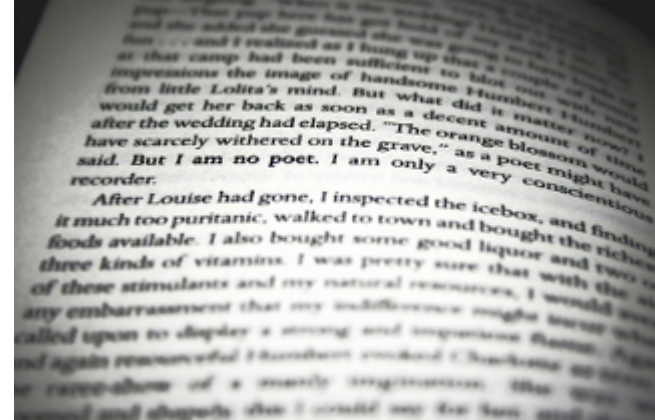


Traitement Automatique des Langues

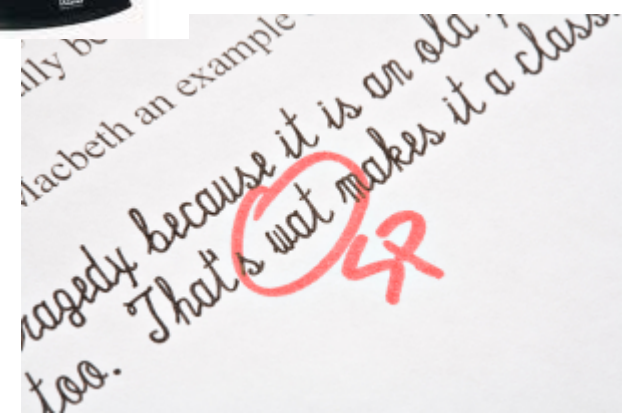
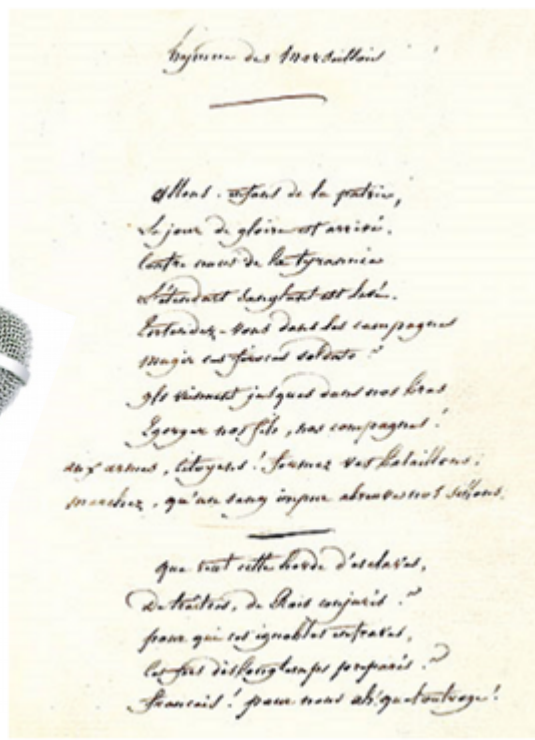
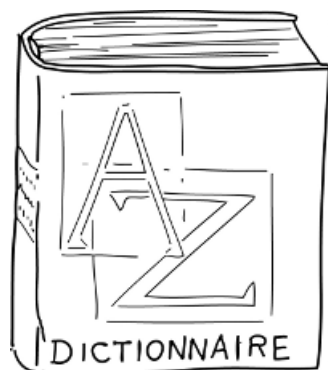
Indexation et Recherche
d'Information

Anne-Laure Ligozat / Xavier Tannier



Qu'est-ce que le TAL ?

(brainstorming)

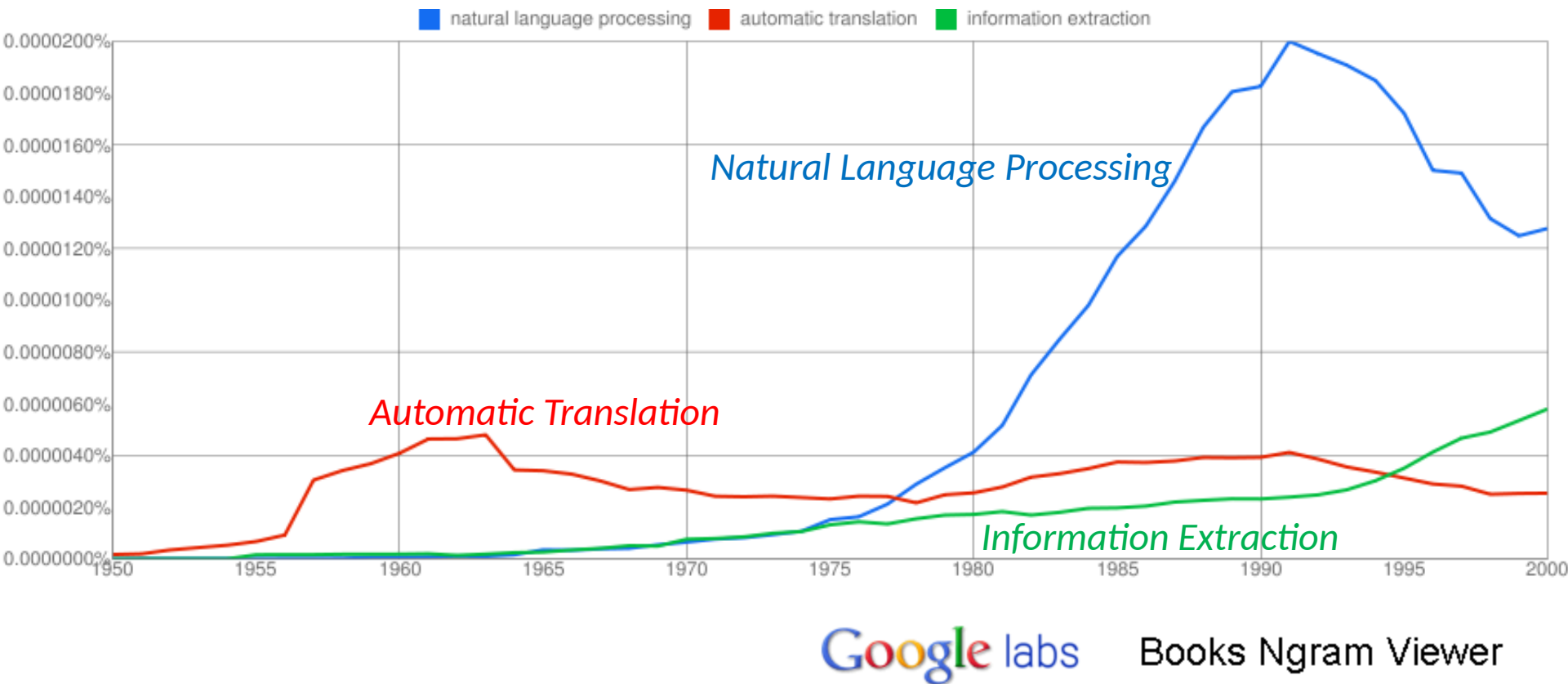


Qu'est-ce que le TAL ?



- **Langage naturel :**
 - Non formel
 - Ambigu
 - Implicite
 - Redondant
- Le **TAL** : une meilleure "compréhension" de la langue naturelle par la machine
- Discipline à la frontière de la linguistique et de l'informatique
 - Née à peu près en même temps que l'informatique (années 1950)
 - Une ambition initiale : la traduction automatique
 - *The spirit is willing but the flesh is weak* (l'esprit est fort mais la chair est faible)
 - *The vodka is strong but the meat is rotten* (la vodka est forte mais la viande est pourrie)

Historique



À quoi sert le TAL ?

- La traduction automatique
- La correction orthographique
- La recherche d'information et la fouille de textes
- Le résumé automatique
- La génération automatique de textes
- La synthèse de la parole
- La reconnaissance vocale

À quoi sert le TAL ?

- La veille technologique (extraction d'information...)
- L'aide aux handicapés (claviers autocorrecteurs, synthèse de la parole, ...)
- L'aide à la rédaction
- La détection de registre
- La reconnaissance de l'écriture manuscrite
- Les agents conversationnels
- ...



Les acteurs du domaine

Des gros éditeurs :

- IBM, Microsoft, Xerox, Apple, Toshiba, Sony, Google, Yahoo, Orange, etc.
- **Des intégrateurs / utilisateurs :**
- Ford, Symantec, EADS, Thalès/Arisem, BBN, SRI, EC, etc
- **Des PME françaises :**
- Exalead, Temis, ACapella, Lingway, Sinequa, Synapse, Systran, Reverso/Softissimo, Vecsys, Pertimm, Mondeca, etc.
- **Des labos de recherche :**
- John Hopkins, Stanford, Berkeley, MIT, U. Maryland, Columbia, NYU, Cambridge, Edimbourg, AixlaChapelle, Stuttgart, Paris Diderot/INRIA, Paris Sud/LIMSI etc.

Les différents niveaux de la langue

1. La **phonétique** et la **phonologie**

Comment les mots et les phrases sont liés aux sons qui les réalisent à l'oral

2. La **morphologie**

Comment les mots sont construits et quels sont leurs rôles dans la phrase

3. La **syntaxe**

Comment les mots se combinent pour former des syntagmes, puis des propositions et enfin des phrases correctes

Les différents niveaux de la langue

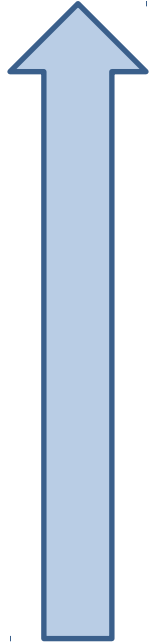
4. La **sémantique**

Comment les mots font du sens lorsqu'ils sont insérés dans une phrase
(indépendamment du contexte)

5. La **pragmatique**

6. Comment les phrases peuvent être interprétées selon leur contexte
d'énonciation (interlocuteurs, phrases précédentes,
connaissance commune du monde...)

Les différents niveaux de la langue



Le sujet de ce cours est le traitement de la langue.

Les grenouilles vertes ont des gros nez.

Les idées vertes ont des gros nez.

Vertes des ont les idées nez gros.



En pratique les frontières sont souvent plus poreuses.

La segmentation (rappel)

Identification des unités élémentaires

(phonèmes, morphèmes, mots, etc.).

- Pour l'écrit, des mots et des phrases.
- Un problème très complexe dans certaines
- langues (chinois...)
- Dans les langues "européennes" :
 - Les délimiteurs de mots et de phrases peuvent être ambigus
 - etc. T.A.L. 21.3 www.sncf.com
 - l'illusion aujourd'hui jusqu'à
 - Jean-Louis donne-t-il 1914-1918 06-13-23-33-12
 - Les mots (noms propres en particulier) peuvent avoir des variantes :
 - Etats-Unis États-Unis
 - France Inter France-Inter



Formes d'un mot, famille d'un mot

- **Flexion**

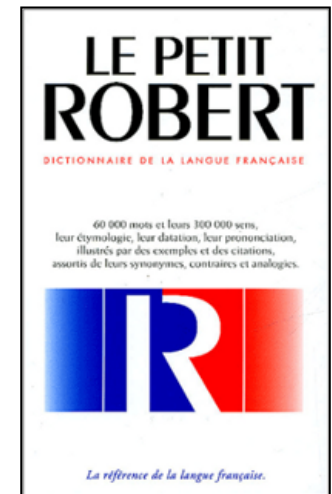
- Verbale : *montrer, montreras...*
- Nominale : *cheval, chevaux...*
- forme canonique (lemme) et formes fléchies

- **Dérivation**

- *penser/V + able = pensable*
- *in + pensable/A = impensable*
- base et dérivé

- **Composition**

- *appendice + ectomie = appendicectomie*
- éléments de formation, mot composé



La morphosyntaxe (rappel)

- Des analyses différentes pour des besoins différents :
 - **Segmentation** : l'étape indispensable.
 - **Lemmatisation** : pour rechercher/extraire de l'information, accéder au sens d'un lemme en faisant abstraction des flexions.
 - **Racinisation** (*stemming*) : pour agréger les dérivations morphologiques à peu de frais, sans souci de la perte du sens et des lemmes initiaux.
 - **Étiquetage** : pour appliquer des techniques de TAL sur les catégories grammaticales plutôt que sur les mots eux-mêmes.
 - Types de **flexions**, de **dérivations** : pour appliquer des traitements plus fins en vue d'une analyse syntaxique et/ou sémantique.
- Des techniques assez bien maîtrisées : un pourcentage d'erreurs faible mais difficilement compressible.

Le niveau lexical

- **But** : identifier les éléments lexicaux, leur structure et leurs caractéristiques ; regrouper les formes d'une même famille. Reconnaissance des **lemmes**, des **flexions**.
- **Moyen** : accès lexical direct, **analyse morphologique** (i.e. décomposition en morphèmes, à partir desquels les propriétés d'une forme sont calculées).
- **Outils** : un lexique, une description des morphèmes et des procédures de décomposition/recomposition associées.
- **Difficultés** : taille du lexique, vitesse d'accès et d'analyse, représentation du lexique, traitement des mots composés.
- **Résultat** : une représentation linéaire ou arborescente du mot, ses caractéristiques morphosyntaxiques, une représentation de sa signification, un représentant de sa famille.


(François Yvon)

Le niveau lexical (à quoi ça sert ?)

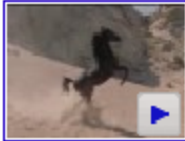
chevaux

Environ 7 760 000 résultats (0,12 secondes)

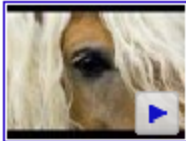
Vidéos correspondant à chevaux



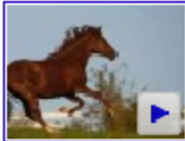
Chute cheval 🔍
2 min - 20 févr. 2007
Importé par cmoimd
[youtube.com](#)



les chevaux 🔍
4 min - 1 juil. 2008
[dailymotion.com](#)



SOS chevaux 🔍
4 min - 18 janv. 2008
Importé par mellemangas36
[youtube.com](#)



Le Cheval 🔍
6 min - 28 avr. 2007
[dailymotion.com](#)

Cheval - Wikipédia 🔍

Le **cheval** (*Equus ferus caballus*, ou *equus caballus*) est un grand mammifère herbivore et ongulé appartenant à l'une des sept espèces de la famille des ...
[fr.wikipedia.org/wiki/Cheval](#) - [En cache](#) - [Pages similaires](#)

Cheval à vendre cheval annonce International 🔍

Chevaux à vendre : acheter un **cheval** ou vendre un **cheval** ? Que ce soit un selle français, un Pur-Sang Arabe, un trotteur français, ici vous trouvez votre ...
[www.horse2buy.fr/](#) - [En cache](#) - [Pages similaires](#)

A cheval net - Les annonces du cheval - chevaux à vendre ... 🔍

Les Annonces du **cheval**, achat et vente **chevaux** et étalons, plus de 7000 annonces avec photos et vidéos de **chevaux**, selles, vans, matériel équestre, ...
[www.acheval.net/](#) - [En cache](#) - [Pages similaires](#)

Le niveau lexical (à quoi ça sert ?)

péché

Environ 29 900 000 résultats (0,09 secondes)

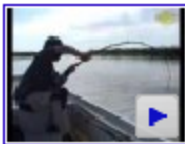
► [Péché - Wikipédia](#) 🔍

Dans les religions et certaines sectes, le **péché** est une transgression volontaire ou non de la loi divine. Il est souvent défini comme une désobéissance, ...

Étymologie - Péché, religions et athéisme - Péché et droit - Expressions

fr.wikipedia.org/wiki/Péché - En cache - Pages similaires

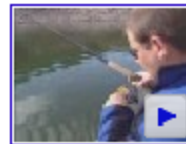
[Vidéos correspondant à péché](#)



[peche au silure](#) 🔍

2 min - 13 mars 2007

Importé par linda66750
youtube.com



[Techniques Modernes de la](#)

[Pêche du Sandre](#) 🔍

2 min - 3 avr. 2008

dailymotion.com



[Pêche à la carpe-explication-](#)

[HD-16.9](#) 🔍

10 min - 29 août 2009

Importé par eucalyptus51
youtube.com



[Les 11 commandements -](#)

[Pêche au Bazooka](#) 🔍

1 min - 29 oct. 2007

Importé par escobarlusitano
youtube.com

[PECHE-DIRECT.COM : VENTE ARTICLES ET MATERIEL DE PECHE, LE ...](#) 🔍

vente **peche** canne, vente **peche** moulinet, vente **peche** materiel, vente **peche** discount, vente **peche** barque, vente articles **peche**.

www.peche-direct.com/ - En cache - Pages similaires

[Fédération Nationale pour la Pêche en France](#) 🔍

Vous cherchez des informations sur la **pêche** de loisir en eau douce en France ? Pratique de la **pêche** dans chaque département, cartes de **pêche**, ...

www.federationpeche.fr/ - En cache

Le niveau lexical (à quoi ça sert ?)

restau végétarien

Environ 18 500 000 résultats (0,22 secondes)

▶ Restaurants végétariens Paris 🔍

Restaurants végétariens Paris - cuisine Végétarienne.

www.annuaire-parisien.com/restaurant-vegetarien-paris,40.html - En cache - Pages similaires

Restaurant Végétarien - Cityvox : Actualité, cinéma, restaurants ...

www.fra.cityvox.fr/restaurants_paris/LieuxTrouves?GLI... - Pages similaires

Restaurant végétarien, cuisine végétarienne : la sélection des ... 🔍

Restaurant végétarien, cuisine végétarienne : la sélection des meilleurs **restaurants végétariens**.

www.linternaute.com/Restaurant - En cache - Pages similaires

Adresses pour restau végétarien près de Paris



[godjo restaurant](#) traditionnel d'éthiopie 🍴 🔍 - ★★★★★ 19 avis - Page Google Adresses


www.godjo.com - 8 rue de l'école polytechnique, paris - 01 40 46 82 21



[Lemoni Café](#) 🔍 - 4 avis - Page Google Adresses

www.lemonicafe.fr - 5 Rue Hérold, Paris - 01 45 08 49 84

Le niveau lexical (à quoi ça sert ?)


 Web Images Vidéos Wikipédia

claud chabrol


Accueil > Résultats Web 1-10 de 582 047 pour claud chabrol, Page 1 - Page suivante

annonces Google

Votre publicité ici.
Utilisez Google AdSense pour augmenter vos revenus!
www.google.com



Le réalisateur Claude Chabrol est mort - LeMonde.fr
Claude Chabrol s'impose rapidement en tant qu'auteur, réalisateur.
Claude Chabrol lui confiera le rôle principal dans cinq autres films (1988), La
www.lemonde.fr/cinema/article/2010/09/12/le-realisateur-cla
12 Sep 2010 - En cache - Raccourci



Claude Chabrol - AlloCiné
- Connect with facebook Gagnez le jackpot ! Rechercher Ex.
69001 , Marseille ... Accueil Cinéma Séries TV Vidéos Comm
Stars
www.allocine.fr/personne/fichepersonne_gen_cpersone=97

Votre publicité ici.

ucteur de films [...]
e Affaire de femmes
st-mort_1410157_3476.htm

, Pathé , Gaumont ,
Dossiers | DVD | VOD |

fin de 1957 , Claude
ms, Le Beau Serge puis

9 - Garçon - ...
os, vidéos, blog
friends Membres ▼


Type de site :
» Blog
» Forum

Multimédia :
» Audio

Type de fichier :
» pdf
» word

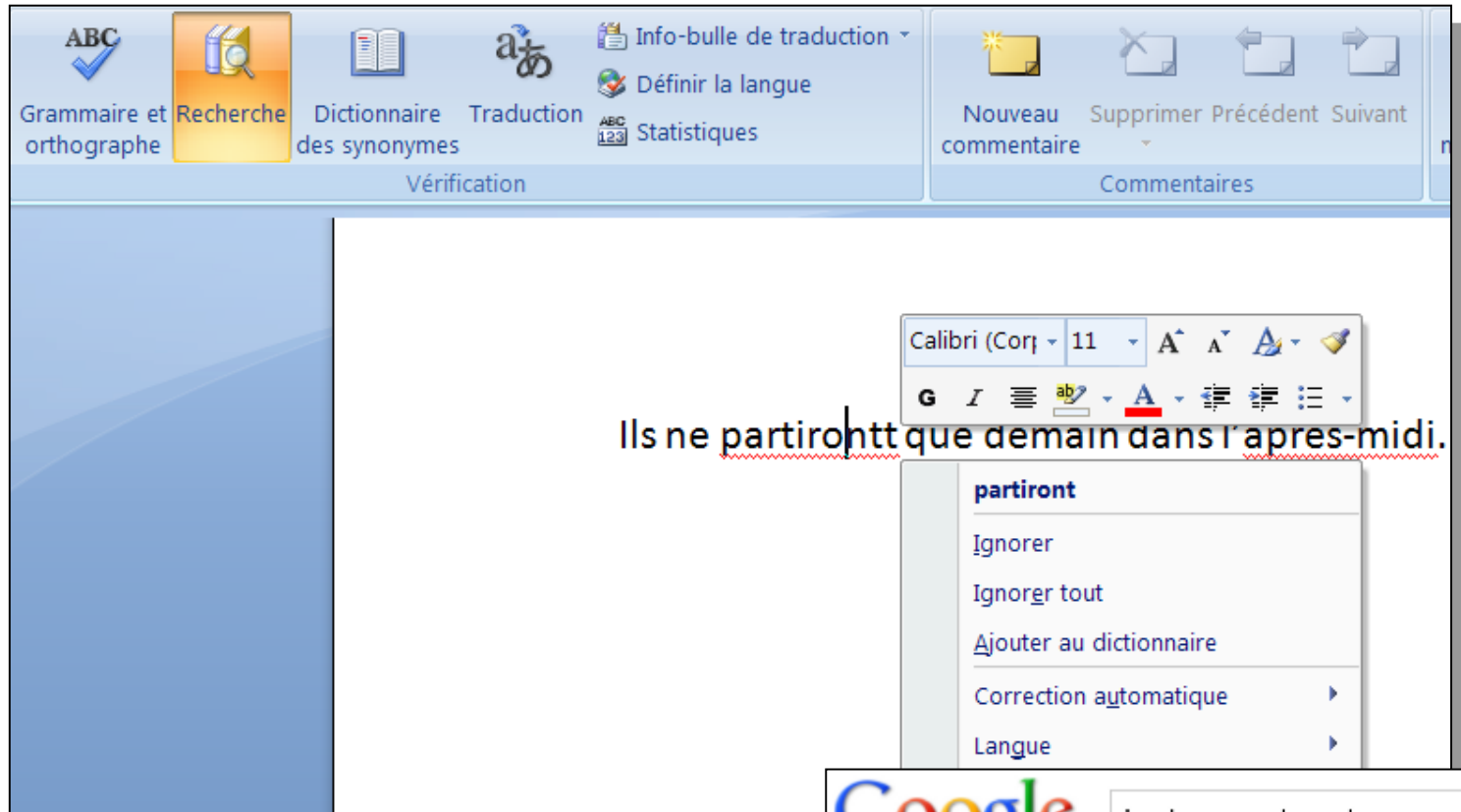
Termes associés :
» Cahiers du Cinéma
» Cinema francais
» Francois Truffaut
» Gerard Depardieu
» Isabelle Huppert
» Jean Poiret
» Jean Yanne
» Madame Bovary
» Stephane Audran

Langues :



Langue	Pourcentage
Français	81%
Anglais	18%
Autres	1%

Le niveau lexical (à quoi ça sert ?)



(techniques TRÈS
différentes !)



Le niveau lexical (à quoi ça sert ?)



[LA UNE](#)[LABO](#)[REBONDS](#)[libé](#)[BET](#)

[MONDE](#)[POLITIQUES](#)[SOCIÉTÉ](#)[ECONOMIE](#)

Éditions régionales [Bordeaux](#) - [Lille](#) - [Lyon](#) - [Marseille](#) - [Orléans](#) - [Rennes](#) - [Strasbourg](#)

Libération



Une politique familiale menacée

Société 22/12/2010 à 00h00

Les menaces sur la politique familiale, fuser (gauche) n'est plus là pour prendre son... dissoudre la politique familiale dans la politique sociale. La dépense p... l'objet de la verve populiste et haineuse de Jean-Luc Mélenchon est u... en cette veille de Noël. [...] En 2005, une employée d'un cabinet... pronostic de cancer du sein. Aujourd'hui, après cinq ans de combat ju...

TAGS

Américaine

Police

Série

Autocars

Lignes SNCF

Boutique

Chaussures

Sport

Grandes marques

Caricature

Sarkozy

Chaîne Info

CNN

Américaine

BBC

Chiffres

d'affaires

Magazine

Claire Chazal

Conseil

d'Etat

Retrait

Salaire

Cour

d'appel de Paris

Décision

Conseil

Prud'hommes

Culture

Soja

Brésil

Alimentation

École

Collège

Aubervilliers

Seine-Saint-Denis

Écrivains

Portraits

Auteurs

XXe siècle

Effet de serre

Carbone

Charbon

Pétrole

Gaz naturel

Enquête

Médiamétrie

Audience

Radio Skyrock

Espérance

de vie

Cadre

Ouvrier

Études

Étiquetage et chunking (la syntaxe du pauvre)

- **But** : désambiguïser les **étiquettes morphosyntaxiques** ambiguës ; identifier les frontières de groupes (**chunking**).
- **Moyen** : règles (patrons) de désambiguïisation ; modèles statistiques (Modèles de Markov cachés, Champs conditionnels aléatoires)
- apprentissage de règles de désambiguïisation.
- **Outils** : règles, patrons, corpus annotés manuellement (pour l'apprentissage).
- **Difficultés** : les mots inconnus ; combinaison de connaissances symboliques et de règles de décision numériques.
- **Résultat** : l'identification des étiquettes morphosyntaxiques (tagging) ; les frontières de groupe (chunking).

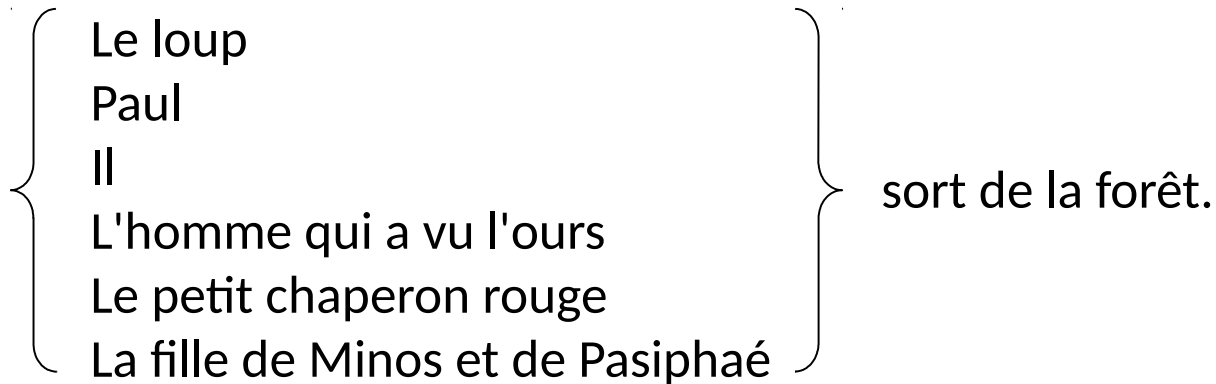
(François Yvon)

Du mot à la phrase : les syntagmes

- Les **syntagmes** (ou **constituants**) sont qualifiés par le type de l'élément principal (la **tête**). On a donc des syntagmes :
 - **nominaux** (le loup, Paul, l'homme qui valait 3 millions)
 - **verbaux** (vendra, valait 3 millions)
 - **adjectivaux** (rouge, [une classe] pleine d'étudiants)
 - **adverbiaux** (bien, conformément à la loi)
 - on parle aussi de syntagmes **prépositionnels** ([le chat] de ma mère)
- Les autres éléments sont :
 - les **spécifieurs** (déterminants...)
 - les **qualificateurs** (adjectifs, adverbes...)
 - les **compléments** (compléments du nom, propositions relatives...)
- Un syntagme a la même fonction que sa tête dans la phrase

Du mot à la phrase : les syntagmes

Les syntagmes de même type sont syntaxiquement **substituables** entre eux...



-
- Les syntagmes peuvent s'imbriquer les uns dans les autres :



Tests de constituency

Il lui avoue que la voiture de la sœur de Jean a été abîmée

- **Pronominalisation**

- *Il le lui avoue, Il lui avoue qu'elle a été abîmée, Il lui avoue que sa voiture a été abîmée, Il lui avoue que la voiture de sa sœur...*

- **Transformation** (construction clivée)

- *Il lui avoue que c'est sa voiture qui a été abîmée.*

- **Transformation** (interrogation)

- *Il lui avoue que quelle voiture a été abîmée ?*

Il lui avoue [que [[la voiture [de la sœur [de Jean]]] [a été abîmée]]].

Chunking

Les gendarmes interpellent un conducteur en état d'ivresse.

- Pas d'analyse de la structure interne
- **Pas de liens de dépendances** entre les **chunks**
- Les **ambiguïtés de rattachement** sont implicites

Bill vit l'homme sur la colline avec un télescope.

Bill vit l'homme sur la colline avec un télescope.

Le chunking (à quoi ça sert ?)

The screenshot shows the Exalead search engine interface. At the top is the Exalead logo and navigation links: Web, Images, Vidéos, Wikipédia, and Plus. A search bar contains the query 'traitement automatique des langues'. Below the search bar, associated research terms are listed: 'Traduction Automatique', 'Traitement Automatique', 'La Langue Naturelle', and 'Traitement De Texte'. The search results section shows 'Accueil > Résultats Web 1-10 de 218 942 pour traitement automatique des...'. Below this is a Google AdSense advertisement with the text 'Votre publicité ici. Utilisez Google AdSense pour augmenter vos revenus! www.google.com'. The first search result is for 'Atala', with a description: 'Rubriques ATALA L'association Adhésion Conseil d'administration Réunion d'Administration Assemblées générales Statuts Communiqués Revue de communication Notes de' and the website 'www.atala.org'. The second search result is for 'la poesie d'amour: Le traitement automatique des langues (tal)', with a description: 'la poesie d'amour recherche sur le web mardi 5 mai 2009 Le traitement automatique des (TAL) Le traitement automatique des langues (TAL) a pour objectif de traiter des données linguistiques'.

Exalead Web Images Vidéos Wikipédia Plus ►

traitement automatique des langues

Recherches associées : Traduction Automatique, Traitement Automatique, La Langue Naturelle, Traitement De Texte

Accueil > Résultats Web 1-10 de 218 942 pour traitement automatique des..., Page 1 - Page suivante

annonces Google

Votre publicité ici.
Utilisez Google AdSense pour augmenter vos revenus!
www.google.com

Atala
Rubriques ATALA L'association Adhésion Conseil d'administration Réunion d'Administration Assemblées générales Statuts Communiqués Revue de communication Notes de
www.atala.org
En cache - Raccourci

la poesie d'amour: Le traitement automatique des langues (tal)
la poesie d'amour recherche sur le web mardi 5 mai 2009 Le **traitement automatique des langues** (TAL) Le **traitement automatique des langues** (TAL) a pour objectif de traiter des données linguistiques

» [word](#)



Termes associés :

- » [Analyse Syntaxique](#)
- » [Archives Ouvertes](#)
- » [Base de donnée](#)
- » [Intelligence Artificielle](#)
- » [Langage Naturel](#)
- » [Linguistique Informatique](#)
- » [Universite de Paris](#)

Les entités nommées

- Les **entités nommées** sont des éléments qu'il est intéressant de pouvoir distinguer du reste du texte :
 - Entités : **personnes**, **organisations**, **lieux**
 - Dates : **dates**, **heures**
 - Quantités : **montants financiers**, **pourcentages**, **etc.**
- **Reconnaissance** des entités nommées :
 - **Identifier** ces unités dans un texte
 - Les **catégoriser**
 - Éventuellement, les **normaliser**

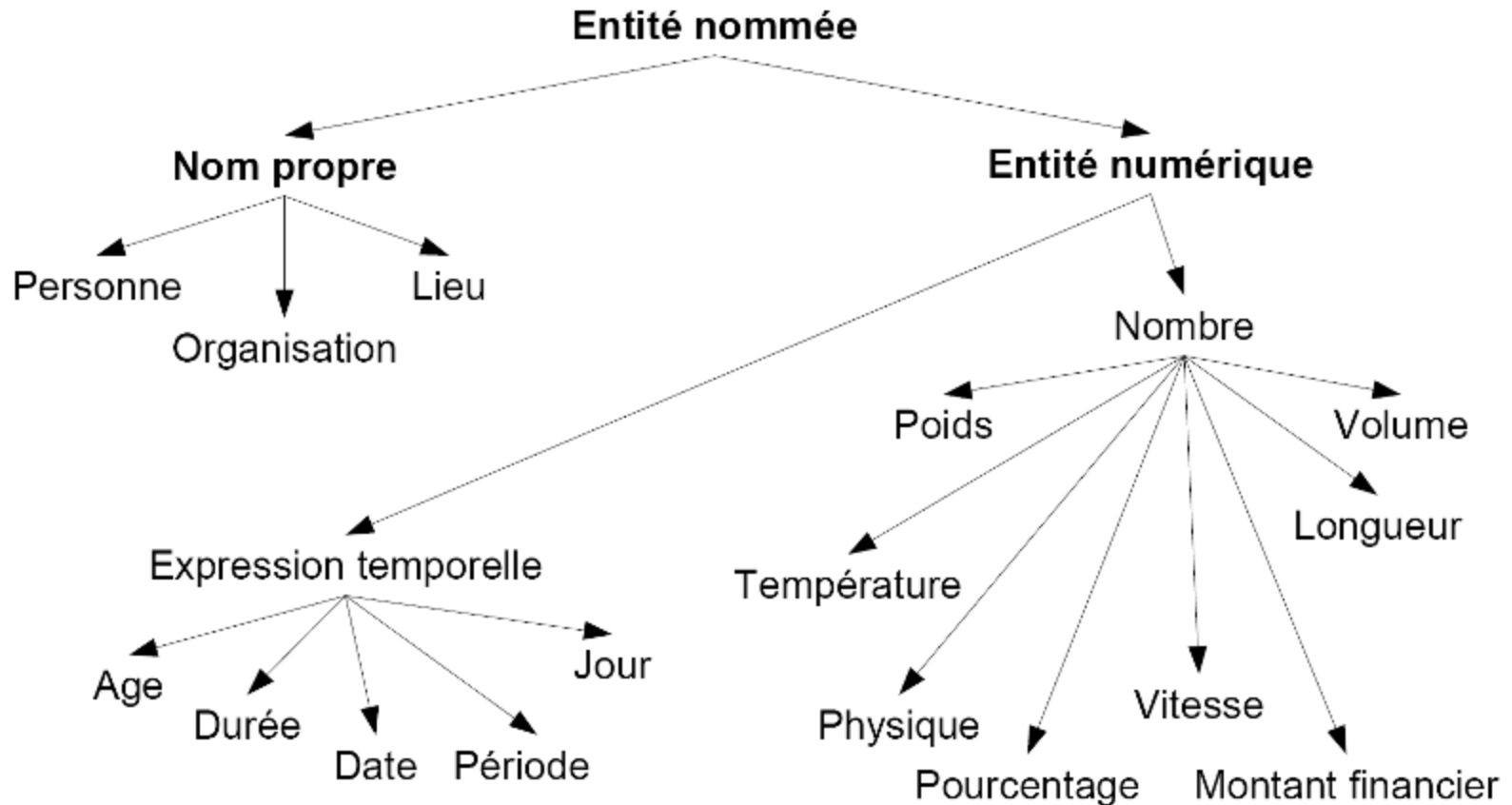
Les entités nommées

- L'ancien premier ministre socialiste Lionel Jospin a confirmé, jeudi 28 septembre, sur RTL, qu'il ne sera pas candidat à l'investiture socialiste pour la présidentielle de 2007.
- **Identification** : Lionel Jospin, jeudi 28 septembre, RTL, 2007.
- **Catégorisation** : L'ancien premier ministre socialiste <PERS>Lionel Jospin</PERS> a confirmé, <DATE>jeudi 28 septembre</DATE>, sur <ORG>RTL</ORG>, qu'il ne sera pas candidat à l'investiture socialiste pour la présidentielle de <DATE>2007</DATE>.
- **Normalisation** : L. Jospin \Rightarrow Lionel Jospin

Les entités nommées

- Plus de finesse ?
- **<PERS><FONCTION>***L'ancien premier ministre socialiste***</FONCTION>** *Lionel Jospin***</PERS>** *a confirmé,* **<DATE val="20060928">***jeudi 28 septembre***</DATE>**, *sur* **<ORG type="radio">***RTL***</ORG>**, *qu'il ne sera pas candidat à l'investiture socialiste pour la présidentielle de* **<DATE val="2007">***2007***</DATE>**.
- Le niveau dépend des capacités du système mais aussi de l'application
- La reconnaissance d'entités nommées est issue de la tâche plus générale de l'**extraction d'information**

Entités nommées



Les entités nommées (à quoi ça sert ?)

Google translate

From: English ▼



To: French ▼

Translate


English to French translation

Jack London is a famous writer.
He was not born in London.

Jack London est un célèbre écrivain.
Il n'était pas né à Londres.

 Listen

Les entités nommées (à quoi ça sert ?)

 exalead®

Web Images Vidéos Wikipédia

claudio chabrol

Votre publicité ici.

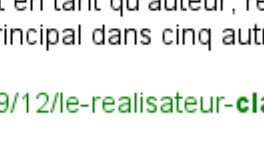
Accueil > Résultats Web 1-10 de 582 047 pour claudio chabrol, Page 1 - Page suivante

Annonces Google

Votre publicité ici.

Utilisez Google AdSense pour augmenter vos revenus!

www.google.com



Le réalisateur Claude Chabrol est mort - Le Monde

Claude Chabrol s'impose rapidement en tant qu'auteur, réalisant des films à la fois populaires et d'art. Claude Chabrol lui confiera le rôle principal dans cinq autres films (1988), La

www.lemonde.fr/cinema/article/2010/09/12/le-realisateur-claudio-chabrol_1410157_3476.html

12 Sep 2010 - En cache - Raccourci


Claude Chabrol - AlloCiné

- Connect with facebook Gagnez le jackpot ! Rechercher Exalead 69001 , Marseille ... Accueil Cinéma Séries TV Vidéos Commentaires


www.allocine.fr/personne/fichepersonne_gen_cpersonne=97...

9 - Garçon - ...

os, vidéos, blog

 **Type de site :**


- » [Blog](#)
- » [Forum](#)

 **Multimédia :**


- » [Audio](#)

 **Type de fichier :**

- » [pdf](#)
- » [word](#)

 **Termes associés :**

- » [Cahiers du Cinéma](#)
- » [Cinema francais](#)
- » [Francois Truffaut](#)
- » [Gerard Depardieu](#)
- » [Isabelle Huppert](#)
- » [Jean Poiret](#)
- » [Jean Yanne](#)
- » [Madame Bovary](#)
- » [Stephane Audran](#)

 **Langues :**



Langue	Pourcentage
Français	81%
Anglais	18%
Autres	1%

Les entités nommées (à quoi ça sert ?)

Cher Confrère,

Votre patiente madame [REDACTED], âgée de [REDACTED], née le [REDACTED], a été hospitalisée aux soins intensifs de cardiologie le [REDACTED] pour la prise en charge d'un syndrome douloureux thoracique.

Les principaux antécédents de cette patiente sont représentés par :

- une thyroïdectomie en [REDACTED].
- une hystérectomie en [REDACTED].
- un stripping des varices en [REDACTED].
- une hernie discale en [REDACTED].

Ses principaux facteurs de risque cardio-vasculaire sont représentés par une hypertension artérielle traitée par ATACAND, une dyslipidémie non traitée, un tabagisme estimé à 5 paquets année, et une légère surcharge pondérale ([REDACTED] pour [REDACTED]).

Les entités nommées (à quoi ça sert ?)

FIDJI

RESULTATS

☐ **Question N° 5** : Quel animal est sur un arbre ?

1. **Réponse N° 5_1** : *corbeau*

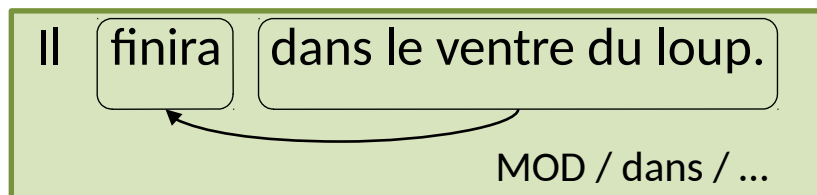
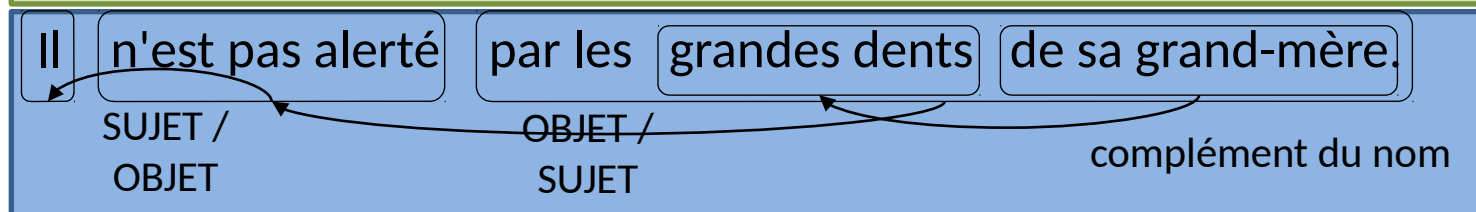
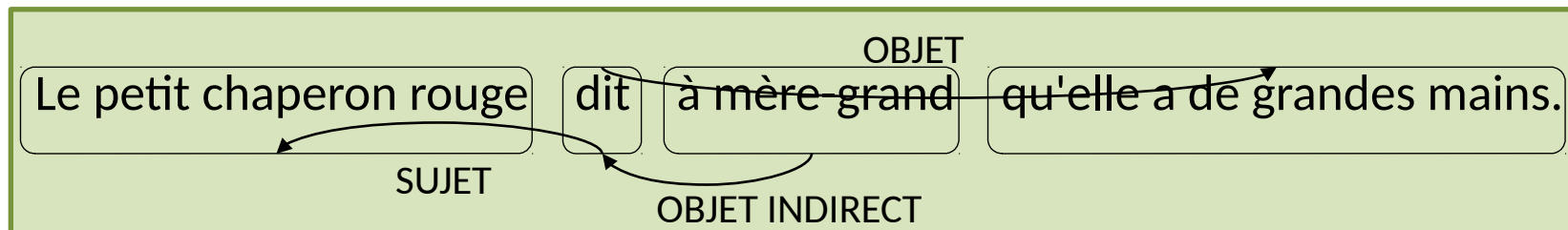
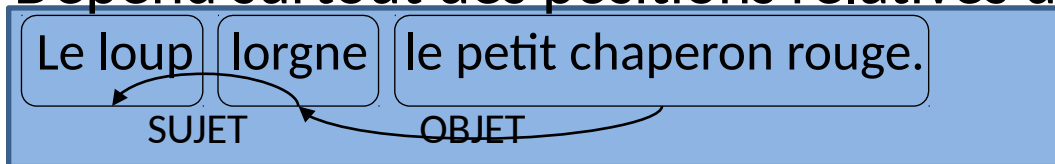
Passage : *Maître corbeau , sur un arbre perché , Tenait en son bec un fromage .*

La phrase

- Une ou plusieurs **propositions** :
 - Deux pigeons s'aimaient d'amour tendre.
 - Deux sûretés valent mieux qu'une, et le trop en cela ne fut jamais perdu.
(coordination)
 - Vous savez que nul n'est prophète en son pays.
(subordination)
- Une succession de phrases forment le **discours**.

Les fonctions grammaticales

- Le **rôle syntaxique** que joue un constituant par rapport à un autre
- Dépend surtout des positions relatives dans l'énoncé :



Le niveau syntaxique

- **But** : identifier les composants syntaxiques (**syntagmes**), leur fonction et les relations qu'ils entretiennent entre eux.
- **Moyen** : analyse syntaxique, qui fournit une représentation arborescente des composants de l'énoncé.
- **Outils** : un analyseur syntaxique, c'est-à-dire un formalisme de description des règles syntaxiques, des règles valides pour un (sous)-langage donné, et un système d'analyse (un parseur) capable d'exploiter ces règles.
- **Difficultés** : compromis entre richesse de description, vitesse d'analyse, et prolifération des ambiguïtés, complexité des phénomènes à décrire, robustesse aux entrées "bruitées" (coquilles, casse...).
- **Résultat** : un (ou des) arbres syntaxiques représentant la phrase, des relations entre les éléments.

(François Yvon)

Grammaire hors-contexte

GN \rightarrow DET ADJ? NOM ADJ?	(le petit? garçon roux?)
GP \rightarrow PREP GN	(de mon voisin)
GN \rightarrow GN GP	(la fille de mon voisin)
GV \rightarrow VER	(mange)
GV \rightarrow VER GN	(mange une pomme)
GV \rightarrow VER GP	(assistera à tous les cours)
GV \rightarrow VER GN GP	(donne un conseil à un étudiant)
GV \rightarrow VER GP GP	(se plaint de sa note à son prof)
P \rightarrow GN GV	(l'étudiant mange une pomme)

Grammaire hors-contexte

GN \rightarrow DET ADJ? NOM ADJ?

GP \rightarrow PREP GN

GN \rightarrow GN GP

GV \rightarrow VER

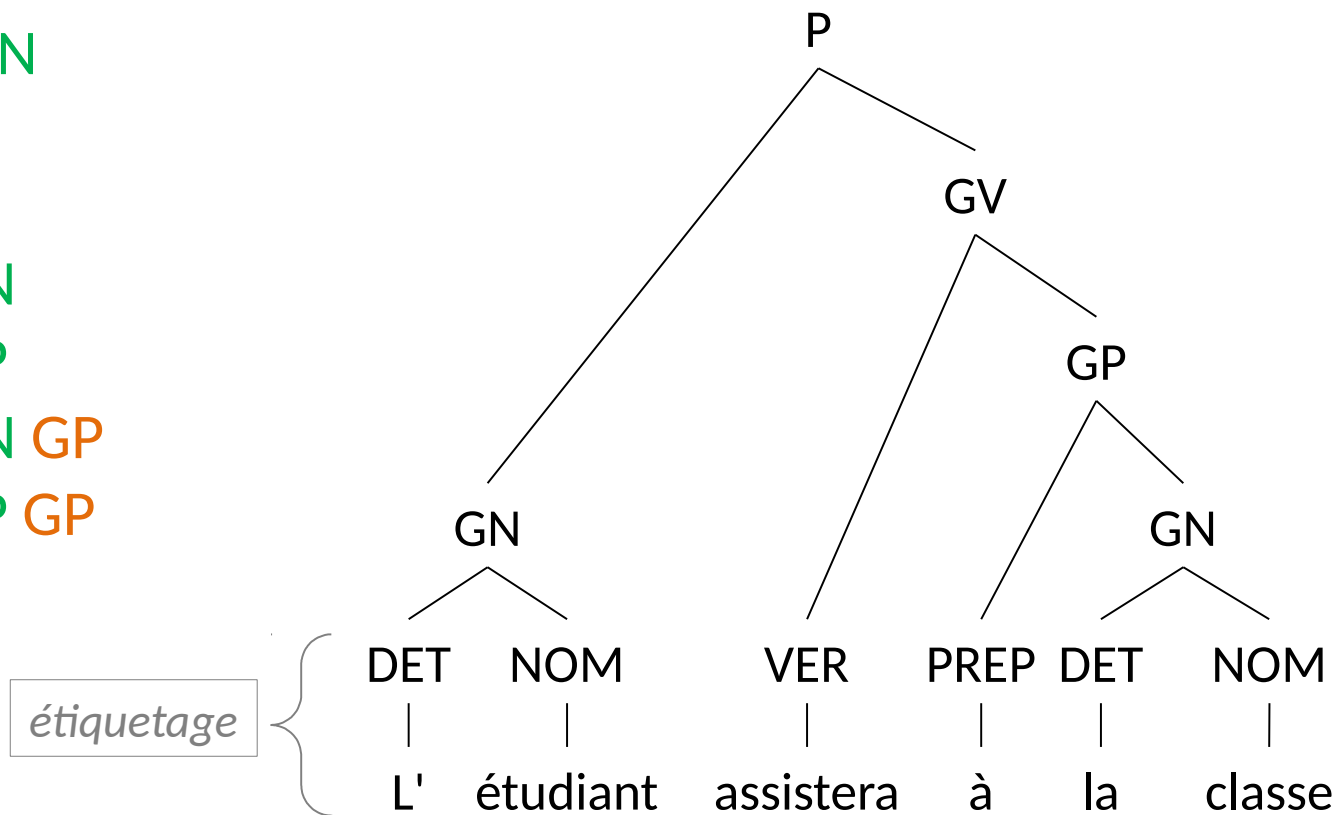
GV \rightarrow VER GN

GV \rightarrow VER GP

GV \rightarrow VER GN GP

GV \rightarrow VER GP GP

P \rightarrow GN GV



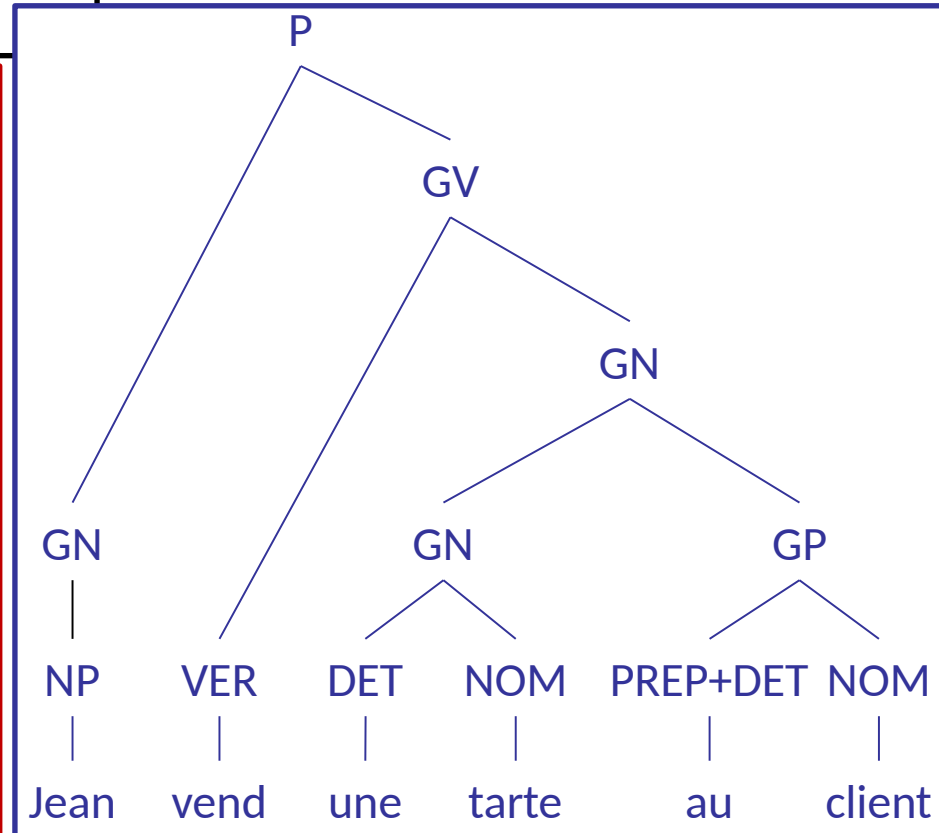
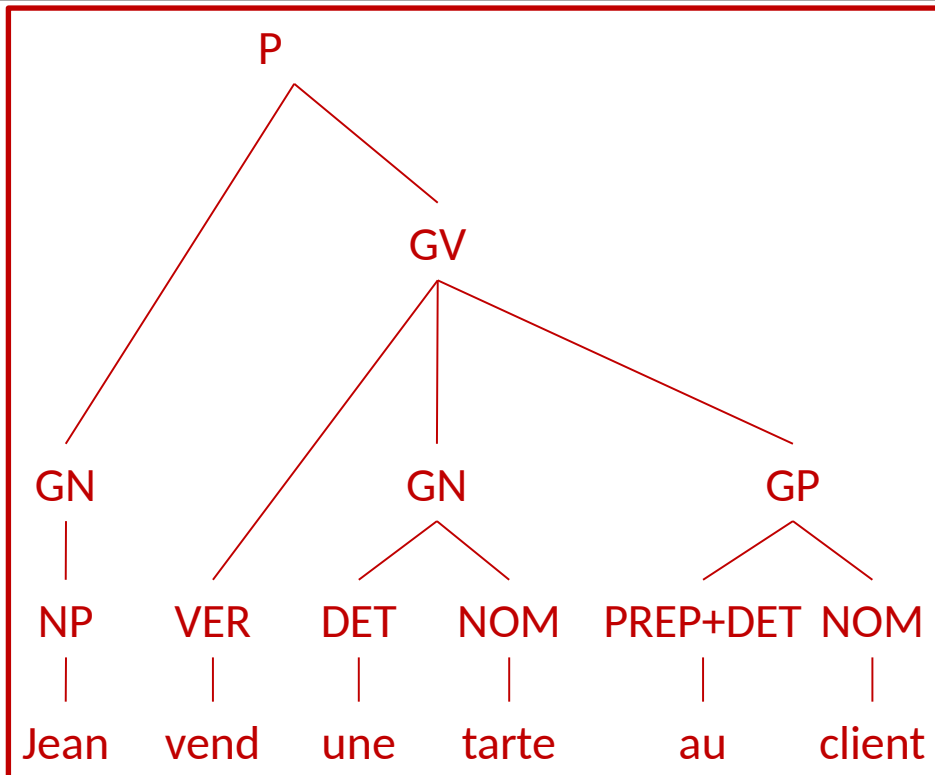
Les ambiguïtés syntaxiques

- On peut les classer selon les connaissances nécessaires pour la désambiguïsation
- Connaissances **pragmatiques**
 - *Jean a rapporté un vase de Chine.*
 - *Jean a rapporté un vase de Chine (des puces de St-Ouen).*
- Connaissances **sémantiques**
 - *Jean vend une tarte aux pommes.*
 - *Jean vend une tarte aux clients.*
- Connaissances **syntaxiques**
 - *Un jus d'oranges fraîches.*
 - *Un jus d'oranges frais.*
- Parfois des centaines de combinaisons possibles pour une phrase.

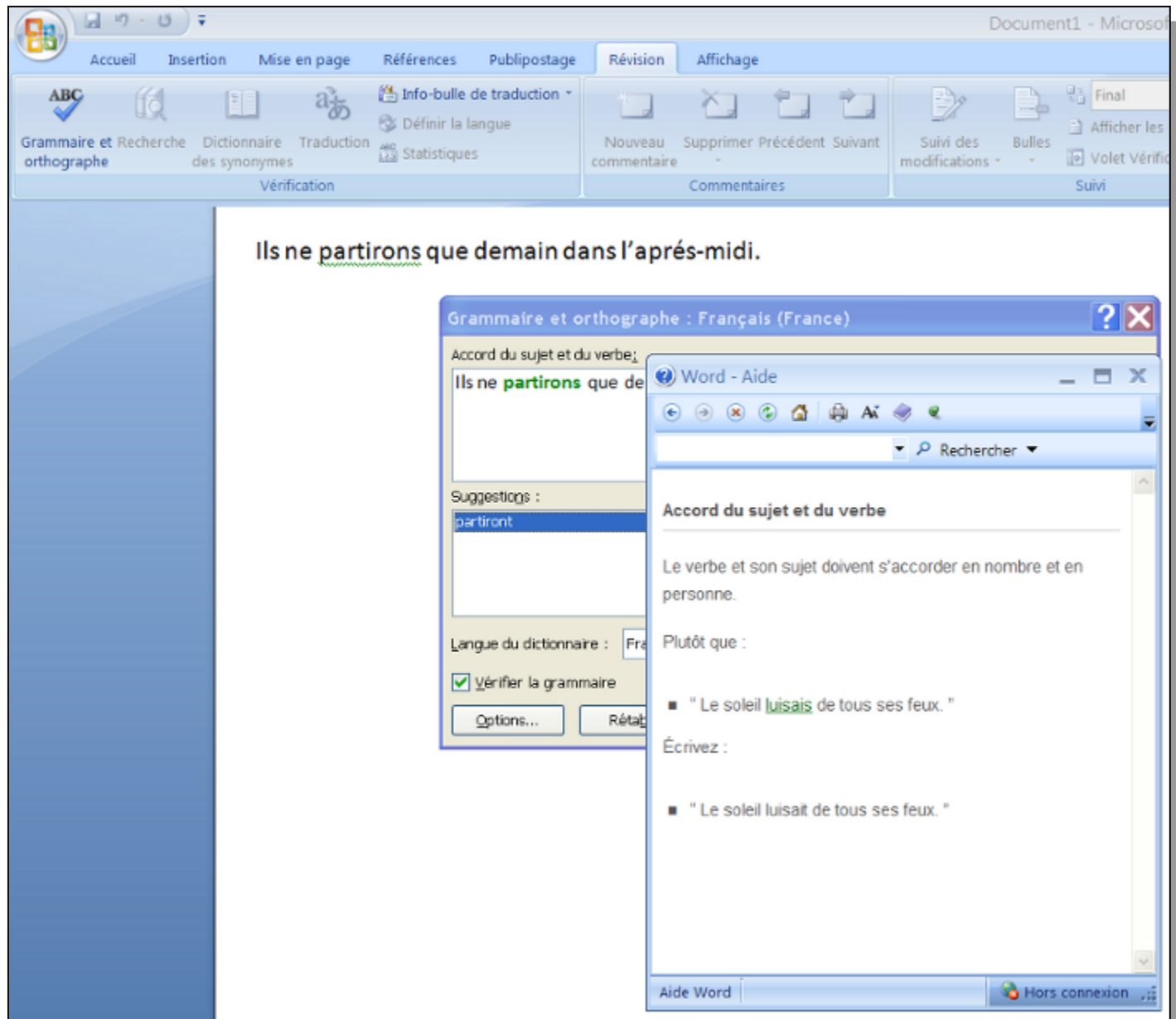
Les ambiguïtés syntaxiques



GN → DET ADJ? NOM ADJ?	GV → VER GN
GN → NP	GV → VER GP
GP → PREP+DET NOM	GV → VER GN GP
GN → GN GP	P → GN GV




Le niveau syntaxique (à quoi ça sert ?)



Le niveau syntaxique (à quoi ça sert ?)

The diagram illustrates the syntactic level of translation using Yahoo! Babel Fish. It shows two overlapping screenshots of the website. The top screenshot shows the English input: "I have nothing to offer but blood toil tears and sweat." The bottom screenshot shows the French translation: "Je n'ai rien à offrir mais les larmes et la sueur de travail dur de sang." A blue arrow points from the English input to the French output. The bottom screenshot also shows the French input being translated back to English, resulting in the same English sentence. The interface includes the Yahoo! Babel Fish logo, a search bar, and a "Translate again" button.

YAHOO! BABEL FISH


Translate a block of text  (Enter up to 150 words)

I have nothing to offer but blood toil tears and sweat.

YAHOO! BABEL FISH

En français

Je n'ai rien à offrir mais les larmes et la sueur de travail dur de sang.

Translate again  (Enter up to 150 words)

I have nothing to offer but blood toil tears and sweat.

Le niveau syntaxique (à quoi ça sert ?)

FIDJI

RESULTATS

☐ **Question N° 1** : Qui se trouve dépourvue quand la bise fut venue ?

1. **Réponse N° 1_1** : **cigale**

Passage : *La cigale , ayant chanté Tout l' été , Se trouva fort dépourvue Quand la bise fut venue .*

Le niveau sémantique

- **But** : obtenir une **représentation conceptuelle** de l'énoncé dans un langage formel, en l'articulant avec le monde « physique » de la scène.
- **Moyen** : calcul sémantique couplé à l'analyse syntaxique ou traduction de la représentation arborée dans un langage formel.
- **Outils** : une description sémantique au niveau lexical (relations de synonymie, méronymie, hyper/hyponymie, etc), des règles de composition, des outils de représentation du monde physique.
- **Difficultés** : explicitation partielle de l'implicite (problèmes de coréférence) ; ambiguïtés sémantiques (portée des quantifieurs) ; taille et précision de la connaissance nécessaire ; choix du formalisme de représentation (temporalité, croyances, etc).
- **Résultat** : un ensemble de représentations formelles des énoncés.

(François Yvon)

Les ambiguïtés sémantiques : l'anaphore




- L'anaphore pronominale

- Jacques₁ était furieux. **Il**₁ s'était disputé avec Georges.
- Dominique₁ rencontra Collins₂ à un congrès. **Ils**₁₊₂ se réconcilièrent.
- Nicolas₁ rencontra Dominique₂ dans un couloir. **Il**_? **lui**_? en voulait toujours.
- Pierre₁ empoisonna Sam₂. **Il**₂ mourut.
- Pierre₁ empoisonna Sam₂. **Il**₁ fut arrêté.
- Si votre bébé ne supporte pas le lait cru, faites-**le** bouillir.

- Autres

- La cage du gorille s'ouvrit. **Sa** serrure devait être mal fermée.
- Le gorille accéléra le pas vers le juge. **Le quadrumane** avait une idée derrière la tête.
- Je ne peux donner la suite de l'histoire. **Cela** serait pourtant délectable.

Les ambiguïtés sémantiques : l'ellipse

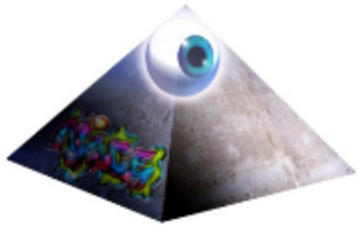
- Omission d'un élément dans une structure syntaxique
- Cet élément peut être n'importe de quel type
- Les Stéphanois portent des écharpes vertes et les Toulousains préfèrent les rouges et noires.

- Les Stéphanois aiment le football et les Toulousains le rugby.

- Les Stéphanois détestent les Parisiens et les Toulousains aussi.


Les ambiguïtés sémantiques : quantificateurs

Le problème : estimer la portée des quantificateurs utilisés

- *Un chien suit chaque homme qui passe la porte du bar.*
(combien de chiens ?)
- *Un problème d'environnement préoccupe tous les politiciens sérieux.*
(combien de problèmes ?)
- *Tous les hommes soulevèrent un piano.*
(combien de pianos ?)

Le niveau sémantique (à quoi ça sert ?)



A. L. I. C. E. Artificial Intelligence Foundation

Promoting the development and adoption of **ALICE** and **AIML Free Software**

A.L.I.C.E. Silver Edition

DAVE E.S.L. bot

C.L.A.U.D.I.O Personality Test

GET SITEPAL AVATARS

Get your own free chat bot: **Be Your Own Bot**

The ALICE A.I. contains material suitable for all ages.



Human: Is Alice your real name ?

ALICE: I don't thank anyone can be her own

Human: What do you want to talk about ?

ALICE: movies, philosophy, and your person

Human: what is your favorite movie ?

ALICE: My favorite movie is Teknolust. Hav

Human: nope, never heard about

ALICE: I would recommend you check it ou

You say:

Aide Fnac.com

> Contactez-nous

Une question ? Comment joindre notre


Avant de contacter notre service rela
avez-vous pensé à inter

Clara est notre conseillère en ligne à votre disposition 7 j
répondre à toutes vos questions concernant le suivi de vot

Par mail ou par téléphone

LIVRES, MUSIQUE, DVD, JOUETS, JEUX VIDÉO, LOGIC
TV, VIDÉO, HI-FI, CAMÉSCOPES :

Suivez votre commande
Bonjour, je suis là pour vous aider. Que puis-je faire pour vous ?



Écrivez ici pour discuter avec Clara

Envoyez

igod

repenting made easy

You are now chatting with God...

I exist as much as you exist, .

Me:

why can't we see you ?

God:

Perhaps it is invisible.

Me:

what ?

God:

Are you confused?

REPENT

Le niveau sémantique (à quoi ça sert ?)

FIDJI

RESULTATS

☐ **Question N° 2 : Qui alla crier famine ?**

1. **Réponse N° 2_1 : mouche**

Passage : *Pas un seul petit morceau De mouche ou de vermisseau Elle a la crier famine Chez la fourmi sa voisine , La priant de lui prêter Quelque grain pour subsister Jusqu'à la saison nouvelle Je vous paierai , lui dit -elle , foi d' animal , Intérêt et principal .*

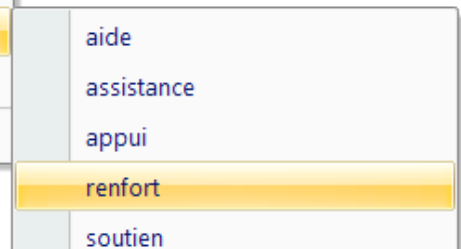
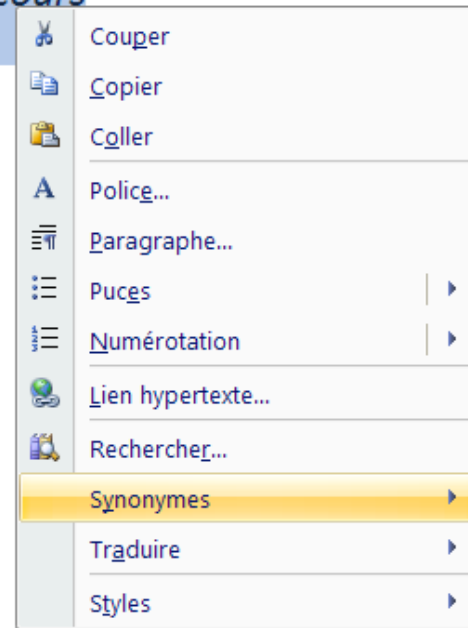
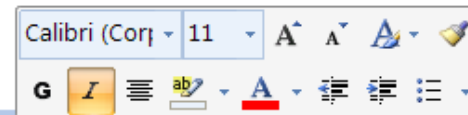
La sémantique lexicale

- Une difficulté majeure en sémantique, connaître :
 - le vrai **sens** des mots
 - les **relations** entre les mots
 -
- Les relations :
 - **Polysémie** (*bureau*) et **homonymie** (*son*)
 - **Synonymie**
 - **Hyponymie** (*tulipe / fleur*) (relation entre hyponyme et hyperonyme)
 - **Méronymie** (*serrure / cage, pièce / bâtiment*) (relation entre méronyme et holonyme)
 -
- Des bases existent (pour l'anglais surtout) qui répertorient les relations entre les mots (exemple, *WordNet*)

La sémantique lexicale (à quoi ça sert ?)


Nous partîmes cinq cents ; mais par un prompt secours

Nous nous vîmes trois mille en arrivant au port



La sémantique lexicale (à quoi ça sert ?)

YAHOO! BABEL FISH

Translate a block of text  (Enter up to 150 words)


The spirit is willing, but the flesh is weak.

English to French

YAHOO! BABEL FISH

En français


L'esprit est prompt, mais la chair est faible.

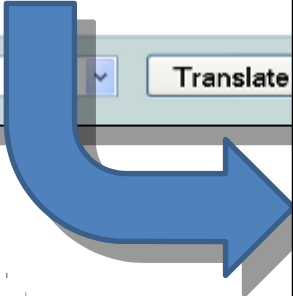
Translate again  (Enter up to 150 words)

The spirit is willing, but the flesh is weak.

SAISISSEZ UN TERME

DÉFINITIONS SYNONYMES CONJUGAISON

 **> spirit**
(nom)
esprit [masc.].
âme [fém.].
fantôme [masc.].
disposition [fém.].
caractère [masc.].
énergie [fém.].
entrain [masc.].
alcool [masc.].



Vers le niveau pragmatique

- **But** : achever la désambiguïsation de l'énoncé en y intégrant ce qui est implicite. Comprendre la fonction argumentative dans le contexte plus général de l'interaction (ou de la narration) : quelle information nouvelle apporte-t-il, au sujet de quoi dit-il quelque chose, ...
- **Moyen** : une théorie des activités humaines ; une théorie des interactions langagières (la pertinence, les conditions de félicité) ; une théorie des structures discursives...
- **Outils** : représentation des actions humaines, "grammaire" des interactions, logique.
- **Difficultés** : taille de la connaissance à représenter, spécification de la "grammaire" des interactions
- **Résultat** : une représentation formelle contextualisée de l'énoncé, des connaissances nouvelles... Mais applications rarissimes.

(François Yvon)

Pragmatique : quelques phénomènes

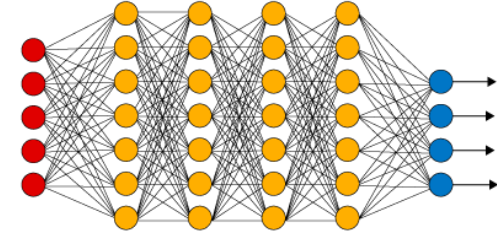
- La **déictique** : allusion directe au contexte de l'énonciation
 - *J'ai rencontré la reine d'Angleterre ici même. C'était l'année dernière.*
 - *Séparons-nous. Je pars avec toi, toi et toi, et vous partez dans ce sens.*
- L'**implicature conversationnelle** : ce qu'on peut déduire d'un énoncé en plus de son sens littéral
 - A : Le voisin est-il chez lui ?
– B : Sa voiture est devant le portail.
 - A : Je suis en panne d'essence.
– B : Il y a une station-service à deux pas d'ici.
 - A : François et Nicolas vont-ils s'arrêter là ?
– B : Sont-ils des hommes modérés ?

Pragmatique : quelques phénomènes

- La **présupposition** : ce qu'on peut déduire sur le contexte
 - *Le roi de France est sage*
⇒ Il existe un roi de France
 - *Jean regrette d'avoir fait ses études à Toulouse*
⇒ Jean est une personne identifiable par les deux interlocuteurs. Il a fait ses études à Toulouse.
 - *Si le vice-chancelier invite Simone de Beauvoir, il regrettera d'avoir une féministe à sa table. (Karttunen)*
⇒ Simone de Beauvoir est une féministe.
 - *Si le vice-chancelier invite le président des États-Unis, il regrettera d'avoir une féministe à sa table. (Karttunen)*
⇒ Il a également invité une féministe.

Le TAL actuellement

- Approches robustes
- grande quantité de textes
- approches statistiques



- Évaluation 
- Disponibilité d'outils et ressources
 - nombreuses langues mais anglais prévalent

images : <https://www.xenonstack.com>