

Cours Traitement du langage naturel

Chapitre 1 Introduction au TLN



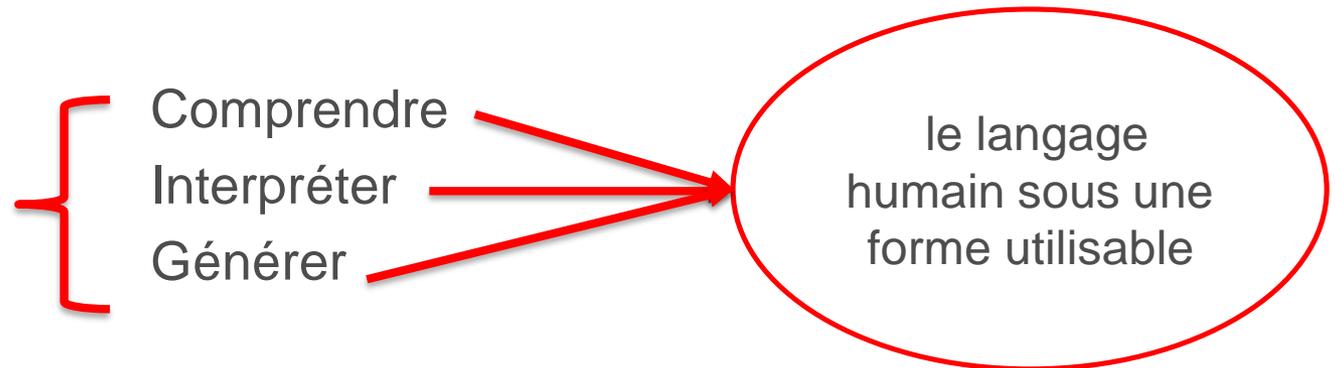
Contenu du premier chapitre

1. Définition et importance du TLN
2. Niveaux de traitement d'un langage :
 - Phonétique, phonologie ,orthographe
 - Morphologie et syntaxe
 - Sémantique
 - Pragmatique et discours
3. Applications du TLN



1. Définition et importance du TLN

- ❑ Le traitement du langage naturel (TLN) est un domaine interdisciplinaire de l'intelligence artificielle (IA) et de la linguistique informatique qui permet aux ordinateurs de:



- ❑ L'objectif est de faciliter **l'interaction** entre les machines et les humains en utilisant un langage que nous comprenons tous les jours.



1. Définition et importance du TLN

Qu'est-ce que le TLN ?

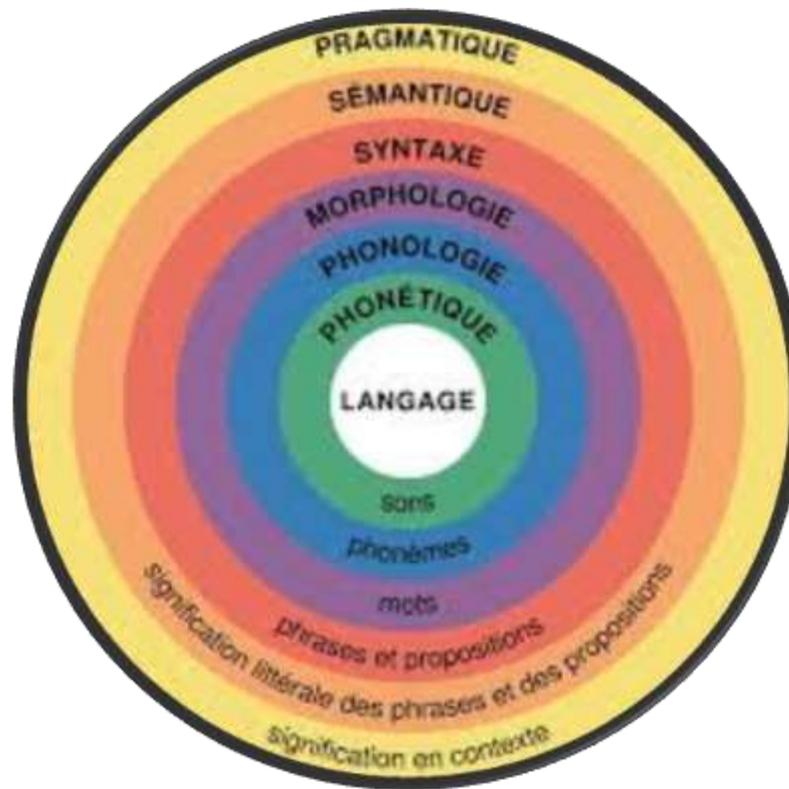
- ❑ Le TLN regroupe des **méthodes** et des **algorithmes** permettant aux ordinateurs de traiter **texte** et **parole** comme un être humain.

- ❑ Tâches principales :
 - Traduction automatique
 - Génération de texte
 - Analyse des sentiments
 - Reconnaissance vocale

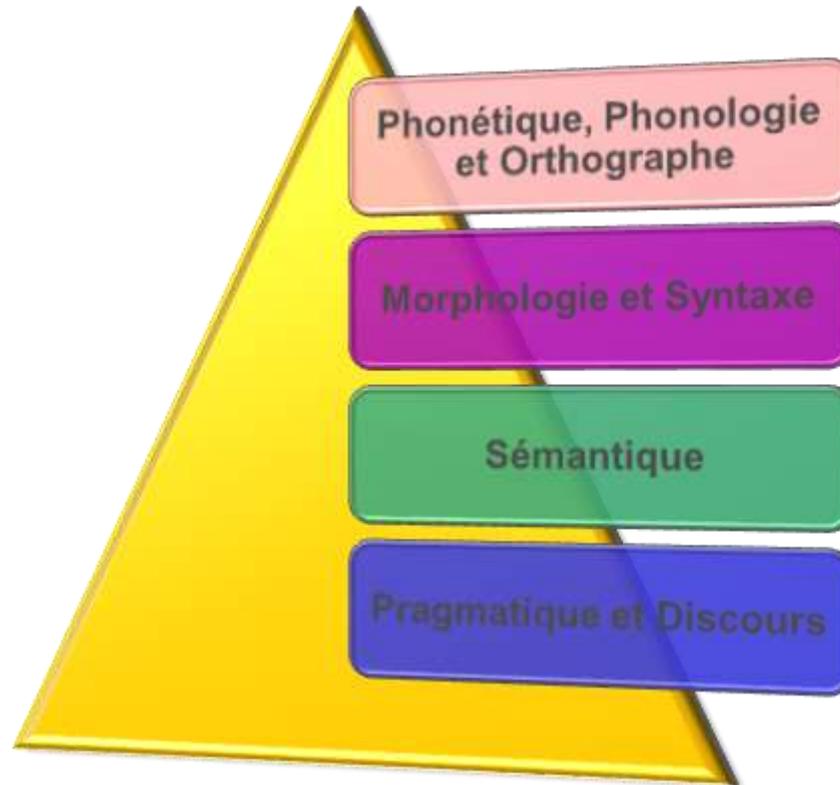


2. Niveaux de traitement d'un langage

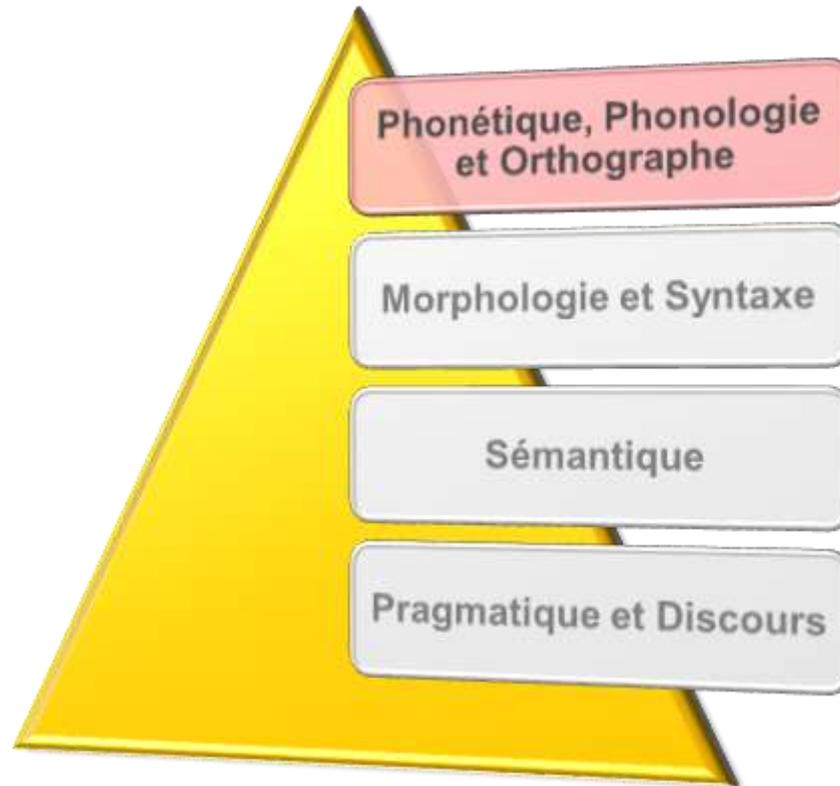
Le traitement du langage se divise en plusieurs niveaux d'analyse, chacun ayant pour but de décomposer et de comprendre les éléments fondamentaux du langage humain.



2. Niveaux de traitement d'un langage



2. Niveaux de traitement d'un langage



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

A. La phonétique

- La **phonétique** est la branche de la linguistique qui étudie les **sons du langage** sous leurs aspects acoustiques et articulatoires.
- Elle se concentre sur la manière dont les sons sont produits, transmis et perçus.
- Elle est divisée en trois grandes catégories :



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

La phonétique

Phonétique
articulatoire

Phonétique
acoustique

Phonétique
auditive



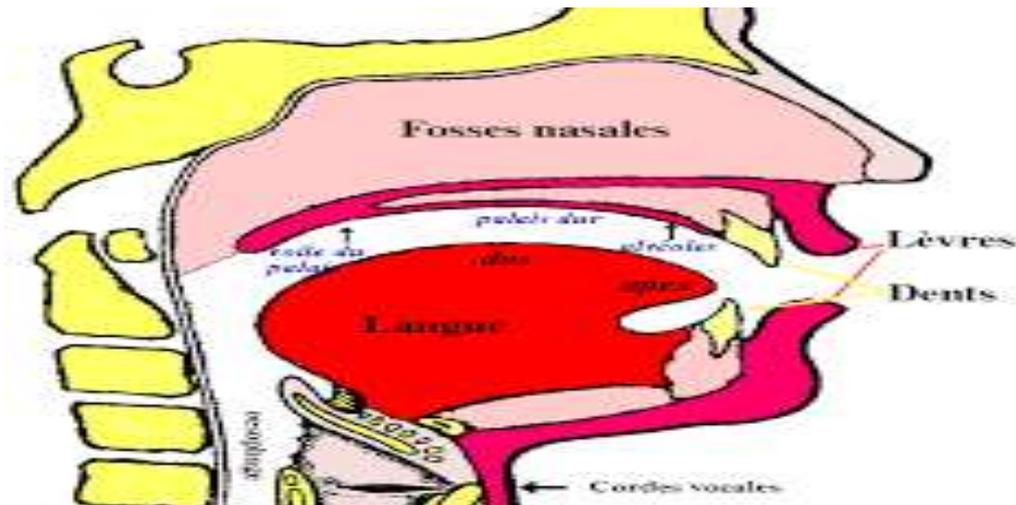
2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

A. La phonétique

Phonétique articulatoire

- Elle analyse la production des sons par les organes de la parole (les lèvres, la langue, le palais, les cordes vocales, etc.).
- Cela inclut l'étude des différents mouvements musculaires nécessaires à la création de sons spécifiques.



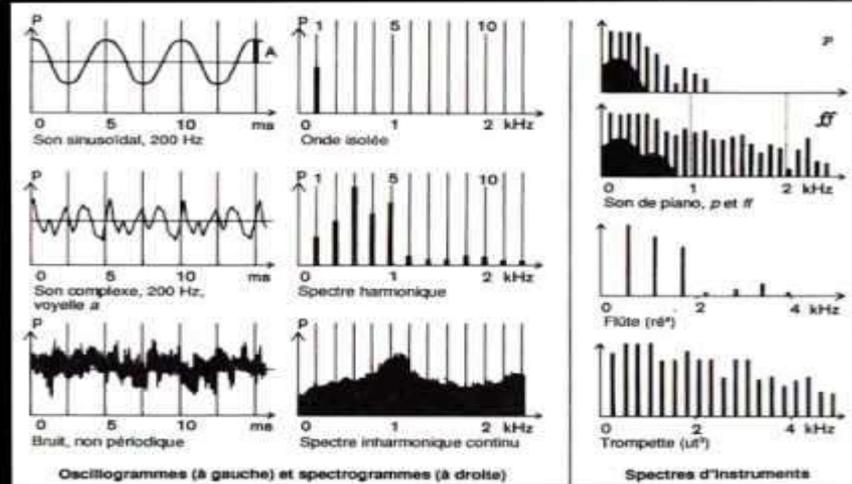
2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

A. La phonétique

Phonétique acoustique

- Elle se focalise sur les propriétés physiques des sons eux-mêmes, comme la fréquence, l'amplitude, et la durée.
- Ce qui permet de comprendre la manière dont les sons voyagent sous forme d'ondes sonores dans l'air



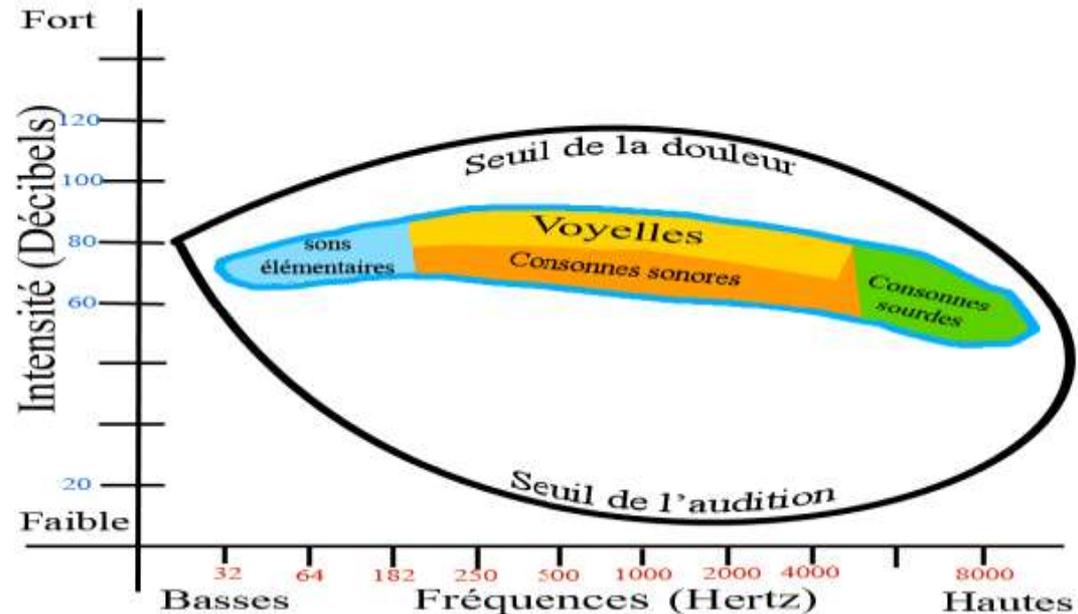
2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

A. La phonétique

Phonétique auditive

- Elle s'intéresse à la façon dont les sons sont perçus par l'oreille humaine et interprétés par le cerveau, en analysant des facteurs comme la perception de la hauteur, du timbre et du rythme des sons



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

B. Phonologie

- ❑ La **phonologie** est l'étude des **sons du langage en tant qu'éléments abstraits** et de leur organisation dans une langue.
- ❑ Contrairement à la phonétique, qui se concentre sur les aspects physiques des sons, la phonologie cherche à comprendre comment les sons sont utilisés dans un langage donné et quelles règles régissent leur agencement.



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

Concepts clé de la Phonologie

Phonèmes

Comportement
des phonèmes

Règles
phonologiques



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

B. Phonologie

Phonèmes

- ❑ La phonologie s'intéresse aux **phonèmes**, les unités minimales de son qui peuvent distinguer un mot d'un autre.
- ❑ Par exemple, en français, les mots «"pas" » et «"bas" » diffèrent par un seul phonème, à savoir /p/ et /b/.



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

B. Phonologie

Comportement des phonèmes

- ❑ La phonologie étudie aussi les **variantes phonétiques** d'un phonème.
- ❑ Par exemple, un phonème peut avoir plusieurs réalisations en fonction du contexte : altération du son, assimilation, élision, etc.

Dans "loup", le // est prononcé avec la pointe de la langue au contact des dents.

Dans "ville", il peut être légèrement influencé par le son suivant et paraître plus "clair" ou "foncé".



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

B. Phonologie

Règles phonologiques

Expliqué
en
python

- ❑ La phonologie analyse aussi les règles qui régissent les interactions entre les phonèmes dans une langue.
- ❑ **Par exemple**, dans certaines langues, un phonème peut changer en fonction de la position dans le mot ou de son environnement sonore

On ne dit pas **/le ami/**, mais **/lez_ami/**.

La règle : le son final **/z/** de "les" est prononcé uniquement parce qu'il y a une voyelle ("amis") qui suit. C'est une interaction entre les phonèmes **/z/** et **/a/**.

2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

C. Orthographe

Qu'est-ce que l'orthographe ?

- Représentation correcte des sons du langage à l'aide de **lettres** et **symboles**.

- Influencée par :
 - **Évolution historique** de la langue.
 - **Règles grammaticales**.



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

C. Orthographe

Rôle de l'orthographe dans le TLN

Reconnaissance des mots écrits :

- Reconnaissance de l'écriture manuscrite.
- Correction orthographique.



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

Les principaux aspects de
l'orthographe

Orthographe
phonétique

Orthographe
grammaticale

Orthographe
lexicale



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

C. Orthographe

Orthographe
phonétique

Expliqué
en
python

- ❑ **Correspondance univoque** : Chaque phonème est transcrit par un seul symbole, sans exceptions.

Dans un système phonétique, le son /f/ serait toujours écrit de la même manière, quel que soit le mot.

- ❑ En revanche, d'autres, comme l'orthographe française, sont moins réguliers et comportent de nombreuses exceptions.



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

C. Orthographe

Orthographe
grammaticale

Elle implique l'accord des mots avec le reste de la phrase (accord en genre, en nombre, conjugaison des verbes, etc.).

Je mange des banane.

des bananes

Ignorer



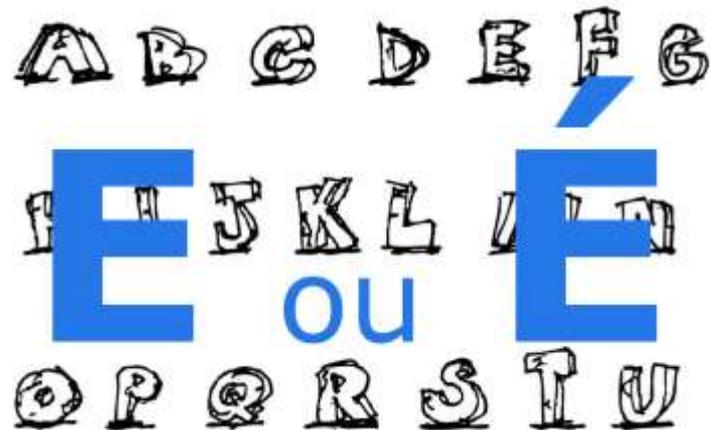
2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

C. Orthographe

Orthographe lexicale

Chaque mot a une orthographe spécifique qui doit être apprise, comme les règles d'accentuation ou de majuscules, et les exceptions à ces règles.



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

C. Orthographe

Orthographe lexicale

Désigne la manière standardisée d'écrire les mots d'une langue dans leur forme correcte selon les règles et conventions établies.

Dans le TLN, l'orthographe lexicale est cruciale pour des tâches telles que :

La correction orthographique : Détection des erreurs dans un texte et suggestion de la forme correcte des mots.

Par exemple, des outils comme **Grammarly** ou **Microsoft Word** utilisent des modèles pour analyser un texte et proposer des corrections.



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

C. Orthographe

Orthographe
lexicale

Dans le TLN, l'orthographe est cruciale pour des tâches telles que :

La reconnaissance de texte manuscrit (OCR) : Transformation de l'écriture manuscrite en texte numérique à savoir : la reconnaissance des lettres et des mots, souvent en prenant en compte les variations de style d'écriture.

La recherche et l'indexation de texte : Les moteurs de recherche doivent comprendre l'orthographe des mots pour indexer et retrouver des documents de manière efficace.



2. Niveaux de traitement d'un langage

2.1 Phonétique, Phonologie et Orthographe

Relations et Interactions

Bien que la phonétique, la phonologie et l'orthographe soient des domaines distincts, ils sont étroitement liés et interagissent dans le cadre du traitement du langage naturel :

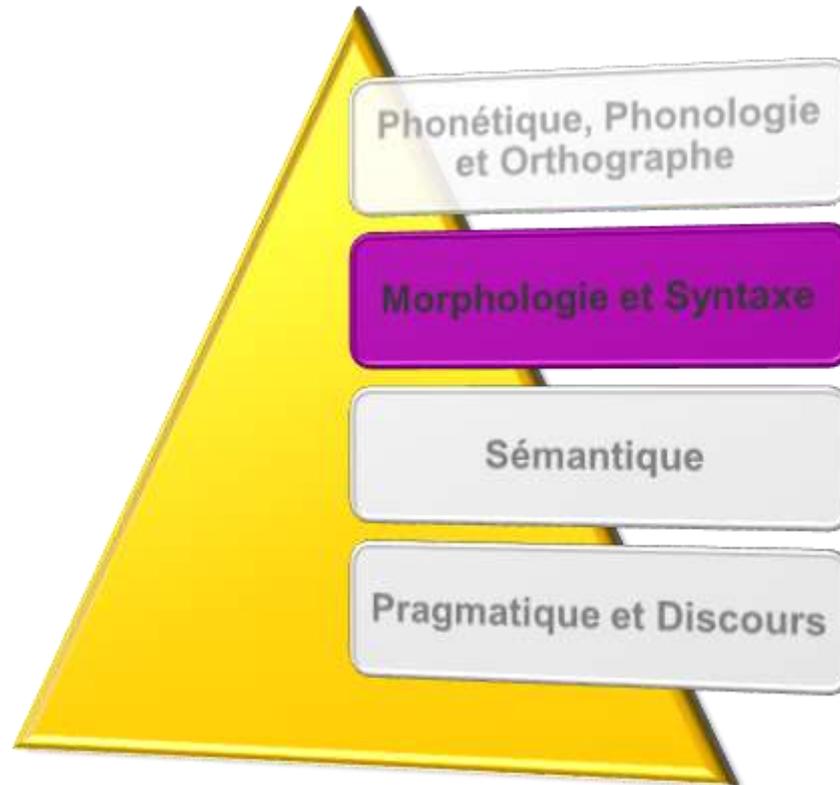
De la phonétique à l'orthographe : Dans des systèmes comme la **synthèse vocale**, les modèles doivent être capables de transformer un texte écrit en une séquence de sons à prononcer, ce qui nécessite une compréhension de l'orthographe et de la phonétique.

De la phonologie à la phonétique : La phonologie guide la manière dont les phonèmes sont interprétés et articulés, alors que la phonétique se charge de leur représentation physique.

De la phonologie à l'orthographe : La phonologie aide à comprendre les règles de l'orthographe, en particulier dans les systèmes de transcription phonétique qui cherchent à coder les sons du langage de manière écrite.



2. Niveaux de traitement d'un langage



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

La **morphologie** et la **syntaxe** occupent des rôles clés pour comprendre la structure des mots et des phrases dans une langue.



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Morphologie

- La **morphologie** est la branche de la linguistique qui étudie la structure interne des mots.
- Elle se concentre sur la manière dont les mots sont formés à partir de plus petites unités significatives appelées **morphèmes**.
- Un morphème est la plus petite unité linguistique dotée de sens ou de fonction grammaticale.



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Types de morphologie

Morphèmes lexicaux : Ils portent le sens principal d'un mot.

Par exemple, dans "manger", le morphème "mang-" exprime l'idée principale de l'action.

Morphèmes grammaticaux : Ils ajoutent des informations grammaticales telles que le temps, le nombre ou la personne.

Par exemple, dans "mangerai", le suffixe "-rai" indique le futur.



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Types de morphologie

Morphologie inflectionnelle :

Elle modifie un mot pour exprimer des variations grammaticales sans en changer la catégorie grammaticale ou le sens fondamental.

Exemple : "chat" → "chats" (singulier → pluriel), "mange" → "mangerai" (présent → futur).



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Types de morphologie

Morphologie dérivationnelle :

Elle permet de créer de nouveaux mots en modifiant le sens ou la catégorie grammaticale.

Exemple : "chant" (nom) → "chanteur" (nom) → "chantant" (adjectif).



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Rôle de la morphologie en TLN

- ❑ **Analyser les variantes d'un mot** : Identifier que "mangera", "mangé" et "manger" se réfèrent à la même racine.
- ❑ **Normaliser les données textuelles** : Grâce à des techniques comme la lemmatisation et le stemming, la morphologie aide à réduire les mots à leur forme de base.
- ❑ **Comprendre les langues morphologiquement riches** : Les langues comme l'arabe ou le turc, où un seul mot peut contenir des informations sur le temps, le sujet, et l'objet, nécessitent une analyse morphologique approfondie.



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

La syntaxe

La **syntaxe** s'intéresse à l'organisation et à la structure des phrases.

Elle examine comment les mots et les groupes de mots sont combinés pour former des phrases grammaticalement correctes et significantes.



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Concepts Clés de la syntaxe

Structure
hiérarchique

Règles
syntaxiques

Ambiguïté
syntaxique



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Concepts Clés de la syntaxe

Structure
hiérarchique

Une phrase est composée de plusieurs unités structurelles appelées **constituants**.

Par exemple, dans "**Le chat dort**", "Le chat" est un syntagme nominal (SN), et "dort" est un syntagme verbal (SV).

Expliqué
en
python



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Concepts Clés de la syntaxe

Règles
syntaxiques

Les règles syntaxiques définissent comment les mots peuvent être combinés dans une langue donnée.

en français, l'ordre typique est Sujet-Verbe-Objet (SVO) :
"Jean mange une pomme."



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Concepts Clés de la syntaxe

Ambiguïté
syntaxique

Une phrase peut avoir plusieurs interprétations syntaxiques.

"Le professeur lit le livre dans la bibliothèque"

peut signifier que le professeur est dans la bibliothèque
ou que le livre s'y trouve.



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Importance de la Syntaxe en TLN

❑ Compréhension de la structure des phrases :

Elle est essentielle pour des tâches comme la traduction automatique, où il est crucial de conserver les relations grammaticales entre les mots.

❑ Résolution des ambiguïtés :

Une analyse syntaxique approfondie aide à interpréter correctement des phrases ambiguës.

❑ Amélioration des modèles génératifs :

Les systèmes de génération de texte, comme ceux utilisés dans les chatbots ou la rédaction automatique, utilisent la syntaxe pour produire des phrases grammaticalement correctes.



2. Niveaux de traitement d'un langage

2.2 Morphologie et syntaxe

Interaction entre Morphologie et Syntaxe

La morphologie et la syntaxe sont interdépendantes.

- ❑ *La morphologie d'un mot affecte son rôle syntaxique dans une phrase, comme :*

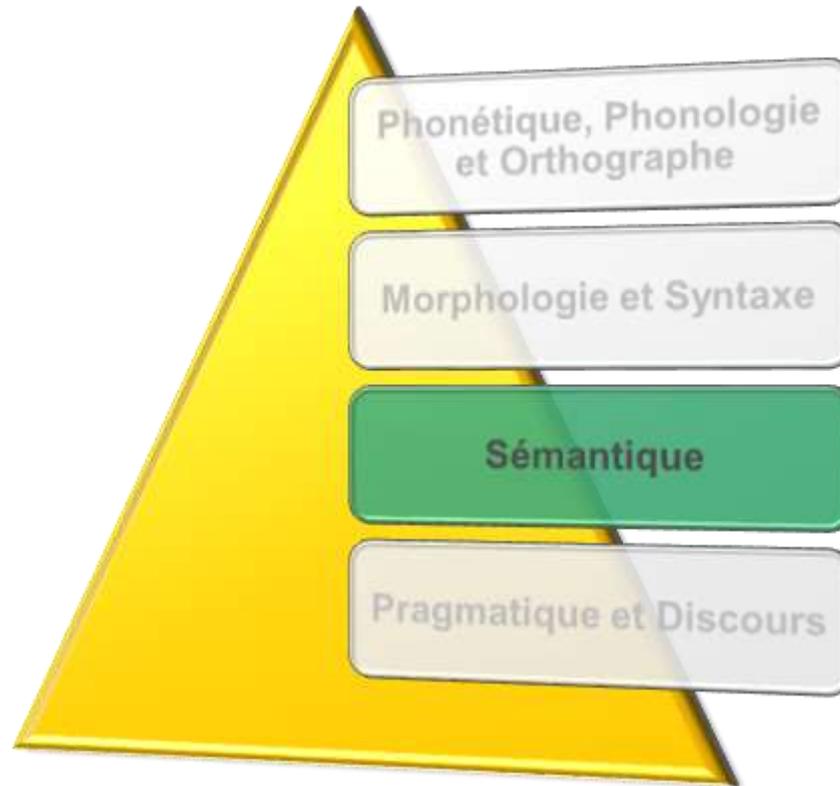
"chante" (verbe) et "chanteur" (nom) seront utilisés différemment dans la structure syntaxique.

- ❑ La syntaxe peut également influencer la morphologie.

les accords grammaticaux dépendent de la structure syntaxique par ex : "Les chats mangent".



2. Niveaux de traitement d'un langage



2. Niveaux de traitement d'un langage

2.3 La sémantique

Définition de la Sémantique

La sémantique étudie les relations entre les **symboles linguistiques** (mots, expressions, phrases) et leur **signification**.

Elle répond aux questions suivantes :

- Que signifie un mot ou une phrase ?
- Comment les significations changent-elles en fonction du contexte ?
- Quelles relations existent entre les différents éléments du langage en termes de sens ?



2. Niveaux de traitement d'un langage

2.3 La sémantique

Techniques de Représentation du Sens

Modèles
Symboliques

Modèles Distribués
(Word Embeddings)

Modèles
Contextuelle



2. Niveaux de traitement d'un langage

2.3 La sémantique

Applications Pratiques de la Sémantique en TLN

❑ Désambiguïsation Lexicale

Déterminer le sens approprié d'un mot dans un contexte donné.

Exemple : "La souris est grise" (animal) vs "La souris est sans fil"
(objet informatique).

❑ Traduction Automatique

Restituer le sens d'un texte d'une langue à une autre tout en respectant les nuances sémantiques.



2. Niveaux de traitement d'un langage

2.3 La sémantique

Applications Pratiques de la Sémantique en TLN

❑ Résumé Automatique

Synthétiser un texte en conservant ses informations principales.

❑ Analyse des Sentiments

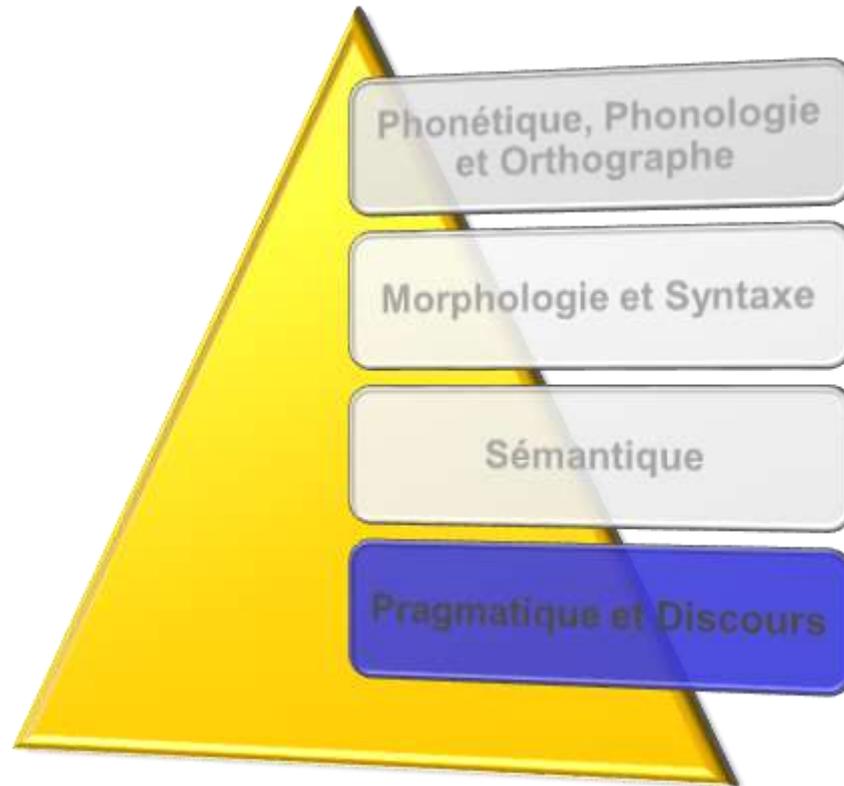
Évaluer l'émotion ou l'attitude exprimée dans un texte, par exemple dans des avis clients ou des publications sur les réseaux sociaux.

❑ Recherche Sémantique

Améliorer les moteurs de recherche pour comprendre les intentions des utilisateurs au-delà de simples correspondances de mots-clés.



2. Niveaux de traitement d'un langage



2. Niveaux de traitement d'un langage

Pragmatique

ET

Discours



2. Niveaux de traitement d'un langage

Pragmatique

Est la branche de la linguistique qui s'intéresse à l'étude de l'usage du langage dans des contextes de communication réels. Elle se focalise sur l'interaction entre les énoncés et leur contexte d'énonciation. Elle s'intéresse à la manière dont les personnes utilisent le langage pour accomplir des actes de communication dans des situations spécifiques



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

La Pragmatique: Actes de Discours (Speech Acts)

Les actes de discours, introduits par Austin (1962) et Searle (1969), sont les unités fondamentales de la pragmatique.

Un acte de discours est une action réalisée par un locuteur lorsqu'il énonce une phrase.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

La Pragmatique: Actes de langage (Speech Acts)

Assertifs : Ils expriment des propositions qui peuvent être vraies ou fausses. Exemple : "Le ciel est bleu"

Directifs : Ils ont pour but de faire agir l'interlocuteur.
"Donne-moi le sel".

Commissifs : Ils engagent le locuteur à accomplir une action future.
Exemple : "Je promets de venir demain".

Expressifs : Ils expriment l'état mental ou émotionnel du locuteur.
Exemple : "Je suis désolé".

Déclaratifs : Ils changent l'état du monde par l'énonciation. Exemple :
"Je vous déclare mari et femme".



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

La Pragmatique: Actes de langage (Speech Acts)

La détection des actes de langage dans le TLN permet aux machines de comprendre l'intention derrière un énoncé, essentielle dans des applications telles que les chatbots ou les assistants vocaux.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

La Pragmatique: **Implication Pragmatique et Sous-entendus**

Cela fait référence aux informations non explicites, qui sont sous-entendues dans un énoncé, mais qui sont comprises par les interlocuteurs en fonction du contexte.

- **Présuppositions** : Ce sont des informations supposées être connues par tous les participants de la conversation.

« ali a arrêté de fumer" présuppose que ali fumait auparavant.

- **Implicatures** : Ce sont des informations qui sont implicites, mais qui peuvent être inférées.

"Il pleut" implique que la personne doit prendre un parapluie, même si ce n'est pas dit explicitement.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

La Pragmatique: **Implication Pragmatique et Sous-entendus**

L'identification et l'extraction des présuppositions et des implicatures sont nécessaires pour comprendre des phrases avec significations implicites dans les systèmes de dialogue automatique ou la traduction machine.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

La Pragmatique: **Deixis**

- La **deixis** fait référence à des termes dont la signification dépend du contexte de l'énonciation
- en TLN : La résolution de la deixis est essentielle pour les systèmes TLN, notamment dans les dialogues, où le contexte de la conversation joue un rôle crucial dans la compréhension des termes déictiques.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

La Pragmatique: **Deixis**

Conversation :

Personne A : "Je suis fatigué aujourd'hui."

Personne B : "Toi aussi ?"

Analyse :

Le pronom "**Je**" dans la phrase de Personne A fait référence à Personne A elle-même, mais son sens dépend du locuteur. Sans savoir qui est le locuteur, le pronom "je" n'a pas de signification précise.

Le pronom "**Toi**" dans la réponse de Personne B se réfère à Personne A, mais encore une fois, sa signification dépend du contexte et de la personne à qui Personne B s'adresse.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Le Discours

- Est une chaîne d'énoncés interconnectés, qui peut prendre la forme d'un texte ou d'une conversation.
- L'étude du discours examine la manière dont les éléments linguistiques s'organisent et interagissent pour former des structures plus complexes, et comment les informations sont organisées et présentées dans le discours.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Le Discours : Structure

- La structure du discours s'intéresse à la manière dont les idées sont organisées et reliées entre elles à travers les phrases et les paragraphes.
- Elle se compose de deux éléments principaux :

Cohérence : Le lien logique entre les différentes idées du discours.

“Je vais au magasin. Je dois acheter du pain”

Cohésion : Les liens grammaticaux qui assurent la fluidité du discours, comme les conjonctions, les pronoms, et les adverbes.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Le Discours : Référence et Co-référence

La référence est la relation entre un élément linguistique (par exemple un nom ou un pronom) et un objet, une personne, ou une idée dans le monde ou dans le discours.

La co-référence désigne le phénomène où plusieurs expressions se réfèrent à la même entité.

Exemple : Dans le texte "Marie est venue. Elle a acheté un livre", "Marie" et "Elle" se réfèrent à la même personne.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Le Discours : Analyse

Analyse syntaxique :
Identification des structures grammaticales qui organisent le discours

Analyse sémantique :
Compréhension du sens des énoncés et de leurs relations dans le discours

Pragmatique :
Identification des actes de discours et des implications pragmatiques



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Différence entre Pragmatique et Discours

La **pragmatique** se concentre davantage sur l'interaction entre les individus et les contextes d'énonciation, cherchant à comprendre comment le sens est construit au-delà du sens littéral des mots.

Le **discours**, quant à lui, est plus centré sur l'organisation des énoncés dans une interaction ou un texte global, ainsi que sur les stratégies et les techniques utilisées pour structurer et produire le discours.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Applications en TLN

La pragmatique et le discours jouent un rôle clé dans de nombreuses applications de TLN. Parmi celles-ci, on trouve :

Chatbots et assistants vocaux : La gestion des actes de discours et des implicatures est essentielle pour améliorer la qualité des interactions.

Traduction automatique : La compréhension des sous-entendus et des présuppositions permet de produire des traductions plus naturelles.

Résumé automatique : Maintien de la cohérence et de la cohésion dans les résumés générés.

Analyse des sentiments : La reconnaissance des actes de discours et des nuances pragmatiques permet de mieux comprendre les intentions derrière les avis exprimés.



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Applications en TLN

Traitement et Analyse de Texte

Détection du plagiat

Résumé automatique
de textes

**Correction
grammaticale et
orthographique**

Classification et
catégorisation de
documents



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Applications en TLN

Recherche et Extraction d'Information

Moteurs de recherche
intelligents

Extraction d'entités
nommées

Systèmes de questions-
réponses



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Applications en TLN

Interaction Homme-Machine

Traduction neuronale et statistique

Systèmes de dialogue et agents conversationnels

Résumé et reformulation de texte

Assistants virtuels et chatbots



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Applications en TLN

Analyse des Sentiments et Opinion Mining

Analyse des avis clients

Surveillance des réseaux sociaux



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Applications en TLN

Analyse des Sentiments et Opinion Mining

Analyse des avis clients

Surveillance des réseaux sociaux



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Applications en TLN

Reconnaissance et Synthèse Vocale

Reconnaissance
automatique de la
parole

Text-to-Speech



2. Niveaux de traitement d'un langage

2.4 Pragmatique et Discours

Applications en TLN

Médecine et Biologie

Extraction
d'informations
médicales à partir de
textes

Diagnostic assisté par
IA à partir de dossiers
médicaux



Conclusion

le TLN est un domaine complexe et interdisciplinaire qui combine des connaissances en linguistique, en informatique et en intelligence artificielle. Les progrès réalisés dans ce domaine ouvrent la voie à des technologies de plus en plus avancées, capables d'interagir avec le langage humain de manière naturelle et efficace.

