

Séminaire ASCO Club Cognitique
Jeudi 2 décembre 18h, amphi Pîtres
Université Bordeaux II Victor Segalen
Place de la Victoire, Bordeaux

PARLER ET COMPRENDRE:

**PROCESSUS COGNITIFS ET
MODÈLES INFORMATIQUES
DE LA FACULTÉ DE LANGAGE**

Christian RETORÉ

**(Université de Bordeaux,
INRIA Bordeaux Sud Ouest & LaBRI-CNRS)**

PAS DE DÉMOS:
LES FORMULES POUR LES ORDINATEURS,
ET NON POUR LES HUMAINS
QUI CONÇOIVENT LES PROGRAMMES
MANIPULANT LES DITES FORMULES

- Tout élève suit un cours de langue.

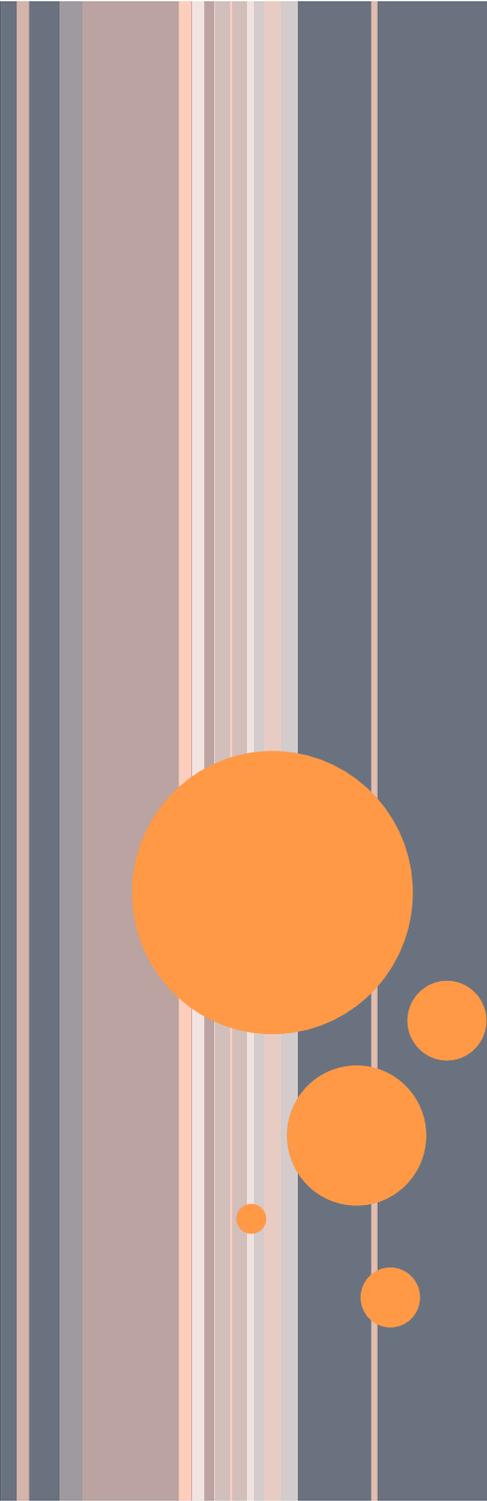
$$\forall e \exists c \exists l (\textit{etudiant}(e) \Rightarrow (\textit{suit}(e, c) \wedge \textit{sujet}(c, l) \wedge \textit{langue}(l)))$$

PLAN

- Historique
- Traitement Automatique des Langues (quelques applications)
- Les domaines de la linguistique
- Grammaire générative, principes,
 - hiérarchie,
 - acquisition
- Logique et sémantique
richesse des questions classiques
- Quelques aspects cognitifs

LINGUISTIQUE ET INFORMATIQUE: UNE LONGUE HISTOIRE

- 1949 Machine Translation
 - The flesh is weak but the spirit is willing
 - (russian)
 - The meat is rotten but the vodka is strong
 - La chair est faible mais l'esprit est fort.
- Nom du domaine:
 - Linguistique informatique
 - Traitement automatique des langues
 - Linguistique computationnelle



TRAITEMENT
AUTOMATIQUE DES
LANGUES

Le versant applicatif

QUELQUES APPLICATIONS

- Le Graal: la traduction automatique (il faut savoir tout traiter pour y parvenir)
- Aide à la traduction:
 - domaine spécifique
 - propose pour chaque mot ou expression des traductions
 - les assemble avec les choix du lecteur
 - (éviter au maximum la représentation des connaissances)

QUELQUES APPLICATIONS

- L'interface homme/machine en langue naturelle par exemple:
 - interrogation de base de données en langage naturel
 - *Quels sont les films des années cinquante qui passent actuellement à Bordeaux?*

QUELQUES APPLICATIONS

- Correcteurs orthographiques :
 - Synapse Word (souligné vert: français, italien,...)
 - Quels livres crois-tu qu'il sait que je pense que tu as lus ?
- Génération automatique de bulletins météo, de comptes-rendus,..
- Résumé automatique:
deux techniques contrastées

QUELQUES APPLICATIONS

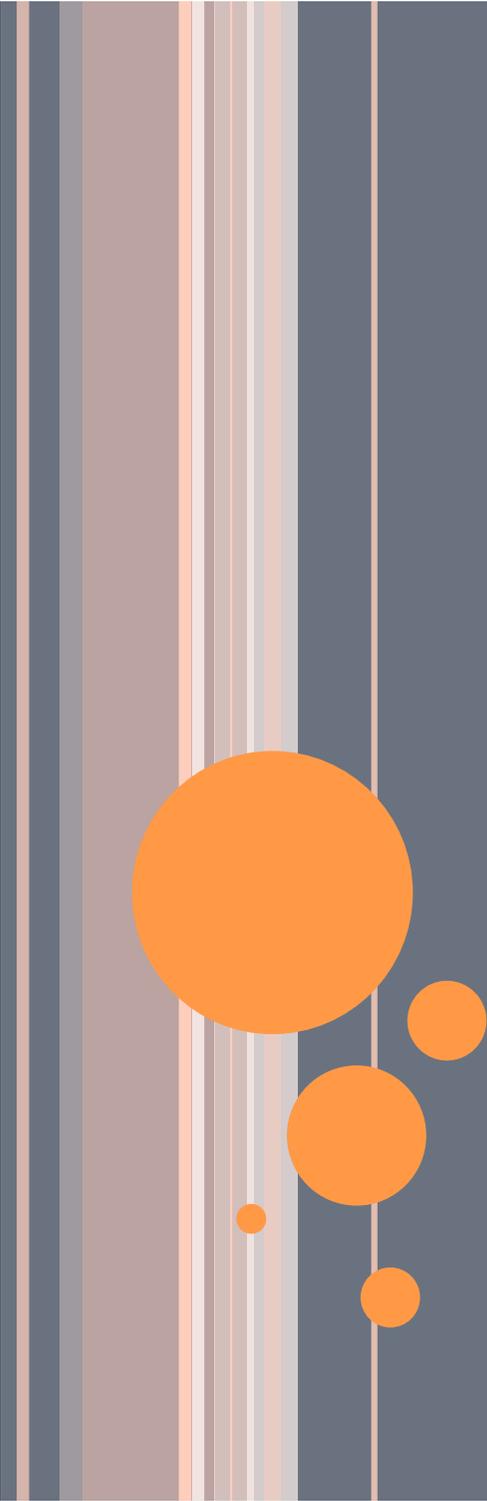
- Recherche d'information
(notamment sur Internet)
 - production laitière / production de lait
 - production minière / production de mine(s) ???
- Reconnaissance de la parole
(par ex. pour sous-titrage)
nécessite une analyse morpho-syntaxique
pour fonctionner en temps réel

UN DOMAINE INTERDISCIPLINAIRE

- Mathématiques
 - Logique
 - Théorie des langages
 - Probabilités
- Informatique
 - Algorithmique
 - Génie logiciel
- Linguistique
 - Descriptions linguistiques
 - Linguistique formelle
 - Philosophie du langage

DES OBJECTIFS VARIÉS

- Réalisation d'outils de traitement des langues
- Formalisation des théories linguistiques, vérification ou réfutation d'hypothèses
- Développement des théories informatiques et mathématiques pour elles-mêmes, éventuellement pour d'autres objectifs (par ex. grammaire du génome)



UN APERÇU DES DOMAINES DE LA LINGUISTIQUE

Diviser pour comprendre

ANALYSE EN GLOBALE: MISSION IMPOSSIBLE

- « Elle remplace Marie-Claire avec le supplément. »
- Questions:
 - Qui est « elle »?
 - S'agit-il de magazines ou de personnes?
 - Supplément: magazine, heures,...
 - « remplace avec le supplément d'heures» ou « Marie Claire avec supplément »....

« COMPRENDRE » ?

- Construire une représentation du sens utilisable par un ordinateur
- Sous tâches:
 - Trouver la structure de la phrase
 - Interpréter les unités de base
 - Assembler ces sens en suivant la structure de la phrase

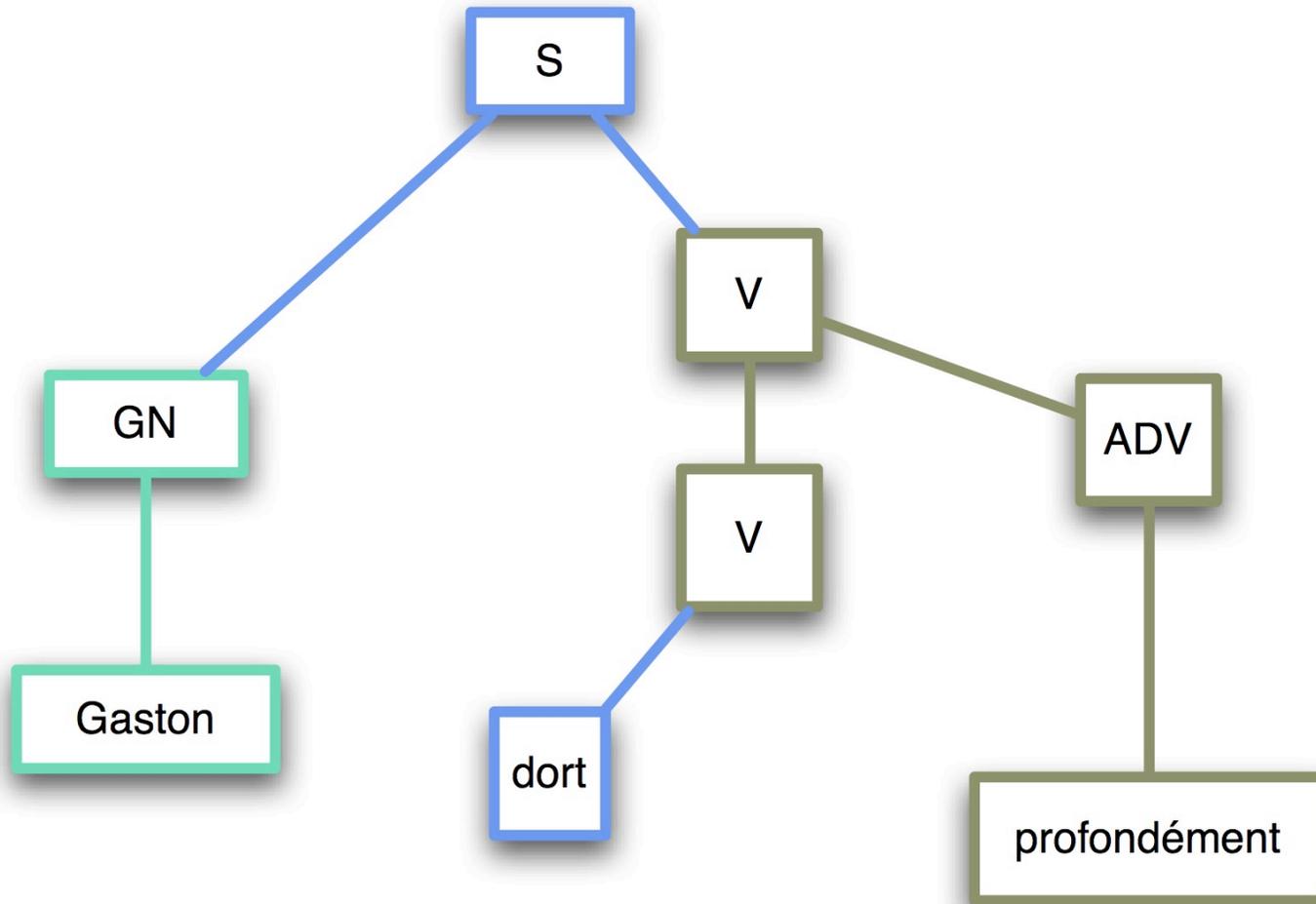
NIVEAUX D'ANALYSE (1/2)

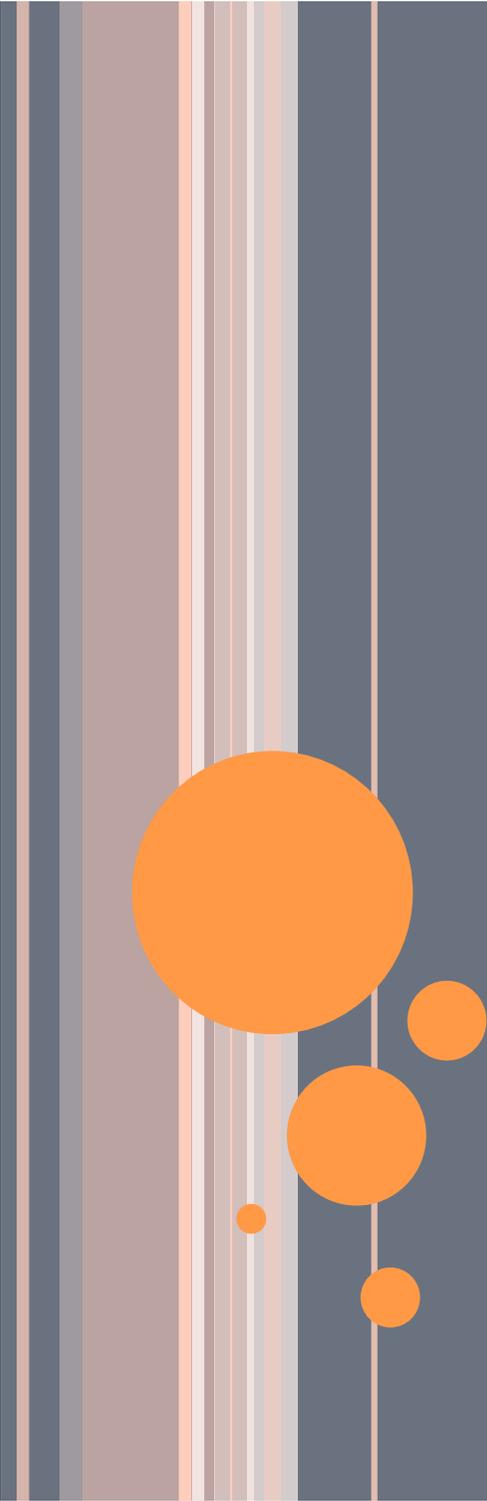
- Phonétique: son, production et réception
- Phonologie: sons comme système
Bali et Paris identiques pour un japonais
- Prosodie
peu formalisée, mais importante pour l'acquisition
- Morphologie structure des mots
 - Flexionnelle, sans changement de catégorie:
cheval -> chevaux
manger -> mangerons
maison-> maisonnette carpe->carpette??
 - Dérivationnelle:
plier-plier, laver-lavage,
porte-manteau, lave vaisselle,

NIVEAUX D'ANALYSE (2/2)

- Syntaxe: structure de la phrase (cf. ci-après)
- Sémantique (sens)
 - Des mots, système conceptuel (cf. ci-après)
 - De la phrase, du discours: logique, compositionnalité:
Le sens du tout est fonction du sens des parties et du mode de composition.
 - *Le chat de la voisine dort sur le toit d'une voiture rouge.*
 - Léger problème: *Si un paysan possède un âne, il le bâte.*
- Pragmatique, énonciation
 - Sens par l'usage, en situation: discours, dialogue et du contexte linguistique et extra linguistique:
 - *Allons plutôt dans ce restaurant.*
 - *Ce n'est pas grave.* Dit par une fillette de 6 ans: *grave???*

SYNTAXE





LA GRAMMAIRE GÉNÉRATIVE

Théorie incontournable du XX^e siècle

LA GRAMMAIRE GÉNÉRATIVE

- Théorie linguistique aussi utilisée en
 - Informatique (compilation, parallélisme)
 - Mathématiques (théorie des groupes)
 - Biologie (génomique)
- Origines
 - (Panini, Inde, V^e siècle avant J.C.)
 - Noam Chomsky 1955

RUPTURE AVEC LE COMPORTEMENTALISME

- Une langue N'EST PAS l'ensemble des énoncés produits par les locuteurs
- MAIS
- Un ensemble fini de règles inconscientes permettant de produire ces énoncés.
- Premier argument
Soit P la phrase la plus longue à ce jour,
« *il croit que P* »
est sans doute aussi une phrase.

RUPTURE AVEC LE COMPORTEMENTALISME

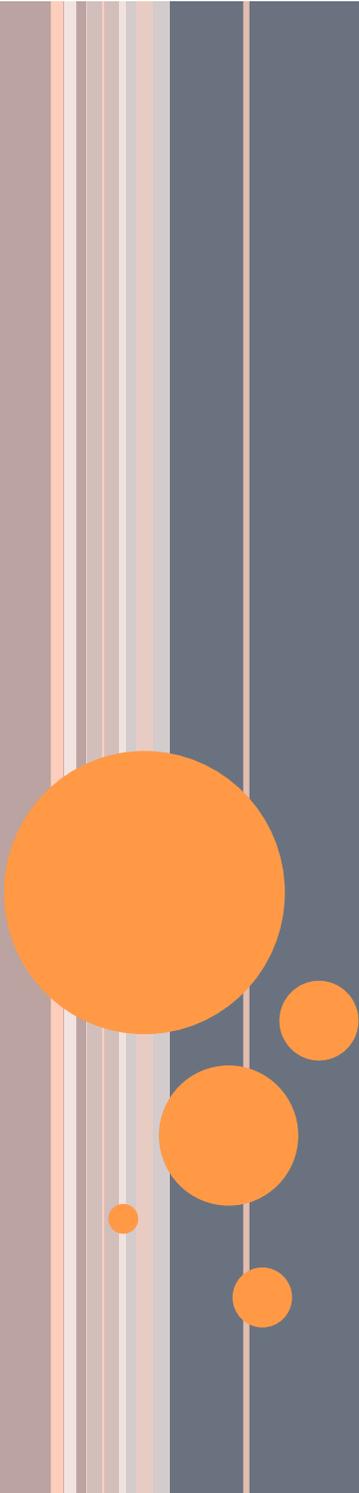
- Règles inconscientes, acquisition:
 - le tout jeune enfant dit « *vous faites* » puis « *vous faites* » puis de nouveau « *vous faites* ».
 - Le « *vous faites* » ne peut provenir que de règles (surgénéralisation).
- Règles inconscientes, coréférence
 - *Le chien de Paul pense qu'il ne l'aime pas.*
Il ≠ tout le reste est possible.
 - *Il a aimé trois des livres que Jean Echenoz a écrit.*
Il ≠ Jean Echenoz.
 - *Combien de livres que Jean Echenoz a écrit a-t-il aimé?*
Il = Jean Echenoz possible

DEUX PRINCIPES

- Les phrases sont analysables (compréhensibles) en temps raisonnable (traduction informatique: en temps polynomial)
- Il existe un bon algorithme d'apprentissage de la grammaires à partir d'exemples positifs en nombre relativement faible.

COMPÉTENCE / PERFORMANCE

- Les règles / Nos limites (mémoire)
 - Le loup a dévoré la chèvre.
 - La chèvre que le loup a dévoré avait mangé le chou.
 - ? Le chou que la chèvre que le loup a dévoré avait mangé appartenait au passeur.
 - ?? Le passeur auquel le chou que la chèvre que le loup a dévoré avait mangé appartenait possède plusieurs bateaux.
 - ??? Les bateaux que le passeur auquel le chou que la chèvre que le loup a dévoré avait mangé appartenait possède sont des barges.
- (néanmoins correct, en prenant son temps et un crayon)



**QUELS
LANGAGES FORMELS
POUR LA
SYNTAXE DU LANGAGE
NATUREL**

**Avec seuls le principe de complexité de l'analyse
et d'adéquation descriptive**

EXEMPLE DE GRAMMAIRE

s → sn sv

sn → det n | np | det n rel _s | det n rel _o | np rel _s | n p rel
_o

rel _s → pro_s sv

rel _o → pro_o sn vt

sv → vi | vt sn

pro_o → que

pro_s → qui

vt → regard e | regardent | mange | mangent

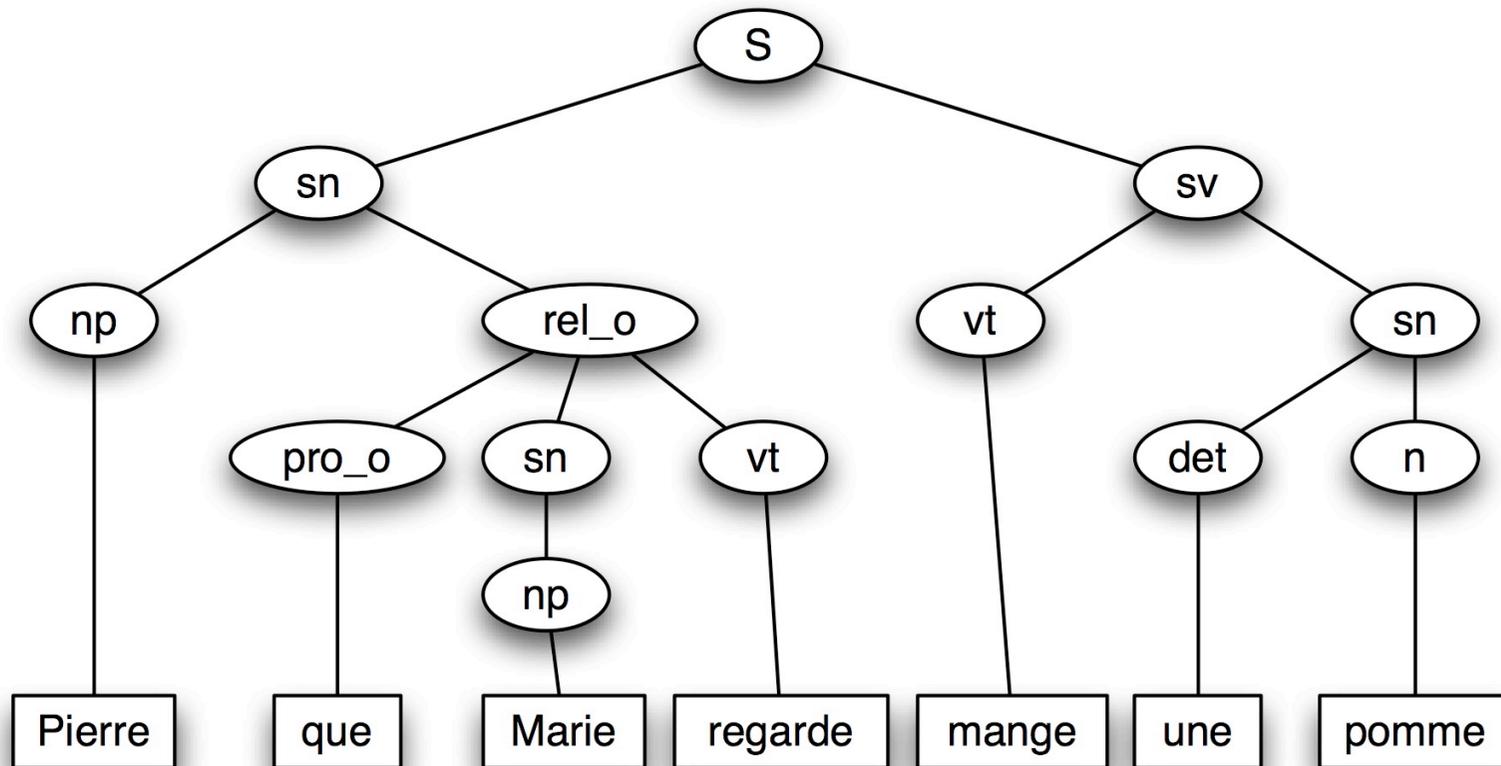
vi → dort | dorment | tombe | tombent

det → une | un | la | le | des | les

n → pommes | pomme | femme | femmes

np → pierre | marie

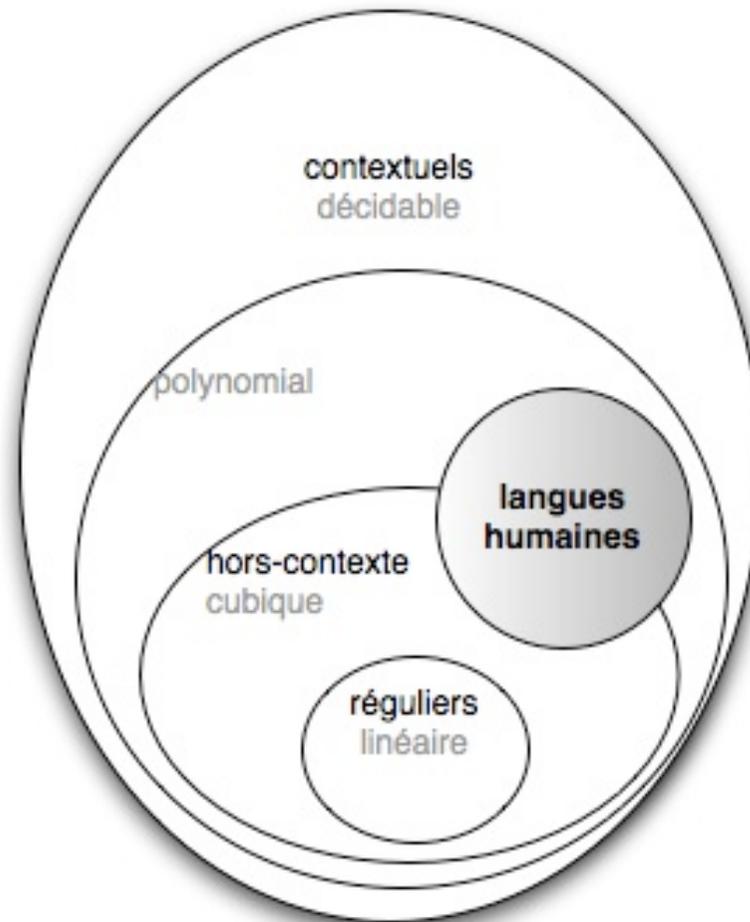
EXEMPLE DE DÉRIVATION



QUEL TYPE DE RÈGLES?

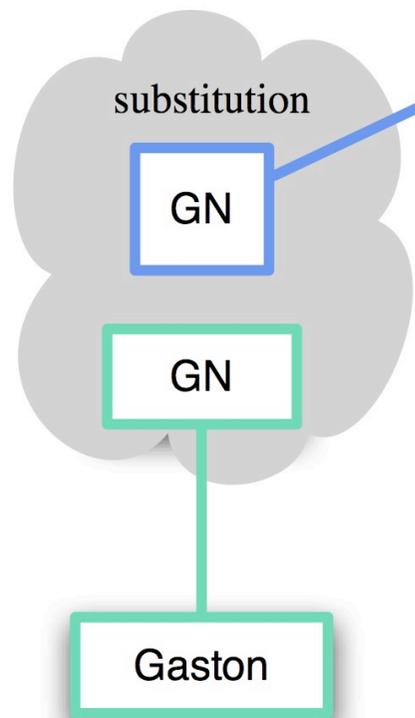
- Sujet1 Sujet2 Sujet3 ...
.....Verbe3 Verbe2 Verbe1
(ex. précédent relative avec « que »)
- Sujet1 Sujet2 Sujet3 ...
.....Verbe1 Verbe2 Verbe3
(complétives NL)

HIÉRARCHIE DES LANGAGES FORMELS

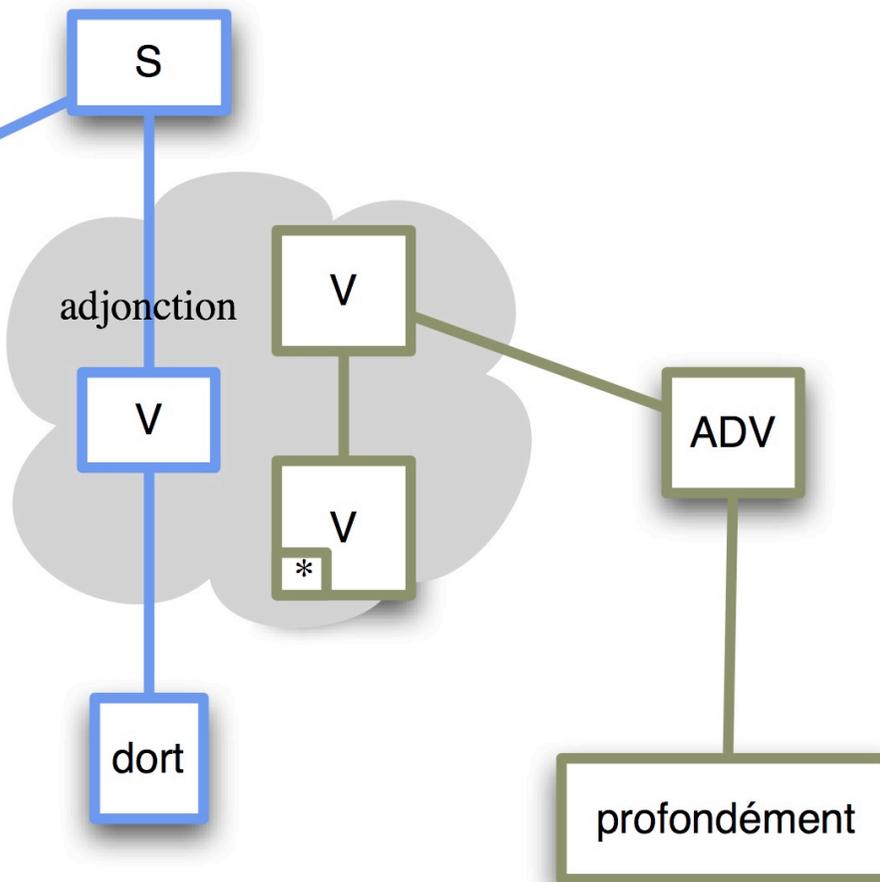


GRAMMAIRES D'ARBRES ADJOINTS

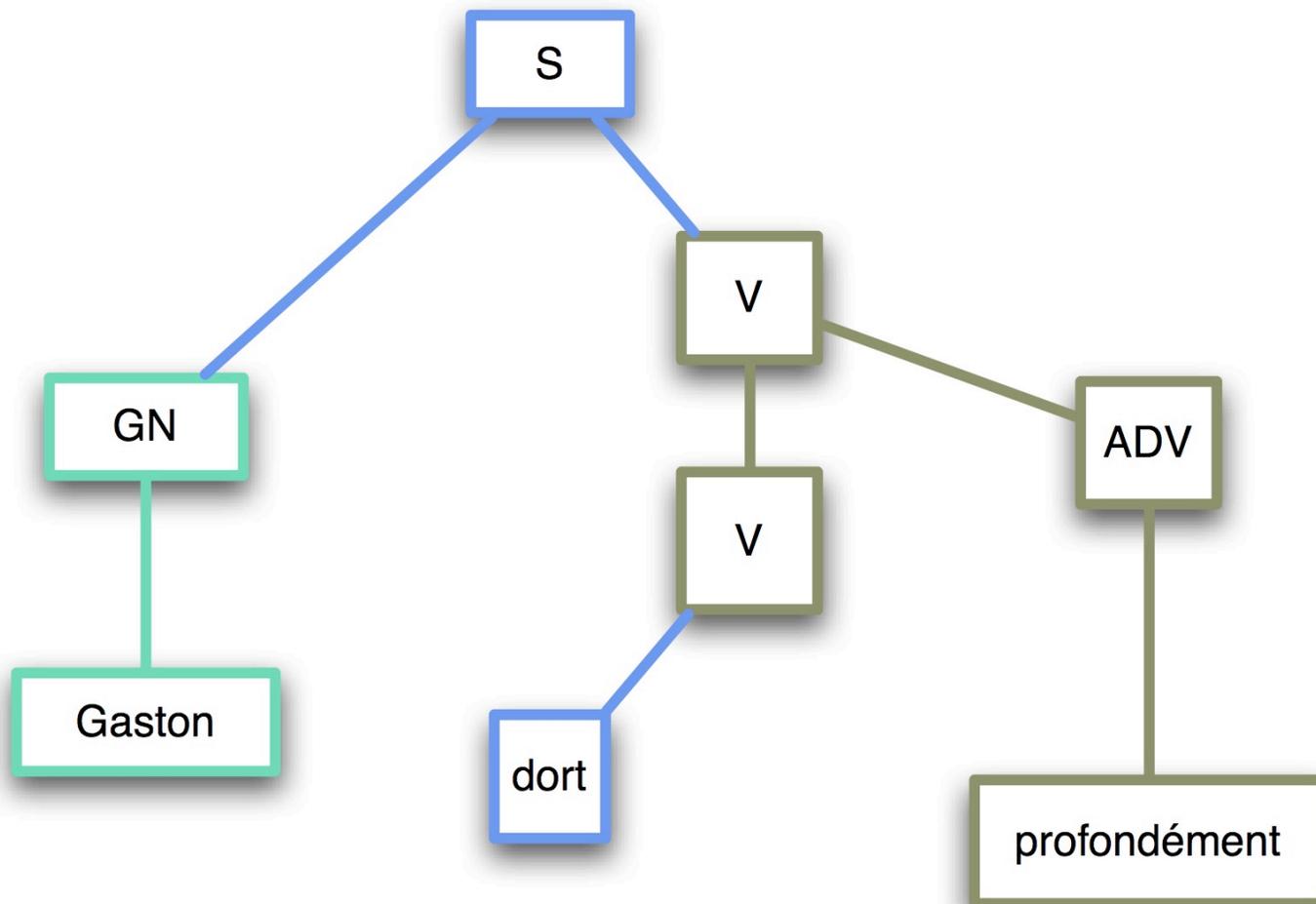
2 arbres élémentaires

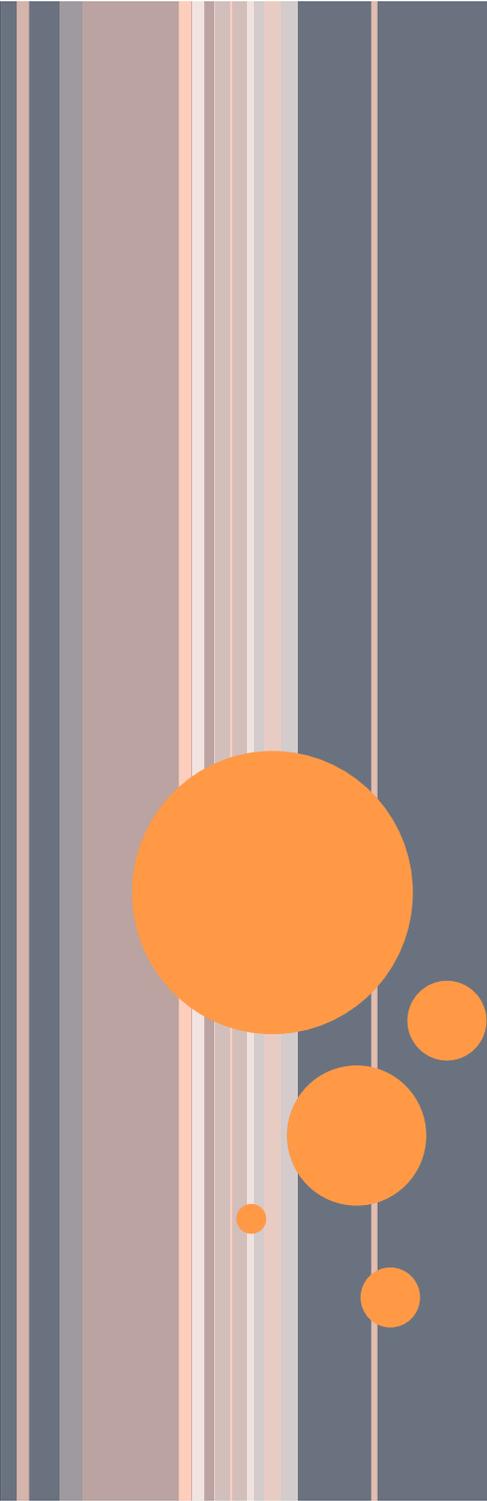


un arbre adjoint



GRAMMAIRES D'ARBRES ADJOINTS





ACQUISITION DE LA SYNTAXE ET GRAMMAIRE UNIVERSELLE

Après l'efficacité de l'analyse,
Deuxième guide de la grammaire générative.

FAITS CONNUS SUR LE JEUNE ENFANT 18-36 MOIS

- Exemples positifs seulement
- Exemples assez peu nombreux par opposition à complexité d'une langue naturelle
- Sens des mots connu au préalable
- Intonation utile

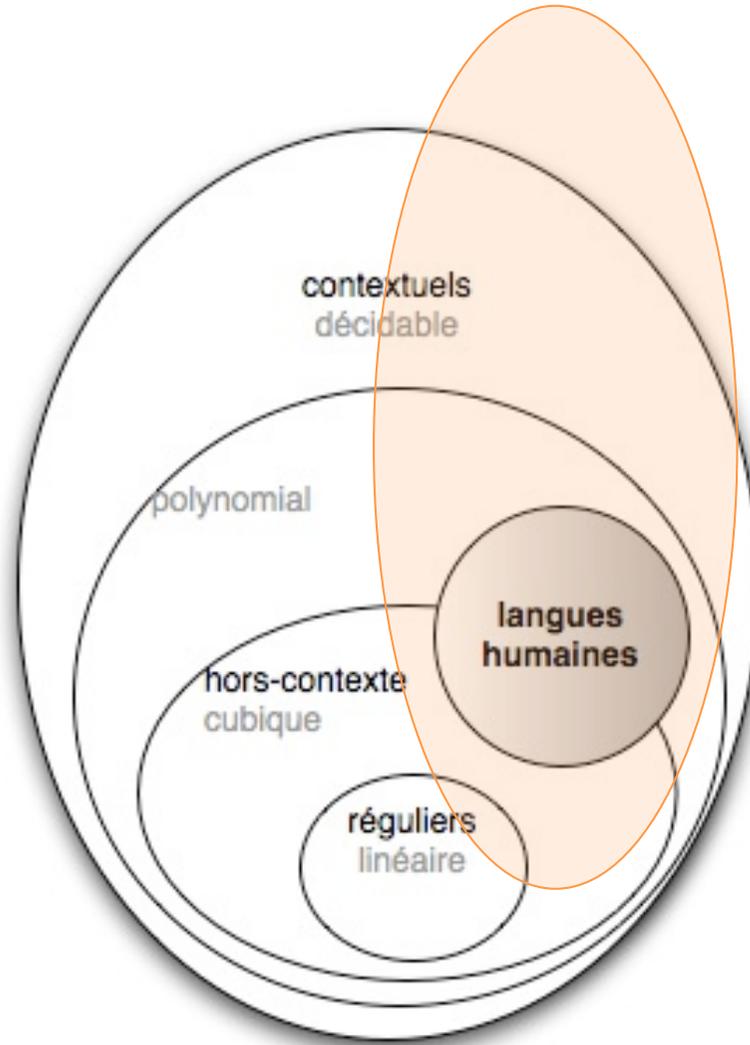
L'HYPOTHÈSE DE LA GRAMMAIRE UNIVERSELLE

- Pas une grammaire au sens usuel mais des contraintes sur la forme des grammaires des langues humaines
- Avec cette hypothèse le processus d'acquisition devient explicable
- Apprentissage par choix de paramètres
- Exemple bête SVO ou SOV?
un exemple suffit:
 - maman conduit la voiture

MODÈLE DE GOLD: CLASSE APPRENABLE

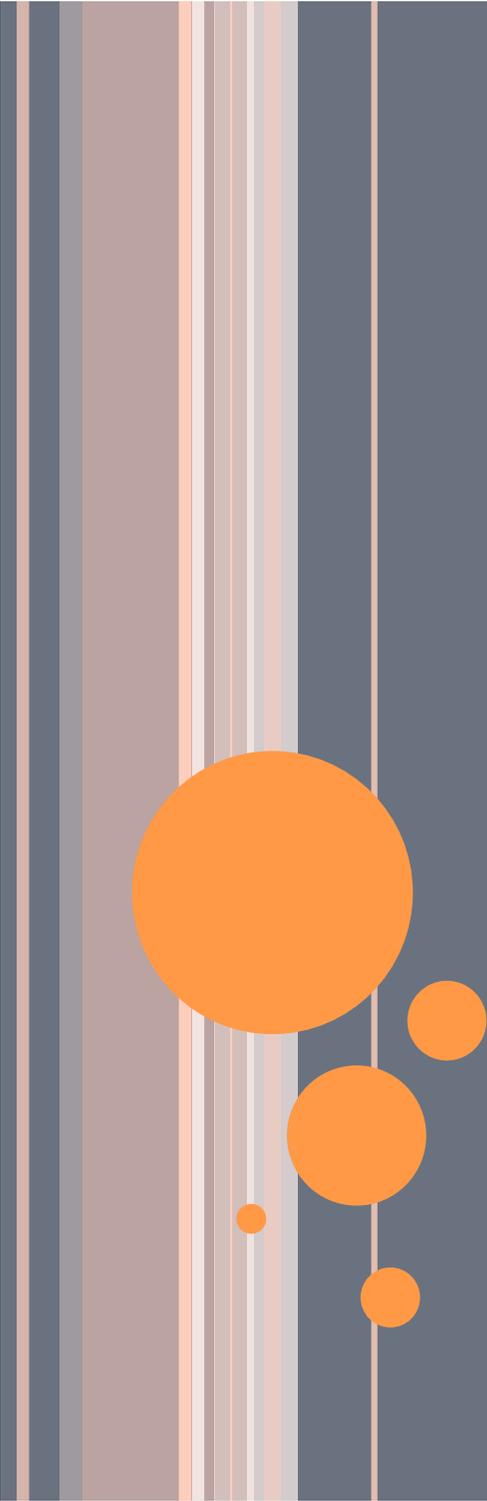
- Algorithme
Phrase₁.... Phrase_n → Grammaire G_n
G_n engendre Phrase₁.... Phrase_n
- SI la totalité des Phrase₁.....
énumèrent un langage L de la classe
- ALORS à partir d'un nombre fini n
d'exemples, l'hypothèse faite G_n
ne varie plus et le langage engendré par
G_n est L.

LANGAGES APPRENABLES ET HIÉRARCHIE



LIMITES DE L'APPROCHE À LA GOLD

- (Pratique) Nécessite des exemples avec beaucoup d'informations, très structurés: données disponibles?
- (Théorique) Les algorithmes, dits par généralisation, font grossir le langage jusqu'au langage cible. Chez l'enfant, c'est l'inverse...
Algorithme d'apprentissage par spécialisation?



LINGUISTIQUE ET LOGIQUE

Une longue tradition (Antiquité, Moyen-Âge)

Vers un calcul informatique ou cognitif du “sens”.

COMPRENDRE LES PHRASES ANALYSÉES

- Arbres, structures calculées automatiquement, mais ensuite, comment leur associer des formules manipulables par une machine?
- Quels problèmes rencontre-t-on?
Ils sont très classiques
... mais pas faciles.

LOGIQUE ET GRAMMAIRE: 23 SIÈCLES D'HISTOIRE

- Depuis l'antiquité (Aristote, Denis de Thrace)
 - puis au Moyen-Âge (scholastique),
 - au 18e (Port-Royal)...
- La phrase a une structure logique, importante en pratique.
- *Les enfants prendront une pizza*
 - Chaque en prend une pizza pour lui?
 - Ils partagent la même?

RICHESSE LOGIQUE DES LANGUES

- *Un, des, certains, tous, tous les, tout, chaque*
+ d'autres quantificateurs:
la plupart, les, un grand nombre de, un petit nombre de,
 - La plupart des étudiants travaillent.
- Les nombres: encore des quantificateurs:
 - *Mettre huit gouttes dans trois cuillères à soupe d'eau.*
 - $3 \times 8 = 24$ gouttes?
 - 8 gouttes?

PROBLÈMES DE PORTÉES, SUITE

○ Lectures *de re* et *de dicto*

- 007 croit que l'un des chercheurs du laboratoire est un espion.
- Il existe un chercheur dont 007 croit qu'il est un espion: 007 pense que le chercheur Blofeld est un espion.
- 007 croit qu'il existe un espion dans le laboratoire, mais il ne soupçonne personne en particulier: 007 a trouvé un microfilm dans le laboratoire.

PROPRIÉTÉS DE PROPRIÉTÉS

- Une fraise rouge vermillon
 - Rouge propriété, vermillon propriété de propriétés
- Adjectif “Bon”
 - tous les médecins sont des conducteurs
 - (donc) tous les médecins bordelais sont des conducteurs bordelais
 - *(donc) tous les bons médecins sont des bons conducteurs

ATTENTION À NE PAS S'ÉCARTER DE CE QUI EST DIT

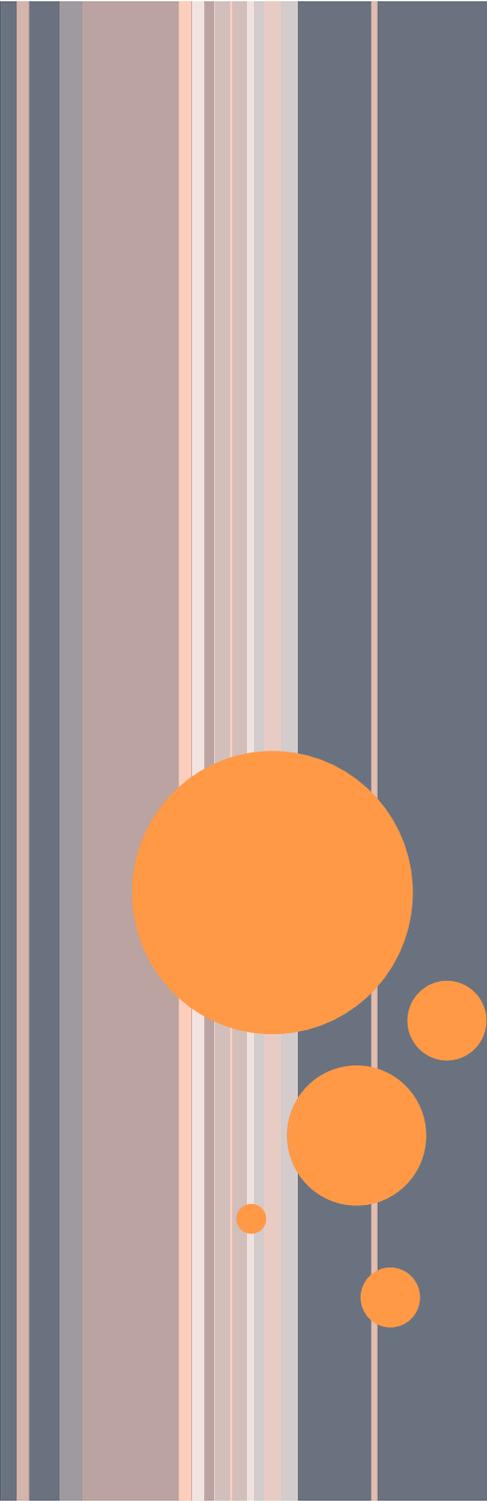
- *J'avais trois trombones dans ma poche, je les ai tous perdus sauf un.
Je le range dans un tiroir.*
- *J'avais trois trombones dans ma poche, j'en ai perdu deux.
* Je le range dans un tiroir.*
- Pourtant, d'un point de vue purement logique, la situation est identique.

LIEN AVEC LA SYNTAXE

- Aspects sémantiques des catégories syntaxiques
- Les catégories ou parties du discours ont une contre partie logique.

LIEN AVEC LA SYNTAXE

- Groupes nominaux: individus
 - *Le chien de ma voisine.*
- Verbes, propriétés
 - 1 arguments *X dort, X regarde Marie*
 - 2 arguments *X regarde Y*
 - 3 arguments: *X donne Y à Z*
- Adjectifs: noms? verbes?
- Groupes prépositionnels :
ni des prédicats, ni des individus:
modifient des propriétés



ASPECTS COGNITIFS: QUELQUES EXEMPLES

Notre utilisation quotidienne du langage:
Les modèles peuvent-ils prédire nos difficultés.

EXPÉRIENCES

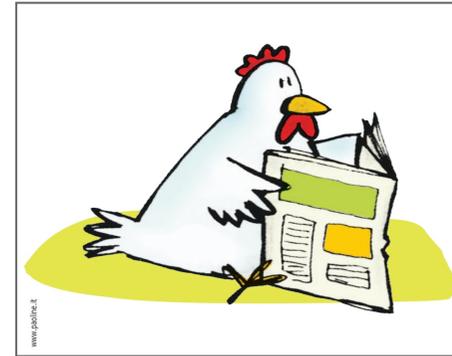


Figure 2: An example of Exp. 1

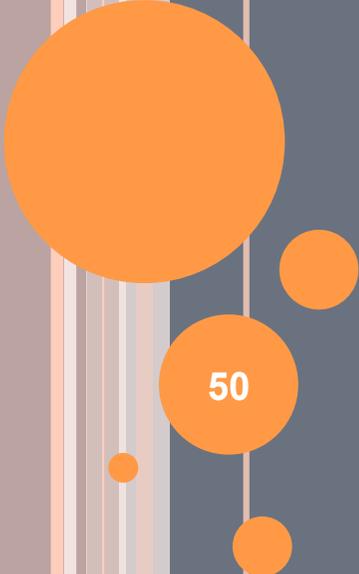
- Phrase, image:
la phrase est-elle vraie dans l'image?
 1. *La poule lit le journal.*
 2. *La poule fait du shopping.*
 3. *La poule ne lit pas le journal.*
 4. *La poule ne fait pas de shopping.*
- Mémoire à court terme:
4 est le pire: imaginer, comparer.
pour enfants dyslexiques
50% oui 50% non

QUANTIFICATION NATURELLE PLUS RICHE QUE LA QUANTIFICATION LOGIQUE

- La plupart...la majorité....
Modèle usuel: cardinal $|\{x \mid A(x)\}| < |\{x \mid \neg A(x)\}|$
Ce modèles est très contestable. Tests:
 - *La plupart des nombres sont premiers.* NON
 - *La plupart des nombres ne sont pas premiers.* OUI
- Le langage contient des quantificateurs de Henkin.
Invérifiable (complexité), pourtant on en entend!
 - *Un membre de chaque conseil de labo connaît membre de chaque conseil municipal de la CUB.*
- L'analyse formelle en terme d'ensembles est très contestable. Il vaut revenir au point de vue Aristotélicien, Avicennien ...
L'usage, preuves et réfutations, dialectique (jeux).

LA SIMPLE QUANTIFICATION EST ELLE AUSSI SIMPLE

- « Tout »:
 - objet générique, universel (scholastique)
 - ou ensembles
 - « la restriction est incluse dans le prédicat »
- Ça dépend:
 - Latifa, Myriam et Nabil ont tous réussi.
 - Mes enfants ont tous déjà eu la varicelle.
 - Tous les étudiants ont réussi.
 - Tout homme est mortel.



PLAN RAPPROCHÉ: SÉMANTIQUE LEXICALE DANS UN CADRE COMPOSITIONNEL

50

Un peu de subtilité dans un modèle trop carré

COMPOSITIONNALITÉ SÉMANTIQUE LOGIQUE

- *Tout barbier se rase lui même. Je le crois.*
- *Tu ne finis jamais un livre. Cela m'énerve.*
- Avantages:
 - *qui fait quoi*
 - logique d'ordre supérieur,...
 - références, coréférences, quantifications,...
- Inconvénients
pas de lien entre *barbier* et *raser* (pour ex. 1, ça va)
ni entre *livre* et les prédicats *lire*, *écrire*, *imprimer*
et c'est un problème pour ex. 2.

SÉMANTIQUE LEXICALE

SENS DES MOTS, SYSTÈME CONCEPTUEL LANGAGIER

- Rôle thématiques:
 - *Une cigarette rapide.*
 - *Un sourire moqueur.*
- Glissement de sens, contenant, contenu, associations d'idées:
 - *Boire un verre.*
 - *Je suis garé en face.*
 - *Le 10 est un as.*
 - *La table 2 prendra un pichet de rouge.*

LES FACETTES EN SÉMANTIQUE LEXICALE

○ Une facette

- *Un livre lourd.*
- *Un bon livre.*
- *Un dîner sympathique.*
- *Un dîner exquis.*
- *Sète est une ville portuaire.*
- *Paris envoie des soldats en Afghanistan.*
- *Bordeaux a battu Marseille.*
- *Un délicieux saumon.*
- *Un saumon vif et rapide.*

LES FACETTES EN SÉMANTIQUE LEXICALE

- Coprédication sur plusieurs facettes.
 - *Un livre lourd mais intéressant.*
 - *Une ville portuaire et cosmopolite.*
 - *Le dîner était délicieux mais a duré des heures.*
 - *? Un saumon rapide mais délicieux.*
 - *? Washington borde le Potomac et a attaqué l'Irak.*
 - *? Bordeaux est sur la Garonne et a battu Marseille.*

FACETTES, RÉFÉRENCE, QUANTIFICATION

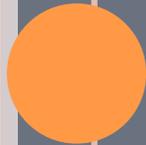
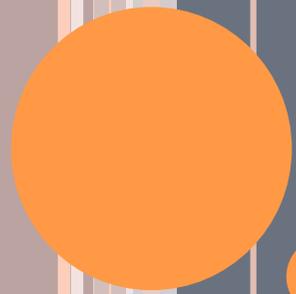
- *Etagère avec les livres de Flaubert*
 - *deux copies de Madame Bovary,*
 - *trois de l'Education sentimentale,*
 - *un volume contenant les Trois contes*
- *Expérience....*
 - *J'ai lu tous les livres de l'étagère.*
 - *Je vais les monter au grenier.*
 - *Combien de livres ai-je lus?*
 - *Combien de livres vais-je porter?*
- *Pronom renvoyant à l'espèce.*
 - *Racoons are in the cellar.*
 - *They quickly multiply.*

TYPES ET ONTOLOGIES CONTRE IDIOSYNCRASIES LANGAGIÈRES

- *Ma voiture est crevée.* (une roue, pas une durite)
- * *Ma voiture est bouchée.* (pourtant toutes les voiture ont un carburateur, des injecteurs etc. susceptibles de se boucher)
- Le langage agit comme un filtre.
Il n'est pas l'ontologie.
Le langage instancie certains lien conceptuels.
- Vice-versa le langage crée des liens conceptuels.
Par exemple ***girouette***:
rencontre en Normandie de ***girare*** (tourner en latin)
et de ***veðr*** (vent, temps en islandais),
puis notion qui prend un sens figuré...

ACQUISITION DU LEXIQUE

- Structure argumentale et sens avant sa syntaxe.
- Sur machine construction de graphes de mots associés,
 - recherche de proximités sur Internet
 - « Jeux de mots » (LIRMM) sur Internet
- Extraction de sens (composantes très connexes)
- Problèmes: mot-prédictat mot-argument
comment trouver les sens spécifiques associés,
et les probabilités des liaisons:
commencer un livre, voiture crevée,...

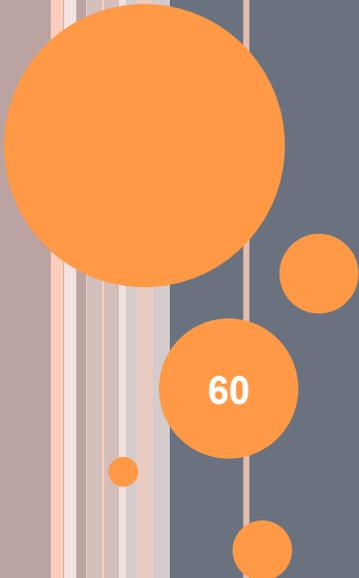


CONCLUSION

de cet aperçu morcelé

BILAN ET PERSPECTIVES

- Modèles mathématiques, algorithmes: réalisations informatiques possibles.
- Outils: manque de ressources linguistiques corpus, grammaires, dictionnaire
- Point de vue théorique, compétence: principes régissant le langage (sémantique encore méconnue).
- Cognition, performance: débuts prometteurs modèles — tests, corpus, expériences



MERCI!

Des questions?

60