



# Sémantique Lexicale pour le TAL

Françoise Gayral

# Un peu de vocabulaire

- **Sémantique** : étude du sens
- **Lexique** : ensemble des mots d'une langue donnée.  
==> **Sémantique lexicale**: étude du sens des mots d'une langue
- Concerne différentes disciplines de la linguistique :
  - **Lexicologie** = étude du **lexique** (mais l'étude du lexique ne traite pas que le sens : morphologie, étymologie, prononciation, syntaxe), discipline **théorique**
  - **Lexicographie** = élaboration des dictionnaires, discipline **appliquée**
- Concerne le TAL

# Etendue du travail

## ■ Quantité d'information à traiter :

- Dictionnaire du français : 60 000 entrées, chacune en moyenne de 4 à 5 sens (240 000 à 300 000 sens !!) + expressions figées

→ 1 million à traiter

Exemple : Petit Robert (1999) 67000 entrées (noms communs)

## ■ Nombreuses informations à fournir pour une unité lexicale

- flexion, prononciation, définition ou représentation sémantique, constructions syntaxiques liées,...

# Pourquoi de la sémantique lexicale en TAL ?

## ■ Pour la traduction automatique

- *Le robinet fuit/Le voleur fuit -> leak/run away*
- *Paul a échoué dans sa tentative/Paul a échoué dans un petit hôtel -> fail in attempt/end up*

## ■ Pour la recherche de documents

- pour éviter le silence      ex : *vendre/céder/donner*
- pour éviter le bruit      ex : *avocat*

## ■ Pour les systèmes de questions/réponses

- *Paul a offert des roses à Marie*
  - Paul a-t-il donné des fleurs à Marie ?
  - Est-ce que Marie a des roses ?



# Pourquoi de la sémantique lexicale en TAL ?

- Pour la constitution de terminologie

*appareil de mesure/système de mesure : variantes*

- Pour la génération de textes

- Choisir les bonnes collocations :

– *Devenir grand*                      *tomber amoureux*

– *Réaliser un souhait*                *tenir une promesse*

- Pour la désambiguïisation et la compréhension de textes

*La pièce de cet auteur est mauvaise*

*La pièce de 1 euro est tombée sous le bureau*

*La pièce de derrière est sombre*



# La sémantique lexicale en TAL : un domaine très ouvert

- Beaucoup de modélisations ad-hoc dans de nombreuses applications
- Peu de descriptions du lexique « grandeur réelle »
  - Sauf wordnet et framenet (mais pour l'anglais et...)

# Questions préalables

## lexique = ensemble de mots ?

- mot = unité de base qui a un sens ???

- Il existe des unités lexicales composées de plusieurs mots

*prendre la fuite = déguerpir*

*≠ prendre le train*

*pomme de terre = carotte, navet*

*≠ pomme de Normandie*

*≠ sac de terre*

→ **problème des expressions figées**

- Une unité lexicale peut se décomposer en unités significatives plus petites que le mot

*désengagement --> dés-engage-ment*

→ **problème de la morphologie**



# Questions préalables :

## nombre de mots du lexique ?

- Peut-on répertorier le lexique d'une langue ?

- Créativité lexicale, néologismes, emprunts à d'autres langues, acronymes

*Mal-bouffe, mal-logé*

*désamiantage, lepénisation, raffarinade, chiraquie, mitterrandisme*

*week-end, jogging, mail*

*OGM*

→ Le lexique d'une langue est potentiellement infini





# Récréation : créativité lexicale

- *Le commissaire divisionnaire Cercaire dispose d'un témoignage accablant au sujet de la drogue, et presque d'un flagrant délit. Or Julie Corrençon a été droguée avant d'être **dépointée**.*
- *« Dépointée » ?*
- *Un néologisme que je m'autorise, monsieur par glissement du verbe «défenestrer».*
- *Je ne sais pas si je dois permettre de pareilles audaces dans mon service, Pastor.*
- Peut-être préféreriez-vous « **empénichée** » , monsieur.*

*La fée carabine, D. Pennac*

# Questions préalables :

## Mais c'est quoi le sens ?

### ■ sens référentiel

*Le vainqueur d'Austerlitz / Le vaincu de Waterloo*

*Le vilain mari / le prince charmant*

### ■ sens différentiel

*Un sou est un sou      à ce prix-là, elle ne devrait pas être à ce prix-là*

### ■ sens inférentiel

*Le vase est tombé*

### ■ sens implicite

*Même Pierre a eu une bonne note*

-> sous-entendu

*Pierre a arrêté de fumer*

-> pré-supposé

# Questions préalables :

## Mais c'est quoi le sens ?

- sens instructionnel

*Je vous déclare mariés*

*Je te promets...*

- sens connotatif

*sans-papier/clandestin*

*policier/flic*

*chaussure/pompe*

- sens en contexte

*L'omelette est partie sans payer*

*il fait chaud*

*Qu'est-ce que tu fabriques ?*

- sens argumentatif

*Je roulais sur la partie droite de la chaussée*

→ Multiples dimensions du sens



# Différents angles d'étude du lexique

- **Forme → Morphologie lexicale**  
segmentation des mots, décomposition en morphèmes
- **Sens → Sémantique lexicale**
- **Distribution → Interface lexicale – syntaxe**

Ici b et c

# Questions à résoudre pour la sémantique lexicale en TAL

- Comment **définir** le sens lexical et quelle **représentation** en donner ?
  - Quel contenu ?
- Comment prendre en compte les **relations sémantiques** entre les unités lexicales ?
  - synonymie, homonymie, antonymie, hyponymie, polysémie, collocation
  - Quelle organisation ?
- Comment **interfacer** niveau **sémantique** et **syntactique** ?
  - Quel lien à la syntaxe ?

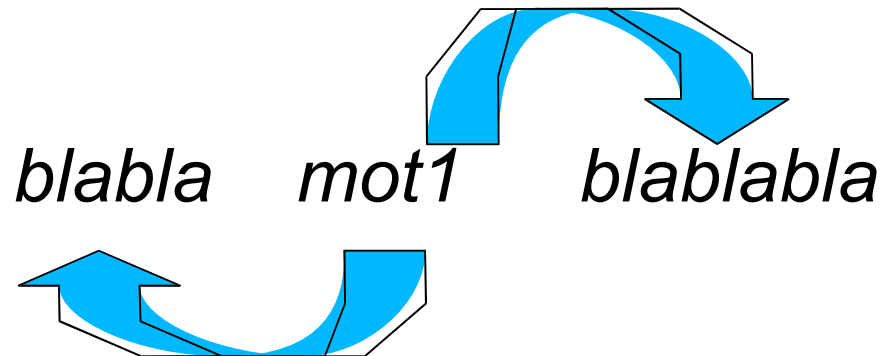
# Relations sémantiques entre unités lexicales à prendre en compte

- Relations paradigmaticques (substitution)

*blabla mot1 blablaba*  
                  ↓  
                  *mot2*

synonymie, hyperonymie, méronymie, antonymie

- Relations syntagmatiques (enchaînement des unités lexicales dans le texte)



arguments, collocations

# Relation paradigmatique : Synonymie

- Synonymie : relie des mots ayant des formes différentes mais ayant le même sens.
- La « vraie » synonymie (substituabilité dans tous les contextes) est rarissime
  - Synonymie partielle (ou contextuelle) = dans certains contextes.
    - Ex : *éprouver de la peur, éprouver de la crainte*  
*avoir peur, \*avoir crainte*
  - Plutôt quasi-synonymie
- Attention : synonymie ≠ dénomination multiple  
identité référentielle, mais pas sémantique  
*Le vilain mari et le prince charmant*



# Synonymie : exercices

- *grave/sérieux*
- *sommet/cime*
- *battre/frapper*
- *technicienne de surface/femme de ménage*
- *fute/pantalon*
- *longue maladie/cancer*



# Relation paradigmatique: Hyperonymie

- **Hyperonymie** : relation hiérarchique entre un mot et un mot plus général

*rose/fleur, chien/animal, voiture/véhicule, grignoter/manger*

*rose est un hyponyme de fleur*

*fleur est un hyperonyme de rose*

- Relation naturelle dans l'activité définitoire :  
décomposition par genre et différence spécifique

*La rose est une fleur qui....*

*grignoter est une façon de manger*

- Relation utilisée dans certains procédés linguistiques :
  - Coordination : *des roses et d'autres fleurs.*
  - Relations anaphoriques : *j'ai rapporté des roses et des tulipes, puis j'ai mis les fleurs dans le vase*
  - Et même : *au lieu d'acheter une voiture, achetez une SAAB quand on regarde Canal +, au moins on n'est pas devant la télé*

# Relation paradigmatique: méronymie

- **Méronymie** : relation partie/tout

*main/bras, pétale/fleur*

*Main* est le méronyme désigne la partie

*bras* est l'holonyme

Composant/objet complet	<i>pédalier/bicyclette</i>
Membre /collection	<i>arbre/forêt</i>
Portion/masse	<i>flocon/neige</i>
Matière/objet	<i>métal/voiture</i>
Sous activité/ activité	<i>payer/acheter</i>
Zone/lieu	<i>oasis/désert</i>



# Méronymie : exercices

- Relation utilisée dans certains procédés linguistiques :

- Anaphore associative : *Il regarde le village; l'église est superbe*

- Compléter le tableau
- Chercher des exemples et contrexemples d'anaphores associatives

# Autres relations paradigmaticques

## ■ Antonymie : contraires

- Complémentarité : on ne peut pas nier les deux en même temps

*pair/impair, marié/célibataire*

- Antonymie scalaire ou graduable : on peut nier les deux en même temps, la gradation est possible

*grand/ petit --> il est (très/assez/plus) grand*

- Réciprocité, inversion

*partir/arriver, augmenter/réduire, mari/femme*

## ■ Actant typique : *acheter/acheteur*

## ■ Instrument typique : *patiner/patin*



## Antonymie : exercices

- Complémentarité
- Antonymie scalaire
- Réciprocité, inversion

# Relations syntagmatiques

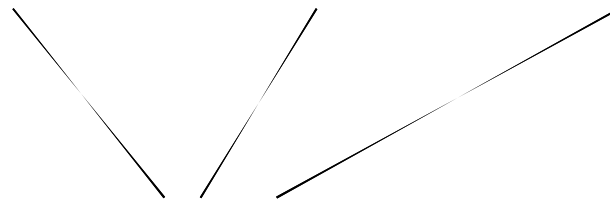
## ■ Collocations

- **corrélats** : mots qui apparaissent dans le voisinage du mot-cible, sans être nécessairement reliés par un rapport de dépendance syntaxique
    - verbes fréquents avec *barrage* : *construire, édifier, démolir, financer*
    - adjectifs fréquents avec *barrage* : *grand, futur, gigantesque, coûteux, haut, gigantesque, monumental, ultramoderne, hydraulique, hydroélectrique*
  - mots présentent des affinités et tendant à apparaître ensemble :
    - argument de poids, faim de loup, tomber amoureux, grièvement blessé*
- Surtout intéressant pour la génération

# Relations syntagmatiques

- **Structure argumentale** : participants (arguments) attendus par le mot

Acheter : qu'un achète qq chose à qq'un



3 actants

*Acheter* est un prédicat qui attend 3 arguments (**arité**) qui jouent le rôle de l'agent, de thème, de source

→ **Intéressant pour le lien avec la syntaxe**



# Relation paradigmatique et interfaçage niveau sémantique et syntaxique

- Le lexique précise des contraintes de sélection concernant son environnement, surtout pour les unités lexicales prédicatives: verbes et prépositions
  - en relation avec la syntaxe
    - Schéma de sous-catégorisation syntaxique
  - en relation avec une structure profonde
    - Structure argumentale
    - Rôles thématiques
  - en relation avec des propriétés sémantiques des arguments
    - Restrictions de sélection ou sous-catégorisation sémantique



# Schéma de sous-catégorisation

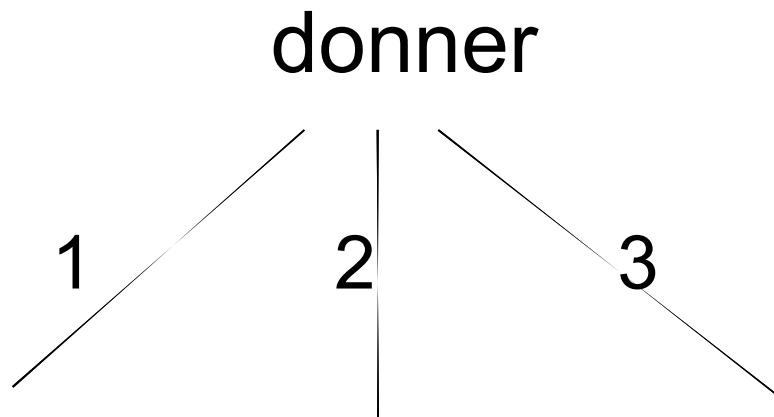
- indique les constructions syntaxiques dans lesquelles entre l'unité lexicale, en particulier le type syntaxique des compléments du verbe et leur caractère obligatoire ou optionnel
  - *Courir* : SN1
  - *Manger* : SN1, SN2
  - *Donner* : SN1, SN2, SP3 (à)
  - *Envieux* : SN1, SP2 (de)
- mais difficultés à définir le caractère optionnel
  - J'ai beaucoup donné cette année*
  - Paul a déjà mangé*

# Structure argumentale

- Méthode 1 : Représentation « logique »

Exemple : *donner* est un prédicat à trois arguments  
(Ex : DEC de Mel'cuk)

donner(X,Y,Z)



# Structure argumentale

- Méthode 2 : cas sémantiques (Fillmore 65) ou rôles thématiques (Gruber)
  - Des rôles sont associés aux arguments
  - Inventaire de rôles thématiques universels (?)
    - **agent** : fait ou cause l'action, intentionnellement ou non
    - **thème** : entité concernée par l'action (souvent non animé)
    - **patient** : entité à qui il arrive qq chose du fait de l'action et qui en est affecté (souvent animé)
    - **expérenceur** : entité impliquée dans un événement psychologique ou dans un état mental, recevant des sensations ou des émotions
    - **bénéficiaire** : entité qui bénéficie de l'action
    - **localisation** : désigne le lieu de l'action
    - **instrument** : moyen utilisé pour l'accomplissement de l'action.
    - **but** : désigne le but final de l'action.
    - **source** : désigne la source de l'action



- Agent

- -- A participant which the meaning of the verb specifies as doing or causing something, possibly intentionally. Examples: subjects of kill, eat, hit, smash, kick and em watch.

- Patient

- -- A participant which the verb characterises as having something happen to it, and as being affected by what happens to it. Examples: objects of kill, eat and smash but not those of watch, hear and love.

- Experiencer

- -- A participant who is characterised as aware of something. Examples: subject of love or object of annoy.

- Theme

- -- A participant which is characterised as changing its position or condition, or as being in a state or position. Examples: objects of give and hand, subjects of walk and die.

# Exemples

*Jean (agent) frappe le ballon (thème)*

*Jean (patient) reçoit un coup (thème)*

*Jean (bénéficiaire) reçoit un cadeau (thème)*

*Jean (experienceur) aime Marie (patient)*

- **Remarque** : La fonction grammaticale *sujet* peut être remplie par un argument jouant des rôles différents suivant le verbe
- Bien distinguer fonctions grammaticales et les cas sémantiques

# Structure argumentale

- L'entrée lexicale d'un verbe spécifie sa structure argumentale ou grille thématique
  - Donner [agent, thème, but]
  - Acheter [agent, thème, source]
  - Aimer [experienceur, patient]
  - Casser [agent, thème, instrument ]
- Ce qui permet le lien à la syntaxe :
  - Rôle thème et patient souvent réalisés en cod
  - Rôle agent souvent réalisé en sujet, si pas d'agent alors sujet peut être l' experienceur, le bénéficiaire,...
  - Autres rôles réalisés par des groupes prépositionnels
- Exemples dans des formalismes syntaxiques (LFG, HPSG)
  - Donner* [subcat : <NP1, NP2, PP3 [à]>, ROLES: <agent, thème, but>]

# Mais...

- Difficultés à s'accorder sur un inventaire préalable de rôles (pas de listes de rôles et de définitions faisant consensus)
- Difficultés à définir la grille thématique
  - nombre d'arguments ? caractère optionnel ?
    - instrument pour *casser* ?  
*Jean (A) a cassé la fenêtre (T)*  
*Jean (A) a cassé la fenêtre (T) avec un marteau (I)*
- Difficultés à définir puis attribuer les rôles
  - effrayer [THÈME, EXPÉRIENCEUR] Ce film effraie Marie
  - effrayer [AGENT, EXPÉRIENCEUR] Paul effraie Marie
  - Jean a donné un livre à Marie (Beneficiaire)*
  - mais *Jean a donné un coup à Marie (But)*

# Mais...

- Deux prédicats peuvent mettre différemment en relation des participants de la même situation
  - 1) *Paul a vendu sa voiture à Jean*
  - 2) *Jean a acheté sa voiture à Paul*
    - *Paul : agent dans 1) et source dans 2)*
- Difficultés de faire un lien systématique à la syntaxe
  - Les arguments d'un même verbe n'ont pas forcément toujours la même position syntaxique
    - alternances
      - Jean (A) a cassé la fenêtre (T) avec un marteau (I)*
      - Le marteau (I) a cassé la fenêtre (T)*
      - La fenêtre (T) a cassé*





# Exercice

**Déterminez le rôle thématique des expressions en italiques**

- a. *Lucie* prend un verre avec ses copines
- b. *Lucie* prit le marteau dans l'armoire
- c. *Lucie* prend un coup sur la tête
- d. *Lucie* a pris parti pour Francine
- e. *Lucie* attrape le voleur
- f. *Lucie* attrape la grippe

# Restrictions de sélection

- Indiquent les sélections sémantiques opérées par le verbe (ou la préposition,...) sur ses arguments, à l'aide de :
  - traits sémantiques : humain/non humain, concret/abstrait, comptable/massif,...
  - catégories sémantiques : lieu, entité\_vivante,...
- Exemple :
  - *manger*      agent : animé      thème : comestible, solide
  - Le chien mange*      agent [+animé]
  - Jean mange le pain*      thème [+solide] [+comestible]

# Restrictions de sélection

## ■ Intérêts :

– générer des attentes

– filtrer ce qui est acceptable

\* *La banane mange*

\* agent [-animé]

\* *Jean mange le vin*

\* thème [-solide] [+comestible]

– détecter les métaphores

*La multinationale a mangé tour à tour toutes les PME du secteur*

– détecter les métonymies

*Jean a mangé la boîte*

\*comestible

## ■ Mais pas si simple...

mais *Jean mange la soupe* \*solide

# Homonymie - Polysémie

- 1 mot : 1 sens **Exceptionnel !!!!**
- 1 mot : plusieurs sens mais distinction :
  - Polysémie : un mot a plusieurs sens reliés (d'une manière ou d'une autre)
    - La construction a duré deux ans*
    - La construction a dix étages*
  - Homonymie : un mot a plusieurs sens totalement disjoints, sans rapport entre eux
    - l'avocat a bien plaidé*
    - l'avocat est dans le frigo*

A decorative vertical bar on the left side of the slide, featuring a series of horizontal stripes in various colors including blue, yellow, black, and grey, with a white background behind the stripes.

Exercice : chercher des homonymes,  
des polysèmes

# Homonymie - Polysémie : problème du mot isolé pas du mot en contexte car... le contexte désambiguïse en général

- En contexte, l'interprétation des mots qui ont *a priori* plusieurs sens (homonymes ou polysèmes) se fait sans problème

*Ce livre est intéressant*

*Ce livre a 500 pages*

*j'ai loué ma maison*

*j'ai loué ses qualités d'organisation*

*l'avocat a bien plaidé*

*l'avocat était dans le frigo*

- Même si ...

*ce livre qui a 500 pages est intéressant*

*Mais \*L'avocat qui était dans le frigo a bien plaidé*

# Le contexte désambiguïse sauf...

- sauf dans les jeux de mots (souvent utilisés dans les slogans publicitaires)
- Les ambiguïtés persistent quand on vise un effet rhétorique

*Huit briques en liquide, c'est toujours bon à prendre* (pub du jus en emballage carton)

*Le figaro s'ouvre tous les jours*

*L'Auvergne vous sert des fromages sur un plateau*

*L'été est fini : redevenez fidèle* (pub pour un hebdomadaire)

*Région-SNCF : vous ferez vite le rapprochement*

*Pour vous, notre énergie est inépuisable* (pub Total)

# Homonymie - Polysémie : conséquences pour le lexique

## ■ Question de l'unité du mot

- homonymie : c'est comme s'il existait plusieurs mots  
→ Le mot a plusieurs entrées (souvent numérotées)
  1. *avocat: fruit...*
  2. *avocat : personne qui...*
- polysémie : l'unité du mot est préservée  
→ le mot a une seule entrée subdivisée en plusieurs acceptations reliées par des relations (abstrait, concret, par métonymie, par extension, par analogie, figuré...)
  - *construction* 1. action de construire
  - *construction* 2. par métonymie l'immeuble obtenu comme résultat de l'action de construire...



# Passer d'une définition intuitive à une définition opératoire....

- Question très difficile
  - Quels critères pour décider sens disjoints / sens reliés ?
  - Quels rapports de “parenté” justifient le rassemblement des sens sous une même unité ?
- Les lexicographes sont obligés de trancher
  - méthode diachronique avec des critères d'étymologie
  - méthode synchronique avec des critères morphologiques (de dérivation), syntaxiques, de restrictions de sélection
- Mais ils ne tranchent pas tous de la même façon !!!
  - Exemple :        *douillet* :    *lit douillet* / *enfant douillet*
  - polysémique dans le dictionnaire *le Robert*
  - homonymique dans le *Lexis*

# Et que font les TAListes ?

## ■ Pas de critères rigoureux et objectifs réutilisables

- Arbitraire dans les découpages
- Divergences suivant les dictionnaires

## ■ Critères du TAL devraient être :

- Economie descriptive
- Critères opérationnels
- Différences exploitables par les applications
- “Calculabilité” du sens

→ Souvent ad-hoc

→ Souvent dégroupement (27 sens dans wordnet pour *give*)



## La recette de l'amour fou

Dans un boudoir introduisez un cœur bien tendre  
Sur canapé laissez s'asseoir et se détendre  
Versez une larme de porto  
Et puis mettez-vous au piano  
Jouez Chopin  
Avec dédain  
Égrenez vos accords  
Et s'il s'endort  
Alors là jetez-le dehors

Le second soir faites revenir ce cœur bien tendre  
Faites mijoter trois bon quarts d'heure à vous attendre  
Et s'il n'est pas encore parti  
Soyez-en sûr c'est qu'il est cuit  
Sans vous trahir laissez frémir  
Faites attendre encore  
Et s'il s'endort  
Alors là jetez-le dehors

Le lendemain il ne tient qu'à vous d'être tendre  
Tamisez toutes les lumières et sans attendre  
Jouez la farce du grand amour  
Dites « jamais » dites « toujours »  
Et consommez sur canapé  
Mais après les transports  
Ah ! S'il s'endort  
Alors là foutez-le dehors

Pour se détendre

Une chanson de  
S. Gainsbourg



# Différentes approches de la sémantique

# Différentes approches de la sémantique

1. La **langue** se rapporte au **monde**



**Sémantique référentielle**

2. La **langue** rapporte au **monde** par l'intermédiaire de **l'esprit**



**Sémantique cognitive**

3. La **langue** est un système autonome

**Sémantique structuraliste**



# Sémantique structuraliste ou différentielle

- La langue est un “système” que l'on peut étudier dans son fonctionnement de manière autonome
  - ==> langue = système dans lequel chaque mot n'est définissable que par les relations d'équivalence ou d'opposition qu'il entretient avec les autres et pas par référence à un système extérieur (le monde extérieur)
  - ==> Un mot ne peut pas être défini isolément, par rapport à ce qu'il désigne. Il doit l'être relativement à d'autres mots.
  - ==> structuralisme linguistique : courant dans lequel la langue est étudiée en elle-même comme un système doté d'une structure décomposable**

De Saussure, Chomsky, Pottier, Greimas, Rastier

# Sémantique structuraliste : exemple

## *espérer, craindre*

- en commun l'idée de /sentiment/ mais qui ne se comprend que relativement à d'autres idées, celles de /pensée/ ou d'/action/ qu'on trouverait dans d'autres mots),
- en commun l'idée de /relatif au futur/ (mais qui ne se comprend que par opposition à /relatif au passé/ qu'on trouverait par exemple dans « regretter »)
- se différencient notamment par l'opposition /agréable/ vs /désagréable/

## *scalpel / bistouri*

s'opposent par le trait 'pour les morts'/'pour les vivants',  
partagent les traits 'instrument', 'pour couper',...



# Sémantique référentielle

- La langue sert à parler du monde extérieur, donc à nommer/désigner les entités du monde
- un mot renvoie à un référent et permet a priori d'identifier une catégorie d'objets particulière
- un référent est une entité matérielle ou conceptuelle (objet, lieu, processus, événement,...) appartenant au monde extralinguistique réel ou fictif (*fantôme*)
- le sens d'un mot est ce qui permet de déterminer sa référence

==> sémantique logique ou référentielle

Montague, Carnap



# Sémantique référentielle ou vériditionnelle

- sens d'un mot = ensemble de conditions que doit remplir un référent pour être dénommé par ce mot de manière adéquate
  - Exemple : pour que le nom *chat* soit attribué à une entité du monde, il faut qu'elle rende vraies les propriétés : *miauler, avoir 4 pattes, avoir une queue, être un animal...*
- ==> Modèle des conditions nécessaires et suffisantes**
- Plus largement, sens d'une phrase = ses conditions de vérité
  - Cf cours de Daniel Kayser

# Sémantique cognitive

- “Le langage est une partie intégrante de la cognition humaine” (Langacker, 1987)
- La construction/production du sens sont des activités mentales, cognitives
- Langage non autonome
  - mais en interaction avec :
    - les activités de perception, d'action, de raisonnement,...
    - la mémoire

Vandeloise, Talmy, Lakoff, Langacker

# Sémantique cognitive

- Beaucoup de travaux sur l'espace car le lien à la perception est très fort
- Décrire les représentations et processus cognitifs nécessaires à la production et à la compréhension des expressions spatiales
  - Exemple : dans quelles conditions perceptives telle préposition sera choisie plutôt que telle autre
    - La cigarette est à gauche du fauteuil.*
    - Le fauteuil est à droite de la cigarette.*
    - pourtant mêmes conditions de vérité !!!

==> Représentations du sens schématiques, icôniques...



# Différentes sources de ce cours

- Agnès Tutin (Grenoble)
- Claire Gardent (Nancy)
- Jacques Moeschler (Genève)
- Benoit Habert (Nanterre)
- Pollet Samvelian (Paris 3)
- Henriette Gezundhajt (Toronto)