoire consulté par le programme. Ce programme va s'appuyer sur lui pour poursuivre les analyses. Prenons un exemple pour illustrer la sollicitation d'un dictionnaire lors d'un traitement par un programme de traduction automatique, ce microdictionnaire contient cinq mots : chat, dame, le, la, griffe.

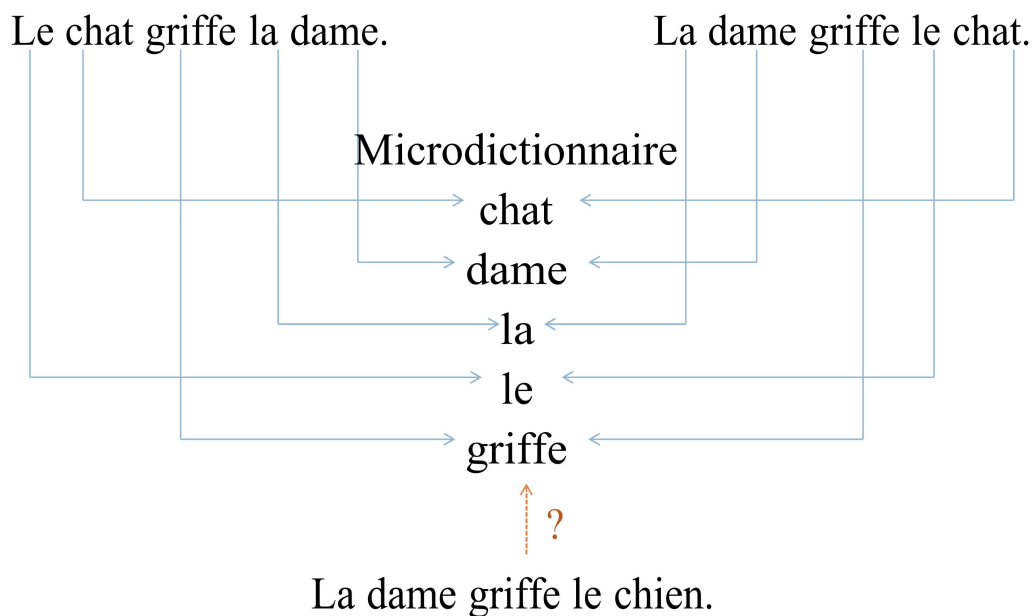
Il reconnaît donc les mots des phrases suivantes :

- Le chat griffe la dame
- La dame griffe le chat

Un programme parcourt ces deux phrases, segmente chaque mot contenu dans les phrases puis les recherche dans le dictionnaire.

Remarque : si la forme n'apparaît pas, le programme ne peut établir de correspondances, si l'on substitue « chien » à « chat », le microdictionnaire ne peut le reconnaître.

Exemple de reconnaissance avec dictionnaire



3.2 Dictionnaire et délimitation du mot

Les mots ne peuvent être tous découpés au moyen de repères graphiques : formes amalgamées (au, du), les mots possédant un caractère non alphabétique ainsi que les mots complexes (pomme de terre). La technique la plus simple pour traiter cette catégorie de mots consiste à ajouter tous ces mots tels qu'ils sont dans le dictionnaire. Toutefois il faut assortir ces mots de renseignements particuliers lors de l'analyse morphosyntaxique (ex : du (de +le)).

De plus, il est impératif d'ajouter des renseignements sur les irrégularités relatives à la flexion des mots et à leur graphie (gratte-ciel/gratte ciel, homme-grenouille/hommes grenouilles). Il faut inclure aussi les mots composés tels que : pomme de terre, chemin de fer et pour lesquels une traduction de surface donnerait : steel path, soil apple ou encore path of steel ou apple of soil.

Dictionnaire
Chemin de fer
Aujourd'hui
du
nom
Pomme de terre
Train d'atterrissage
Prête-nom

Exemple de dictionnaire sans catégorisation

3.3 Mots complexes source d'analyses erronées

Les mots complexes sont souvent une source d'ambiguïtés pour le traitement automatique, d'autant plus lorsqu'ils sont dans un contexte phrastique ambigu incluant plusieurs schémas syntaxiques.

Exemple : pomme de terre, col bleu

- J'ai recouvert la pomme de terre.
- Je mange la pomme de terre.
- J'aime le col bleu de cette chemise.
- Les cols bleus sont en grève.

4 Étiquetage

Jusqu'à maintenant, le dictionnaire élaboré reconnaît les mots corrects sans dire à quelle catégorie ils appartiennent. Cette information est importante pour décrire sa fonction dans la phrase, son mode de flexion ainsi que ses rapports avec les mots avoisinants. Pour ce faire, on indique la catégorie dans le dictionnaire (catégorisation)

4.1 Dictionnaire avec catégorisation

Il existe plusieurs façons de catégoriser les mots dans un dictionnaire. On peut ajouter plus de précision, distinguer les articles définis et indéfinis, possessifs, démonstratifs, verbes transitifs et verbes intransitifs, etc.

Lex_fr	Cat_fr
chat	n
Le	det
mange	v
souris	n
la	dét

Exemple de dictionnaire avec catégorisation

Le linguiste choisit la catégorisation qui convient le mieux à son application mais surtout à la langue vers laquelle il traduit. Un mot peut être grammaticalement polycatégoriel. Un dictionnaire doit tenir compte de ce phénomène. Il doit recenser les interprétations possibles de mots sur le plan catégoriel (homographes)

Exemple :

- Le boucher sale la tranche
- Le <dét>boucher<n> sale<v>la<dét>tranche<n>
- Le <dét>boucher<n> sale<adj>la<pro>tranche<v>
- Le <pro>boucher<n> sale<adj>la<pro>tranche<v>

Lex_fr	Cat_fr
sale	v
la	pro
boucher	n
sale	adj
le	dét
tranche	v
la	dét
le	pro
tranche	n

Exemple de dictionnaire avec mots polycatégoriels

Le système consulte le dictionnaire et donne toutes les interprétations possibles. La simple consultation du dictionnaire ne permet pas au système de choisir entre toutes ces

interprétations. De plus, ces mots polycatégoriels ont des équivalents distincts dans la langue cible.

tranche -- > قطعة لحم

tranche -- > يقطع

4.2 Variation flexionnelle

Les dictionnaires que nous avons établis jusqu'à maintenant ne tiennent pas compte des variations flexionnelles (une seule entrée pour le verbe *manger* [mange] et *trancher* [tranche]. il ne peut pas à ce stade traiter les autres formes fléchies ou la forme canonique du même verbe ainsi que le pluriel des noms et des adjectifs.

- Le chat a mangé la souris
- Les bouchers salent la tranche
- Les chats mangent la souris.
- Etc.

Il existe deux techniques pour traiter les variations flexionnelles :

- Dictionnaires de formes. (en extension)
- Dictionnaires de formes canoniques avec application d'une analyse morphologique. (en intention)

4.3 Dictionnaire de formes (en extension)

On répertorie simplement toutes les formes que le mot peut prendre. La plupart des noms donnent lieu à deux entrées, une forme au singulier et une au pluriel, Les adjectifs donnent lieu à une forme au masculin et au féminin d'une part et au singulier et au pluriel d'autre part. Les formes verbales sont déclinées dans toutes les formes en tenant compte de la personne, du genre, du temps et du mode.

Lex fr	Cat fr
dort	v3psprésind
chat	nms
le	detdefms
mange	v3psprésind
souris	nfs
mangent	v3ppprésind
chats	nmp
souris	nfp
la	detdeffs

Exemple de dictionnaire de formes fléchies

Pour analyser toutes les phrases, on doit enrichir de plus en plus le dictionnaire de formes.

Le dictionnaire de formes :

- est souvent volumineux avec un grand nombre d'entrées.
- ne rend pas compte que des formes telles que « manges, mangent, mangeait,...) alors qu'elles proviennent toutes d'une même forme canonique (manger).

4.4 Dictionnaire de formes canoniques

Dans ce type de dictionnaires, on ne met que les formes canoniques dans le dictionnaire et on ne retient que l'information commune à toutes les variantes du mot. On réduit de ce fait le nombre d'entrées.

Lex_fr	Cat_fr
manger	v
chat	nm
souris	n

Le dictionnaire de formes canoniques risque de ne pas reconnaître l'ensemble des mots. Il faut donc des règles qui associent les formes canoniques et leurs variables apparaissant dans le texte. De plus, Les morphèmes de flexion apportent des informations importantes sur le genre, le nombre, la personne, etc.

5 De l'analyse du mot à celle de la phrase

Précédemment, nous avons décrit des techniques de reconnaissances des mots et de leurs catégories grammaticales. L'analyse morphologique fournit des renseignements grammaticaux supplémentaires (genre, nombre, personne). Cependant, en restant à ce niveau, l'interprétation de la structure de la phrase fait encore défaut. Elle est pourtant essentielle pour comprendre les modalités selon lesquelles les mots se combinent entre eux et régler les ambiguïtés restantes.

5.1 L'analyse syntaxique

La technique menant à la représentation d'une phrase s'appelle l'analyse syntaxique. Le comportement des unités linguistiques d'une phrase est généralisable à des ensembles d'unités. Cela permet la création de règles syntaxiques. L'analyse syntaxique et les règles conçues pour la réaliser prennent diverses formes.

Nous en présenterons deux ici :

- La 1ère envisage les phrase comme des séquences linéaires de mots ;
- La 2ème traite les phrases comme des ensembles structurés hiérarchiquement.

5.2 La phrase, séquence linéaire de mots

Les mots, lorsqu'ils forment des phrases, s'ordonnent en respectant certaines règles. Il existe plusieurs façons d'en rendre compte :

On peut envisager la phrase tout d'abord comme une **séquence linéaire de mots ou de catégories grammaticales**. Ainsi la phrase « Le chat griffe la dame. » peut être présentée par :

Le chat griffe la dame

dét+ n + v + dét + n



La seconde possibilité est le recours, par les analyseurs syntaxiques, à des graphes qui présentent des séquences de catégories grammaticales. Ainsi la phrase « Le chat griffe la dame. » peut être présentée par ce graphe [dét → n → v → dét → n], associé au dictionnaire suivant :

Dictionnaire

chat, n

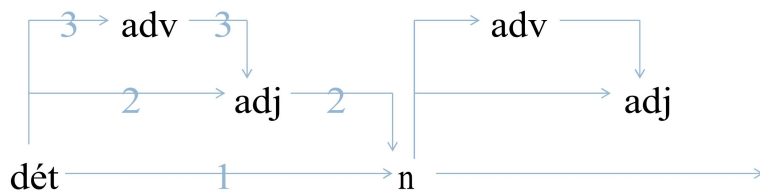
dame, n

le, dét

la, dét

griffe, v

La reconnaissance est ici possible lorsque les dictionnaires indiquent les catégories grammaticales des mots. L'analyseur compare la phrase au graphe. Il compare le 1er mot contenu avec le 1er objet du graphe. Si celui-ci correspond, l'analyse se poursuit ainsi de suite. Si la phrase remplit toutes les conditions du graphe, elle est reconnue. Le graphe reconnaît la phrase si les mots ont la même nature : « Le dame griffe la chat. » sera ainsi reconnue. De même, on peut ajouter facilement des phrases en ajoutant les mots au dictionnaire. Le graphe que nous venons de présenter ne reconnaît pour le moment qu'une seule suite de mots. Les phrases ne sont pas toutes construites sur le même modèle. Il faut donc enrichir le graphe initial en proposant différents parcours :



6 La phrase, unité linguistique structurée hiérarchiquement

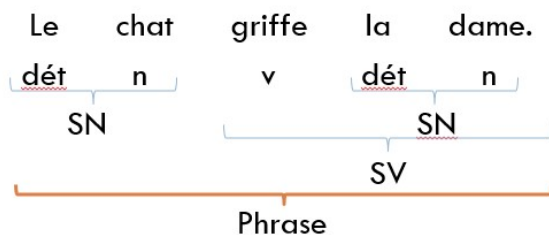
Nous venons de voir une technique d'analyse fondée sur l'examen d'un enchaînement linéaire de catégories grammaticales. La phrase peut également être considérée comme un ensemble structuré hiérarchiquement. Les mots sont groupés dans des ensembles syntaxiques plus grands : des **syntagmes**, qui forment ensuite une phrase.

On distingue différents syntagmes selon l'élément de tête (le plus important du point de vue de l'analyse précédemment effectuée).

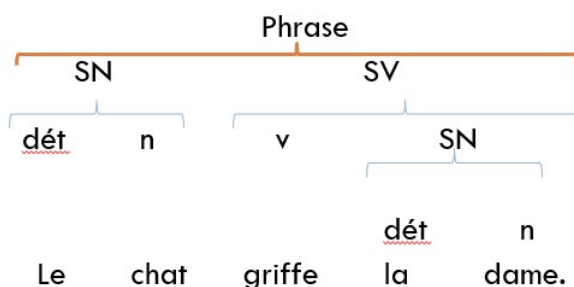
Exemple de différents syntagmes :

- Le syntagme nominal : le petit chat, sa maitresse ;
- Le syntagme verbal : console ;
- Le syntagme prépositionnel : le chat de ma voisine, sa maitresse le garde dans sa maison; Le syntagme adjectival : le chat trop gros ;
- Le syntagme adverbial : le chat sort doucement.

La phrase est alors analysée par paliers, ces derniers étant rangés de façon hiérarchique. De cette manière, la phrase « Le chat griffe la dame. » sera représentée par :



En linguistique ce type de structuration est appelé **arbre syntagmatique**



Certaines grammaires syntagmatiques exploitent des **règles de réécriture**. Prenons un exemple pour les illustrer :

« Le chat griffe la dame. »

Dictionnaire :

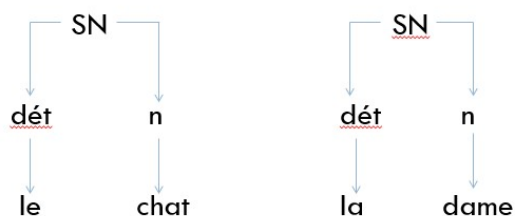
- chat n
- dame n
- le dét
- la dét
- griffe v

Regroupons ces mots en syntagmes et poursuivons jusqu'à l'obtention d'une représentation de la structure complète. Pour notre exemple, il faudra trois règles de réécriture. Elles précisent également l'ordre des éléments.

Elles sont formulées par :

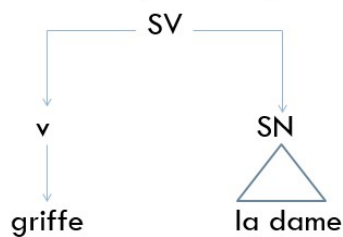
- $P \rightarrow SN + SV$
- $SN \rightarrow \text{dét} + n$
- $SV \rightarrow v + SN$

Prenons la règle $SN \rightarrow \text{dét} + n$, à l'aide de cette règle, l'analyseur syntaxique détermine que « le chat » et « la dame » sont des syntagmes nominaux :

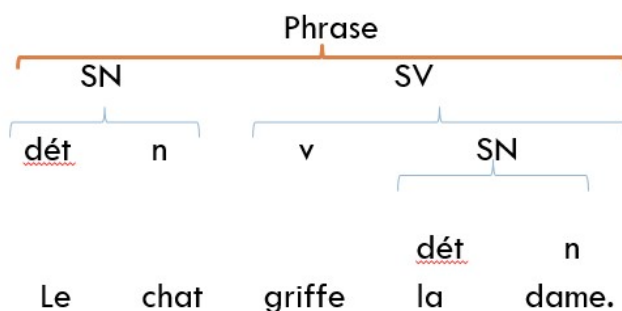


L'analyseur cumule des groupes de mots et tente d'établir des correspondances avec les règles préalablement proposées. Dans les exemples que nous venons de voir, les suites de mots « le chat » et « la dame » sont désormais assimilés à des groupes plus complexes qui sont des syntagmes nominaux. On peut donc ensuite intégrer ces premiers groupes dans d'autres séquences d'un niveau supérieur.

Pour poursuivre l'analyse de la phrase, prenons la seconde règle : $SV \rightarrow v + SN$. Cette règle permet de traiter la partie de phrase « griffe la dame » :



P → SN + SV rend compte de la phrase dans sa globalité « Le chat griffe la dame. ». Elle est représentée par l'arbre suivant :



Ces trois règles permettent d'analyser des phrases du même type comme par exemple : « La dame griffe le chat » ou en complétant le dictionnaire « le chien mord le facteur ». Evidemment, pour analyser un plus grand nombre de phrases, il faut enrichir les règles. Trois règles sont nettement insuffisantes pour rendre compte de l'ensemble des structures de phrases possibles.

Si on ajoute la règle $SP \rightarrow \text{prép} + SN$, on peut analyser les syntagmes prépositionnels. L'analyseur syntaxique pourra donc traiter des phrases telles que : « Le chat du voisin griffe la dame. », « A la maison, le chien mord le facteur. ».

La nouvelle règle $SV \rightarrow v$ complémentaire de la règle $SV \rightarrow v + SN$ permet d'analyser des phrases comme : « Le chien aboie. », « La dame griffe. », « Le chien mord le facteur. ».

Les règles $SN \rightarrow \text{dét} + \text{adj} + n$ ou $SN \rightarrow \text{dét} + N + \text{adj}$ ou encore $SN \rightarrow \text{dét} + n + SP$ rendent compte de syntagmes nominaux plus complexes tels que : « Le chien noir aboie. », « Le chat du voisin griffe la dame. », « La petite dame griffe le chat trop mignon. ».

6.1 Ordre des analyses syntaxiques

L'ensemble des analyses que nous venons de voir sont des **analyses ascendantes**. En effet, à partir des mots eux-mêmes et de l'information grammaticale qu'ils fournissent, l'analyseur remonte jusqu'à la phrase en intégrant ces mots dans des combinaisons dont la complexité va

croissant. On peut, à l'inverse, opter pour une **analyse descendante** dans laquelle l'analyseur suppose que P est composé d'un SN et d'un SV ainsi de suite jusqu'à ce qu'il puisse substituer aux catégories grammaticales les mots reconnus dans la phrase à analyser.

6.2 Correspondance entre les règles de réécriture et les modes de catégorisation

Nous avons vu qu'il était possible de catégoriser les mots de différentes manières : types de verbes (transitifs, intransitifs), types de déterminants (articles, démonstratifs, possessifs), etc. Les règles doivent s'aligner sur le mode de catégorisation choisi. Ainsi des règles de réécriture distinctes peuvent être utilisées pour rendre compte des types de déterminants (SN art + n, SN → poss + n, ...) ou des types de verbes (SV → vt + SN, SV → vi, SV → vi + SP, ...).

6.3 Résolution des ambiguïtés catégorielles

Nous avons vu qu'un programme informatique, à la suite de la consultation d'un dictionnaire, retient toutes les lectures possibles d'une séquence de mots mais ne peut lever les ambiguïtés catégorielles.

Exemple :

Le (dét ou pron)

chien (n)

ronge (v)

son (dét ou n)

os (n).

Les ambiguïtés concernent
les mots « le » et « son ».

L'analyse syntaxique apporte des solutions à ce type de problème. Une des premières méthodes (très utilisée dans les systèmes commerciaux) consiste à trancher pour l'une ou l'autre des catégories après une analyse syntaxique locale. L'analyseur examine les mots qui apparaissent dans le contexte immédiat des mots ambigus et ne retient que les combinaisons de catégories possibles. Si l'on reprend l'exemple juste avant, le mot « son » ne peut pas être un nom car deux noms ne peuvent se suivre (« son os »).

Nous pouvons également confronter ces suites à des graphes qui présentent des phrases admissibles. Si on a un premier graphe qui représente la suite déterminant + nom, certaines des interprétations pour les mots « le » et « son » ne sont plus admises. Si aucun graphe ne représente la suite pronom + nom, certaines d'entre elles aussi ne seront pas admises, etc. Si on confronte ces mêmes possibilités, aux règles syntaxiques posées dans notre exemple, seule une interprétation restera possible. Les règles écrites pourront donc être du type :

- Interdire que « son » en tant que nom puisse être placé entre un verbe et un autre nom.
- « le » pourrait se trouver seulement devant un verbe dans les phrases affirmatives et ne pourrait pas être suivi d'un nom.

La résolution des ambiguïtés catégorielles restent cependant très difficiles à mettre en œuvre lorsqu'une phrase contient plusieurs homographes catégoriels juxtaposés.

Les règles étudiées jusqu'à présent ont pour point commun qu'elles proposent toutes des représentations qui respectent l'ordre des mots dans les phrases analysées. Il peut devenir intéressant de s'écarter de cet ordre dans certaines applications, notamment dans la traduction automatique, puisque cette dernière traite de deux langues dont les mots ne s'ordonnent pas toujours en suivant des règles identiques.

7 Analyse syntaxique et traduction

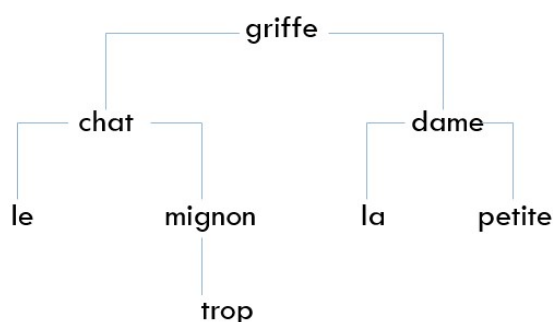
Un programme de traduction doit tenir compte des règles syntaxiques qui régissent au moins deux langues.

Une première série de règles permet de représenter la phrase source → analyse. Une seconde série produit une phrase cible → **génération**, le plus souvent à partir de la représentation de la phrase source.

8 Autres représentations de la phrase

Une première représentation retient, entre les éléments de la phrase, des rapports hiérarchiques différents de ceux qui sont proposés par les règles syntagmatiques. On y identifie les éléments gouverneurs (dont dépendent d'autres unités linguistiques) et les dépendants (qui dépendent d'un gouverneur). Les règles générant ce type de représentations sont appelées **grammaire de dépendance**.

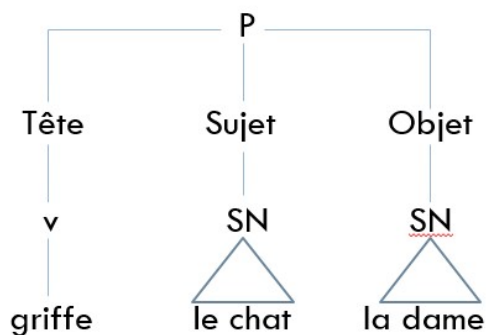
Prenons un exemple avec la phrase : « Le chat trop mignon griffe la petite dame. »



La représentation montre que « griffe » gouverne les noms « chat » et « dame », ces deux noms dépendent donc du verbe. Les déterminants et adjectifs sont gouvernés par les noms « chat » ou « dame », etc. Comme les liens indiquent des rapports de dépendance, le respect de l'ordre des mots dans la représentation n'est plus essentiel (on peut le déduire à partir de celle-ci et de règles d'ordonnance rattachées aux catégories grammaticales).

Une autre forme de représentation identifie les rapports fonctionnels qu'entretiennent les mots entre eux. Les syntagmes, par exemple, sont marqués comme sujet, objet, complément, ...

Ces rapports sont implicites dans la représentation syntagmatique, on peut cependant les rendre explicites par une nouvelle représentation. Il convient de préciser que, même si l'ordre des mots n'est pas explicité dans les deux représentations, il doit figurer dans d'autres parties (grammaires, règles) si l'on veut analyser ou produire des phrases correctes dans une langue donnée. L'analyse de la phrase « Le chat griffe la dame. » produit un arbre retenant les fonctions des groupes par rapport au verbe :



9 Limites des grammaires formelles en TAL

Contrairement à ce que l'on pourrait penser de prime abord, une grammaire formelle contenant l'ensemble des règles syntaxiques ne constitue pas forcément un excellent point de départ pour effectuer l'analyse de n'importe quelle nouvelle phrase. Ce principe, demeurant vrai sur le plan théorique, est en effet difficile à mettre en application dans un programme informatique.

10 Prise en compte du sens

Nous en sommes donc arrivés au stade où notre programme de traitement automatique sait :

- Reconnaître des mots → dictionnaire
- Attacher aux mots des renseignements linguistiques (catégories grammaticales)
- Reconnaître des formes fléchies pour les mots variables → analyse morphologique
- Interpréter la structure de la phrase → analyse syntaxique

Mais il ne connaît pas encore le sens des mots ni la compatibilité des mots entre eux.

10.1 Informations sémantiques sur les mots

La consignation des renseignements sémantiques sur les mots se fait d'une manière semblable à l'indication de la catégorie grammaticale, il faut enrichir les entrées du dictionnaire.

L'identification des **classes sémantiques** se fait quelques fois sous forme de traits. Chacune des entrées du dictionnaire est décrite au moyen d'un nombre fini de traits sémantiques qui rendent compte des sens des mots et qui permettent de les distinguer les uns des autres.

10.2 Marquage des classes sémantiques par des traits

Exemple :

Marquage pour trois mots : « dame », « garçon » et « fille »

- Dame : <+ humain>, <+ femelle>, <+ adulte>
- Garçon : <+ humain>, <- femelle>, <- adulte>
- Fille : <+ humain>, <+ femelle>, <- adulte>

« garçon » ne possède pas les traits <femelle et adulte>.

« fille » ne possède pas le trait <adulte>

Une autre technique consiste à attribuer à chacun des mots une classe sémantique. Par exemple, on peut dire que le mot « chat » appartient à la catégorie des noms concrets comme « dame », « os », « voiture », etc.

On peut aussi dire que le mot « analyse » appartient à la catégorie des noms d'action comme « traitement », « traduction ». Les classes choisies ne valent que pour les noms, pour d'autres catégories les critères ne sont pas les mêmes.

Il est possible d'affiner davantage le classement sémantique. Par exemple, les noms concrets peuvent désigner des êtres animés (dame, chat, ...) ou des objets inanimés (os, voiture, ...). Il peut également y avoir des distinctions dans les êtres animés comme les humains (dame, voisin) et les non-humains (chat, chien) et ainsi de suite.

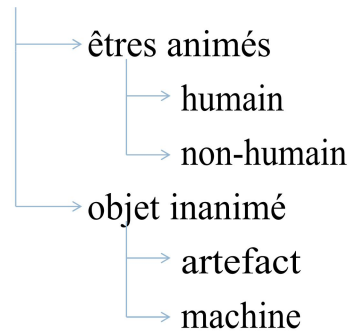
Exemples :

- Chat, nom masculin, concret, être animé, non-humain
- Os, nom masculin, concret, objet inanimé

En outre, les classes sémantiques ainsi définies sont souvent hiérarchisées, cela évite de surcharger le dictionnaire. Si l'on reprend le mot « dame », il suffit de dire qu'il s'agit d'un humain. Le programme qui a accès à des classes hiérarchisées déduit que le mot est lié aux catégories [êtres animés] et [noms concrets]. Les classes spécifiques héritent de toutes les caractéristiques associées à ces classes génériques.

Exemple de hiérarchisation des classes sémantiques :

Noms concrets :



Sans cette organisation hiérarchique, il faudrait expliciter tous les éléments de la hiérarchie dans le dictionnaire. L'explicitation du sens des mots sert à vérifier la compatibilité sémantique des mots et à résoudre des difficultés comme :

- des ambiguïtés sémantiques ;
- des ambiguïtés de structures.

10.3 Vérifications des compatibilités sémantiques entre les mots

Le traitement automatique vérifie les compatibilités sémantiques entre les mots d'une phrase si les classes sémantiques des mots figurent dans le dictionnaire et si les compatibilités entre les mots sont indiquées dans un répertoire (qui peut être complémentaire du dictionnaire).

L'information sémantique, telle qu'elle est exploitée en TAL, est presque toujours tributaire de l'analyse sémantique. Donc, pour voir si deux mots sont compatibles sémantiquement, il faut avant tout comprendre les liens syntaxiques qui les unissent.

Illustrons la technique par un exemple :

- Le chat mange sa pâtée.
- Sa pâtée mange le chat.

Pour rendre compte que le premier exemple est correct et que le second est incorrect, il faut indiquer les classes de mots que le verbe « manger » accepte :

- l'action « manger » est réalisée par des êtres animés
- Les éléments qui subissent cette action doivent appartenir à la catégorie des matières comestibles.

D'un point de vue plus technique, nous dirons, lorsque nous noterons les classes de mots sélectionnés par un autre mot, que nous explicitons ses **restrictions de sélection**.

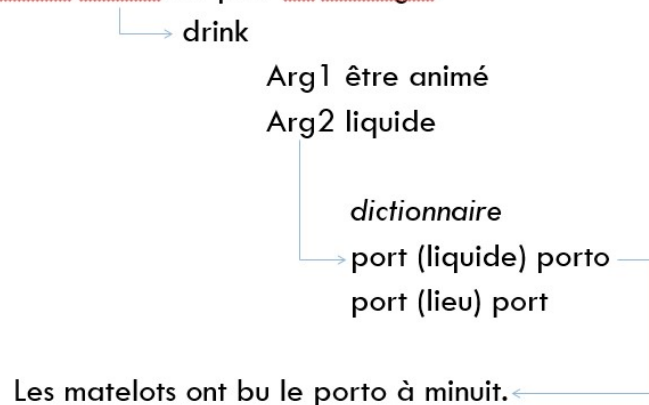
« Manger » sélectionne des arguments dénotant des êtres animés et des matières comestibles. Les restrictions de sélection précisent les conditions que doivent remplir les mots pour être compatibles avec d'autres.

10.4 Résolution des ambiguïtés sémantiques

Certains mots véhiculent plus d'une signification, on parle d'**homographes sémantiques**. La résolution des ambiguïtés sémantiques peut se faire au moyen de la vérification des **restrictions de sélection**. Pour analyser et par conséquent, traduire convenablement la phrase « The sailors drank the port at midnight », on doit distinguer le sens de « port ». Or, un dictionnaire lexical indique les entrées et les équivalences sans lever les ambiguïtés.

Par contre, si la classe sémantique à laquelle se rattachent les deux sens de « port » est indiquée (port(liquide), port(lieu) ainsi que les restrictions du verbe « drink », l'ordinateur va pouvoir résoudre l'ambiguïté. Il consulte le dictionnaire pour connaître les classes sémantiques des mots et constate ainsi que la première acception du mot « port » est classée comme liquide mais pas la deuxième. Il rejette donc la deuxième possibilité au profit de la première car le verbe l'impose.

The sailors drank the port at midnight.



Il existe une autre technique de désambiguïsation des homographes sémantiques utilisée en traduction automatique : associer chacun des sens des homographes à un domaine précis (domaine de spécialité). Le système déclenche alors la consultation d'un dictionnaire relatif à ce domaine.

Exemple :

opération (médecine) -- > opération (militaire).

coquille (typographie) -- > coquille (zoologie).

Il est parfois impossible de lever les ambiguïtés même avec des restrictions de sélection :

Exemple :

they passed the port at midnight --> Les deux mots (passe) et (port) sont ambigus (il n'y a pas de repères)

De plus, la vérification des affinités sémantiques est assez difficile à réaliser en pratique pour l'ensemble des combinaisons de mots.

Les techniques de vérification des restrictions de sélection servent à lever certaines ambiguïtés sémantiques, mais elles peuvent également être mises à contribution pour résoudre des ambiguïtés de structure

- Je mange mes pâtes à la cuillère.
- Je mange mes pâtes au Comté.

D'un point de vue syntaxique, un SP ne peut s'attacher qu'à un nom ou un verbe.

« à la cuillère », « au comté » sont considérés, soit comme des compléments du verbe, soit comme des compléments du nom.

□ Je mange mes pâtes à la cuillère.

□ Je mange mes pâtes au comté.



En réalité, « à la cuillère » est complément du verbe « manger » et « au comté » est complément du nom « pâtes ».

Nous pouvons imposer des restrictions de sélection sur la structure.

- Manger (arg1<animé>, arg2 (matière comestible>, arg3<artefact>)
- Cuillère appartient à la troisième classe (artefact).
- Comté n'appartient pas à cette classe donc il est analysé autrement comme un complément du nom.

11 Conclusion

Tout ce que nous avons vu jusqu'à maintenant s'appuie sur le principe de compositionnalité. Cela suppose que les sens des mots d'une phrase sont cumulés et que le sens global de la phrase résulte de ce cumul. Toutes les phrases ne peuvent pas s'analyser de cette façon, puisque certaines comportent des expressions figées.

12 Bibliographie

Boitet Christian et Nedobejkine Nikolai, « Recent developments in Russian-French Machine Translation at Grenoble », *Linguistics*, 19, 1981, pp. 199-271

Bourquin Guy, « Quel statut épistémologique donner à la traduction automatique », *Contrastes*, A4, Paris, ADEC, 1984, pp. 111-119.

Bourquin-Launey Marie-Claude, « De la traduction automatique à la réflexion linguistique : la traduction de la préposition anglaise ‘through’ », *Contrastes*, A4, Paris, ADEC, 1984, pp. 81-99.

Danlos Laurence, « Degré d’abstraction des représentations intermédiaires en traduction automatique, un exemple : Eurotra », *T.A. Informations*, 31-1, Paris, ATALA, 1990, pp. 25-37.

Delavenay Emile, « La traduction automatique », in *Le Langage*, encyclopédie de la Pléiade, Paris, NRF, 1968, pp. 758-771.

Delavenay Emile, *La Machine à traduire*, P.U.F., coll. *Que sais-je ?*, n° 834, Paris, 1963 (1^{re} éd. 1950)

Fuchs Catherine, et alii., *Linguistique et Traitements Automatiques des Langues*, Paris, Hachette – Supérieur.

Hutchins W. John, « Linguistics models in machine translation », *University of East Anglia Papers in Linguistics*, n° 9, 1979

Isabelle Pierre, « Current research in Machine Translation : A Reply to Somers », *Machine Translation*, 7-4, Kuwer Academic Press, 1992/1993, pp. 265-272.

Loffler-Laurian Anne-Marie, « Remarques à propos de quelques exemples de traduction assistée par ordinateur et la post-édition », *Contrastes*, 2, Paris, ADEC, 1981, pp. 63-70

Loffler-Laurian Anne-Marie, « Remarques à propos de quelques exemples de traduction assistée par ordinateur et la post-édition », *Contrastes*, 2, Paris, ADEC, 1981, pp. 63-70.

Loffler-Laurian Anne-Marie, « Traduction automatique et périphériques : évaluation, post-édition, attitudes, formation », *Contrastes*, A4, Paris, ADEC, 1984, pp. 43-67.

13 Exercices

Exercice 1 :

Quels problèmes de TAL rencontre-t-on dans ce texte ? Trouvez des exemples, des explications, pensez aux solutions et améliorations qu'on peut y apporter.

Texte original :

Michael Douglas and Catherine Zeta-Jones are going to have a busy year.

They'll be tying the knot and welcoming a little bundle of joy.

"Catherine Zeta-Jones and Michael Douglas are pleased to announce they are expecting a child. It will be the first child for Ms. Zeta-Jones. Mr. Douglas has a son, Cameron, from a previous marriage," reads an official statement from the couple.

An unnamed source says the couple, who broke the news to her parents last week, are thrilled.

"Catherine has always said she wants to start a family. And she knows Michael is the man for her," dished the source.

The 55-year-old actor and the 30-year-old actress announced their shotgun engagement earlier this month via Douglas' Website.

"I am very happy to announce that I am engaged to Catherine Zeta-Jones. I proposed to Catherine on New Year's Eve at my house in Aspen." We plan to marry sometime this year, however, no date has been set. Check my Website in the future and you'll be among the first to know."

A date has still not been set for the impending nuptials, though some gossip mavens suggest they'll do it on their joint birthday, September 25, on the Spanish island of Majorca.

Vocabulaire:

To tie the knot: se marier

A bundle of joy: un bout de chou

To dish: répandre des ragots

Shotgun: rapide; fusil

Impending: imminent

Gossip: commérages

Maven: expert (péj)

Traduction Systran:

Michael Douglas et Catherine Zeta-Jones vont avoir une année occupée. Ils attacheront le noeud et feront bon accueil à un petit paquet de joie. " Catherine Zeta-Jones et Michael Douglas sont heureux de les annoncer attendent un enfant. Ce sera le premier enfant pour mme. Zeta-Jones. M. Douglas a un fils, Cameron, d'un mariage précédent, " lit un rapport officiel des couples. Une source anonyme indique les couples, qui ont cassé les nouvelles à ses parents la semaine dernière, sont faites frémir. " Catherine a toujours dit qu'elle veut commencer une famille. Et elle sait que Michael est l'homme pour elle, " a bombé la source. L'acteur 55-year-old et l'actrice 30-year-old ont annoncé leur enclenchement de fusil de

chasse plus tôt ce mois par l'intermédiaire de Douglas Website. " je suis très heureux d'annoncer que je suis engagé à Catherine Zeta-Jones. J'ai proposé à Catherine la veille de nouvelle année à ma maison dans Aspen. " Nous projetons marier autrefois cette année, cependant, aucune date n'a été fixée. Contrôlez mon Website à l'avenir et vous serez parmi le premiers à savoir. " Une date n'a toujours pas été fixée pour les nuptials imminents, bien que quelques mavens de bavardage suggèrent qu'ils le fassent sur leur anniversaire commun, septembre 25, sur l'île espagnole de Majorca.

Traduction Freetranslation :

Michael Douglas et Catherine Zeta-Jones va à a une année occupée. Ils lieront le noeud et l'accueil d'un petit paquet de joie. "Catherine Zeta-Jones et Michael Douglas plu devrait annoncer ils prévoient un enfant. Il sera le premier enfant pour Mme Zeta-Jones. M. Douglas a un fils, Cameron, d'un mariage précédent," lit une déclaration officielle du couple. Une source de unnamed dit que le couple, qui s'est annoncé la nouvelle à sa semaine dernière de parents, sont passionné. "Catherine toujours a dit qu'elle veut au début une famille. Et elle sait que Michael est l'homme pour elle," dished la source. Le 55-ans l'acteur et le 30-ans l'actrice a annoncé leur engagement de fusil plus tôt ce mois via Douglas' Website. "Je suis très heureux à annonce que je suis engagé à Catherine Zeta-Jones. J'ai proposé à Catherine sur la Veille de l'année Nouvelle à ma maison dans Aspen." Nous planifions à épouse un jour cette année, cependant, aucune date a été réglée. Vérifier mon Website à l'avenir et vous serez entre le premier à sait." Une date n'a pas été toujours réglée pour le nuptials imminent, bien que quelque mavens de commérage suggère qu'ils le fassent sur leur anniversaire de jointure, le 25 septembre, sur l'île espagnole de Majorca.

VOILA Traduction :

Michel Douglas et Catherine Zeta-Jones vont avoir une année occupée. Ils lieront le noeud et l'accueil d'un petit bout de chou. "Catherine Zeta-Jones et Michel Douglas ont le plaisir d'annoncer qu'ils attendent un enfant. Ce sera le premier enfant pour Mme Zeta-Jones. M. Douglas a un fils, Cameron, d'un mariage précédent," lit une déclaration officielle du couple. Une source anonyme dit le couple, qui a communiqué les nouvelles à ses parents la semaine dernière, est transporté. "Catherine disait toujours qu'elle veut commencer une famille. Et elle sait que Michel est l'homme pour elle," dished la source. L'acteur de 55 ans et l'actrice de 30 ans ont annoncé leur engagement de fusil de chasse au début du mois via le Site Web de Douglas. "Je suis très heureux d'annoncer que je suis engagé à Catherine Zeta-Jones. J'ai proposé à Catherine sur la Saint-Sylvestre à ma maison dans le Tremble." Nous projetons de nous marier un jour ou l'autre cette année, cependant, aucune date n'a été mise. Vérifiez mon Site Web dans l'avenir et vous serez parmi le premier pour savoir(connaître). " Une date n'a toujours pas été mise pour la noce menaçante, quoique quelques experts de commérage suggèrent qu'ils le fassent sur leur anniversaire commun, le 25 septembre, dans l'île espagnole de la Majorque.