

Traitement Automatique des Langues

LEXIQUE ET MORPHO-SYNTAXE

Université de Bretagne Sud — UFR LSH — Master DTIC



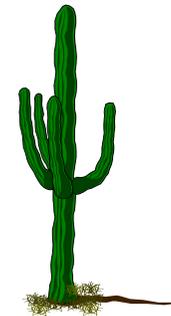
Langue et mots

- **Compositionnalité du langage naturel** : le sens d'une phrase dépend de celui des mots qui la composent
- **Mot** : unité lexicale porteuse de sens \Rightarrow lien entre la langue et l'univers du discours

pomme de terre ?



désespoir du singe ?



- **Français général**

70 000 mots simples

> 300 000 mots composés ou expressions figées

1 mot = 5 à 10 formes fléchies en moyenne

Lexique

Organisation des mots suivant leur forme, leur sens ou leur distribution

- **Propriétés paradigmatisques**

- organisation des mots pris de manière isolée (propriétés intrinsèques)
- **paradigme** : classe de mots suivant un point de vue donné

forme = morphologie

sens = sémantique lexicale

mort
mortel
mourir
mourrant
mortifier

mort
éteint
disparu
trépassé
crevé

- **Propriétés syntagmatiques**

- comportement et **distribution** des mots dans l'énoncé : (morpho)syntaxe

Brassens

sera

toujours

vivant

Les grands auteurs

sont

toujours

vivants

Céline Dion

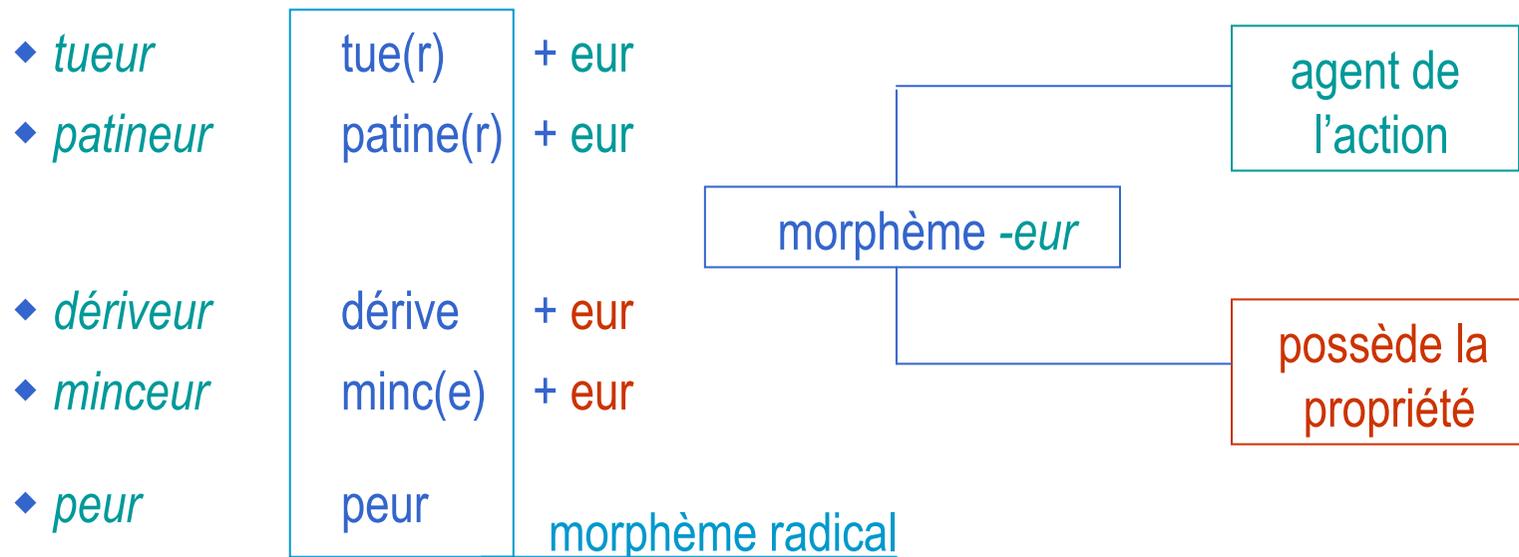
est

encore

vivante

Morphologie

Étude de la construction des formes lexicales à partir d'unités minimales porteuses de sens : **morphèmes**



Morphologie

- **radicaux** : catégorie ouverte de **morphèmes libres** pouvant se trouver seuls
- **affixes** (préfixes, suffixes) : catégorie fermée de **morphèmes liés**

Mot = (préfixe_deriv)* **radical** (radical_composition)* (suffixe_deriv)*
(suffixe_flexion)

**morphologie
inflectionnelle**

**morphologie
dérivationnelle**

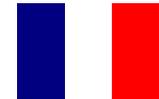
**morphologie
compositionnelle**

Morphologie inflectionnelle

Modification de la forme canonique d'un mot pour lui donner une **forme fléchie** exprimant ses propriétés morphologiques dans l'énoncé

– Pas de modification du sens en règle générale

– Français Genre, nombre, mode, temps, personne



étude ⇒ *études*

étudié ⇒ *étudiée, étudiés, étudiées*

étudier ⇒ *étudie, étudies, étudions, ..., étudiais, étudiait...*

– Anglais Morphologie très pauvre



genre marqué essentiellement par le substantif
marque de la 3ème personne quasi-uniquement
distinction substantif / verbe assez peu marquée

the beautiful houses

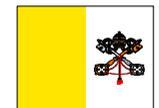
I, you, we, they ... leave

my love / to love

– Cas syntaxique

der, den, dem, des

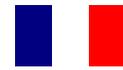
rosa, rosam, rosas, rosae, rosarum, rosis



Morphologie inflectionnelle (2)

Procédés

- En général, ajout d'un **suffixe** à un **radical**
- Exemple du pluriel
 - +s *étude + s ; car + s*
 - +où / ioù *bagad + où ; kêr + ioù*
 - +ed *paotr + ed ; loen + ed*
- Exceptions *cheval / chevaux mais chacal / chacals*
- Défectifs *chaloir, oindre*
trousers , information
dilhad



Cas particuliers langues agglutinatives



Productivité inflectionnelle

Français	70 000 lemmes simples = 550 000 formes fléchies	[Silberztein 1986]
Espagnol	35 000 lemmes simples = 400 000 formes fléchies	[Casajuana et al. 1985]

Application Limitation de la taille du lexique

Morphologie dérivationnelle

Création de **nouveaux mots** à partir d'un radical donné

- **Modification du sens** voire de la catégorie grammaticale
- **Procédé** (préfixe)* + radical + (suffixe)*
Des régularités mais pas de règle absolue
- **Exemples** *brassage, filtrage, remplissage, *évacuage, *remuage*
*masseur, livreur, déménageur, éleveur, *enseigneur, *surveilleur*
- **Dérivations successives** contraintes sur l'ordre des dérivations
nation ⇔ *national* ⇔ *nationaliser* ⇔ *nationalisation* ⇔ *dénationalisation*
mobile ⇔ *mobiliser* ⇔ *immobiliser* ← ? → *mobile* ⇔ *immobile* ⇔ *immobiliser*
- **Parasynthèse** dérivation préfixe + suffixe *lune* ⇔ *alunir*
- **Intérêt** analyse plus difficile que la flexion mais utile pour les mots inconnus

Morphologie compositionnelle

Création de **nouveaux mots** à partir de plusieurs radicaux

- Procédé radical + (radical)*
- Tête du composé porte le sens principal



- Exemples

- $N + N$ ⇒ *homme-grenouille, aller-retour, enseignant-chercheur, mail box*
- $N + Adj$ ⇒ *feu rouge, centre commercial cut-rate, freeway*
- $V + N$ ⇒ *coupe-faim, portemanteaux, porte-bagages, tire-fesses merry-go-round*
- $N de N$ ⇒ *pomme de terre, œil-de-bœuf, gibier de potence*
- $N prep N$ ⇒ *tire-au-flanc, passage à niveaux, boîte à lettres, cul de sac man-at-arms*

- Langues à forte compositionnalité

LLANFAIRPWLLGWYNGYLLGOGERYCHWYRNDROBWLLLLANTYSILIOGOGOGOCH



- Intérêt
 - N de N très fréquents en terminologie technique
 - analyse difficile mais utile pour les mots inconnus

Entités nommées

Lexèmes faisant référence à un élément précis de l'univers (domaine d'interprétation)

– Noms propres

Jacques Chirac

(**métonymie** : L'Élysée)

La Suisse

(**métonymie** : Berne)

Le Professeur Dubernard

– Noms communs ou mixtes

Le président de la République

Ce vieil habitué des arcanes de la politique hexagonale

La confédération Helvétique

Ce très discret pays du chocolat et comptes bancaires anonymes

L'ancien étudiant gauchiste devenu brillant leader régional de l'UMP

Le chirurgien aux grandes premières médicales

– Intérêt

Résolution des anaphores

Traduction

Morpho-syntaxe

- **Catégorie grammaticale**

- **Partie du discours** (POS : *Part-of-Speech*) : classes de mots aux mêmes propriétés distributionnelles (syntagmatiques)

- « **Traces** » sémantiques

nom ≈ élément de l'univers

verbe ≈ action, état

adjectif ≈ propriété

- **POS ouvertes** (mots **lexicaux** : description de l'univers)

- noms

- adjectifs

- verbes

- adverbes

↳ forte productivité lexicale

- **POS fermées** (mots **grammaticaux** : mots outils de la langue)

- conjonctions

- prépositions

- déterminants

- pronoms

↳ évolutivité très lente d'un point de vue diachronique

Morpho-syntaxe

- **Importance des catégories morpho-syntaxique**

- influence sur la prononciation [couvent](#)
- analyse syntaxique POS + morphologie (genre, nombre, personne)

- **Jeu d'étiquettes**

- Au delà d'une première catégorisation grossière, difficulté d'arriver à un consensus sur les POS à retenir.

Spécificités idiomatiques : [MULTEXT](#) www.lpl.univ-aix.fr/projects/multext

Exemples [Penn Treebank](#), [GRACE](#), [Amalgam](#),...

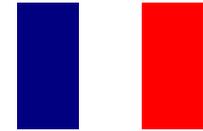
CORDIAL	correcteur grammatical grand public SYNAPSE Technologie
-------------------------	--------------------------------------------------------------------------------------------

POS Penn Treebank



CC	conj. coordination	PRP	pronom personnel
CD	nombre cardinal	RP\$	pronom possessif
DT	déterminant	RB	adverbe
EX	there existentiel	RBR	adverbe comparatif
FW	mot étranger	RBS	adverbe superlatif
IN	préposition + conj. de subordination	RP	particule (take up)
JJ	adjectif	TO	to
JJR	adjectif comparatif	UH	interjection
JJS	adjectif superlatif	VB	verbe infinitif
MD	verbe modal	VBD	verbe au passé
NN	nom singulier / massif	VBG	participe présent
NNS	nom pluriel	VBN	participe passé
NNP	nom propre sg.	VBP	verbe sg. prést sf 3° pers
NNS	nom propre pluriel	VBZ	verbe sg. présent 3° pers
PDT	prédéterminant	WDT	Wh-déterminant (<i>which</i>)
POS	's possessif	WDT	Wh-pronom (<i>who</i>)
		WP\$	Wh-pronom possessif
		WRB	Wh-adverbe (<i>when</i>)

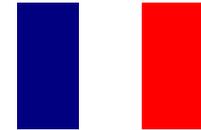
POS Action GRACE



NC	nom commun	PS	pronom possessif
NP	nom propre	DD	déterminant démonstratif
NK	nombre cardinal	DI	déterminant indéfini
VMN	verbe à l'infinitif	DT	déterminant interrog. ou exclamatif
VAN	auxiliaire à l'infinitif	DK	déterminant cardinal
VMPP	verbe participe présent	DR	déterminant relatif
VAPP	auxiliaire participe présent	DS	déterminant possessif
VMPS	verbe participe passé	DA	déterminant article
VAPS	auxiliaire participe passé	RGP	adverbe générique positif
VM <i>mt</i>	verbe au mode <i>m</i> et au temps <i>t</i>	RXP	adv. interrog. / exclamatif positif
VA <i>mt</i>	auxiliaire au mode <i>m</i> et au temps <i>t</i>	RGC	adverbe générique comparatif
AO	adjectif ordinal	RXC	adv. interrog. / exclamatif comparatif
AK	adjectif cardinal	RGN	adverbe générique négatif
AI	adjectif indéfini	RPN	adverbe particule négative (<i>ne</i>)
AS	adjectif possessif	RXN	adverbe interrog. / exclamatif négatif
AFP	adjectif qualificatif	SP	préposition
AFC	adjectif qualificatif comparatif	SD	préposition déictique
PD	pronom démonstratif	CC	conjonction coordination
PI	pronom indéfini	CS	conjonction subordination
PR	pronom relatif	I	interjection
PK	pronom cardinal	F	ponctuation
PX	pronom réflexif	X	autres
PT	pronom interrogatif		
PP	pronom personnel		

+ genre, nombre, personne : 312 POS

POS CORDIAL



ADJ	adjectif	PDS	pronom démonstratif
ADJI	adjectif invariable	PDP	pronom démonstratif
DETDEM	adjectif démonstratif	AINT	adjectif interrogatif
DETPOSS	adjectif possessif	AIND	adjectif indéfini
ADJNUM	adjectif cardinal	NCH	nom débutant par «h» aspiré
ADJORD	adjectif ordinal	PR	pronom relatif
ADV	adverbe	PRI	pronom relatif invariant
DETD	déterminant défini	PI	pronom indéfini
DETI	déterminant indéfini	PII	pronom indéfini invariant
AH	adjectif débutant par «h» aspiré	PP	pronom possessif
COO	conjonction de coordination	NP	nom propre
SUB	conjonction du subordination	NPI	nom propre invariant
INT	interjection	VINF	verbe infinitif
PREP	préposition	Vmt	verbe au mode <i>m</i> et temps <i>t</i>
NC	nom commun	VPARPR	verbe participe présent
NCI	nom commun invariant	VPARP	verbe participe passé
PPER	pronom personnel		

+ genre, nombre, personne = 201 POS

POS MULTEXT



Projet européen : jeu minimal d'étiquettes commun aux langues considérées

<i>Ntgn</i>	nom	+ type <i>t</i> (commun / propre) + genre <i>g</i> + nombre <i>n</i> + cas <i>c</i> (na = inapplicable)
<i>Vtmtpng</i>	verbe	+ type <i>t</i> (ppal / auxiliaire) + mode <i>m</i> + temps <i>t</i> + personne <i>p</i> + nombre <i>n</i> + genre <i>g</i> (na = inapplicable)
A	adjectif	
P	pronom	
D	déterminant	
R	adverbe	
S	préposition	
C	conjonction	
M	numérique	
I	interjection	
X	autres	

principales parties du discours
que l'on retrouve dans toute
théorie linguistique

Morpho-syntaxe

- **Sous-catégorisation**

- **Expression des capacités de rections des mots** : arguments dans la valence d'un verbe et tout autre type de relations dominant / dominé gouvernés par le mot.

- **Exemple** aimer(V ; SUJ = {NC | PP | NP} ; OBJ = {NC, PP, NP})

- **Application** analyse syntaxique lexicalisée

Exemple : structure prédicative dans les grammaires lexicales fonctionnelles (LFG)

PRED = *admirer* < (↑ **SUJ**) (↑ **OBJ**) >

Exemple : Traits de sous-catégorisation (SUBCAT) dans les grammaires HPSG

Exemple : arbre lexical traduisant les capacités d'adjonctions / substitutions du terme dans les grammaires d'arbres adjoints lexicalisées (LTAG)

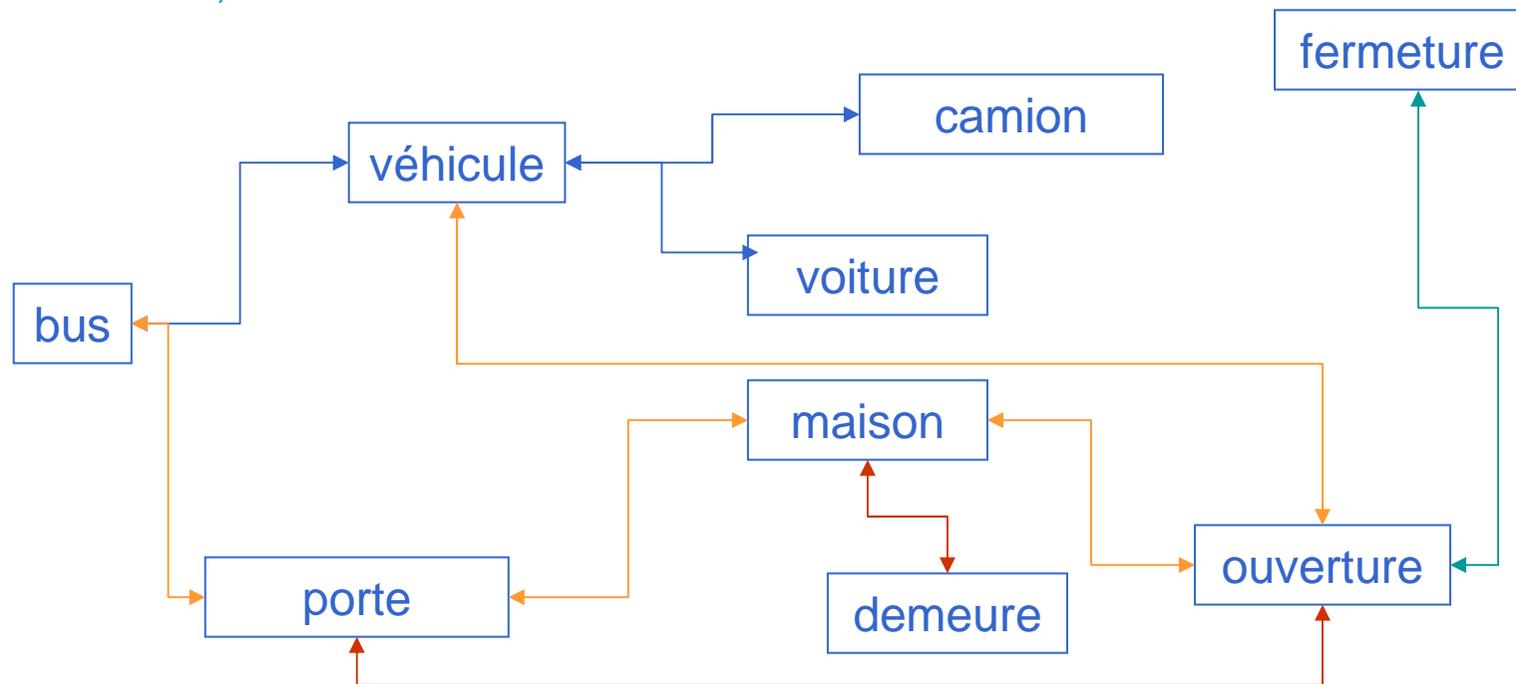
Sémantique lexicale

- **Ontologie**

- liens (hiérarchiques ou non) entre la significations des mots :

- **synonymie**, **antinomie**, **hyponymie (est_un)**, **partie_de**

- WordNet, EuroWordNet



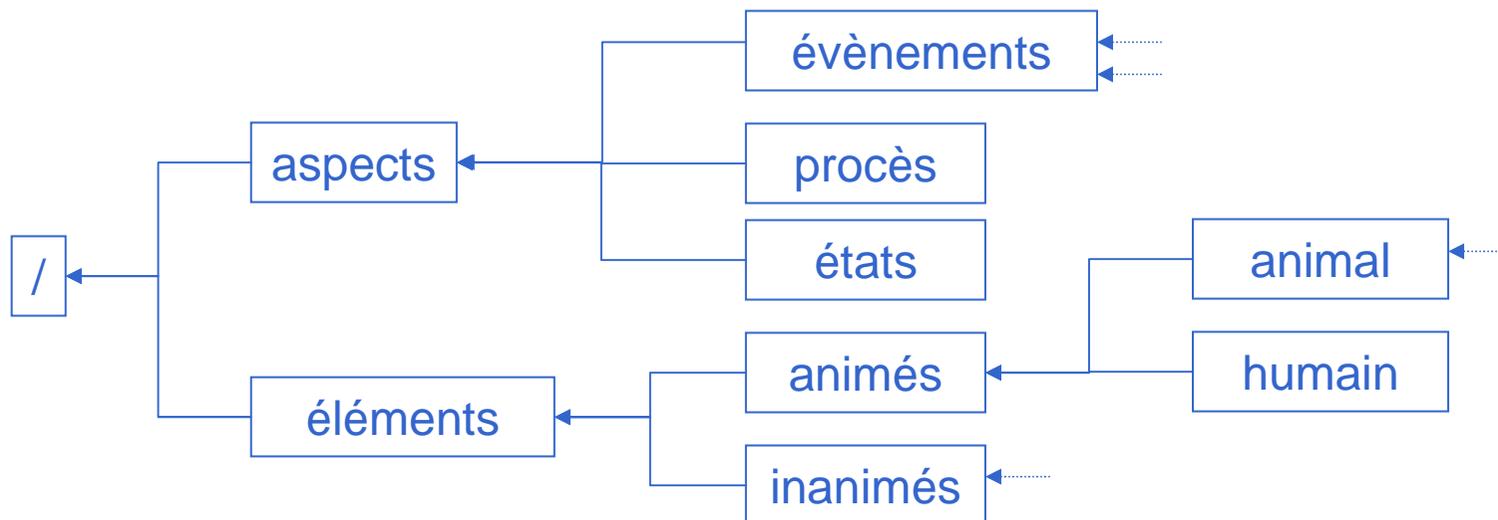
Sémantique lexicale

- **Traits Sémantiques**

- caractérisation de classes sémantiques à l'aide de traits différentiels (*sèmes*)

chat, matou, minou : /animé/ + /animal/ + /félin/ + /domestique/

- **taxonomie** hyponymes / hyperonymes



Propriétés syntagmatiques et sémantique

Lexique sémantique

- **Sens** du mot + relations prédicats / arguments au niveau sémantique

exemple

grammaires de cas sémantiques

grammaires catégorielles + interprétation Curry/Howard

- **Représentation**

logique (LP1)

graphes conceptuels

- **Exemple**

aimer(AGENT,PATIENT)

Bibliographie

- **Ouvrages d'entrée**

Allen J. (1995) *Natural Language Understanding*. Benjamin / Cummings Publ. Comp. Redwood City, CA. (chap. 1)

Huang X., Acero A., Hon H-W. (2001) *Spoken Language Processing : a guide to theory, algorithm and system development*. Prentice Hall, Upper Saddle River, NJ. (chap. 2)

G. Lessard (1996) *Introduction à la linguistique française*. Queen 's University at Kingston. Toile : <http://qsilver.queensu.ca/french/Cours/215>

- **Travaux cités**

Casajuana R., Rodriguez C., Sopena L., Villar C.. (1985) *A general purpose spanish dictionary : application to text composition*. *Internal Report IBM Madrid Scientific Center*, Madrid, Espagne.

Silberztein M. (1986) *The lexical analysis of French*. *Lecture Notes in Computer Science, LNCS 377*, pp. 93-110.