
Introducton au TALN

Isabelle Tellier

LIFO, Université d'Orléans

[http ://www.univ-orleans.fr/lifo/Members](http://www.univ-orleans.fr/lifo/Members)

1. Applications et enjeux du TALN
2. Qu'est-ce que le "langage naturel"
3. Du LN au TA

1. Application et enjeux du TALN

Interfaces hommes/machines

Pour avoir accès à un service

- bornes de réservations de transports, hébergements, spectacles...
- consultation de pages jaunes, bases de données...
- renseignements téléphoniques
- utilisation/apprentissage d'une langue étrangère
- aide aux handicapés
- commande vocale

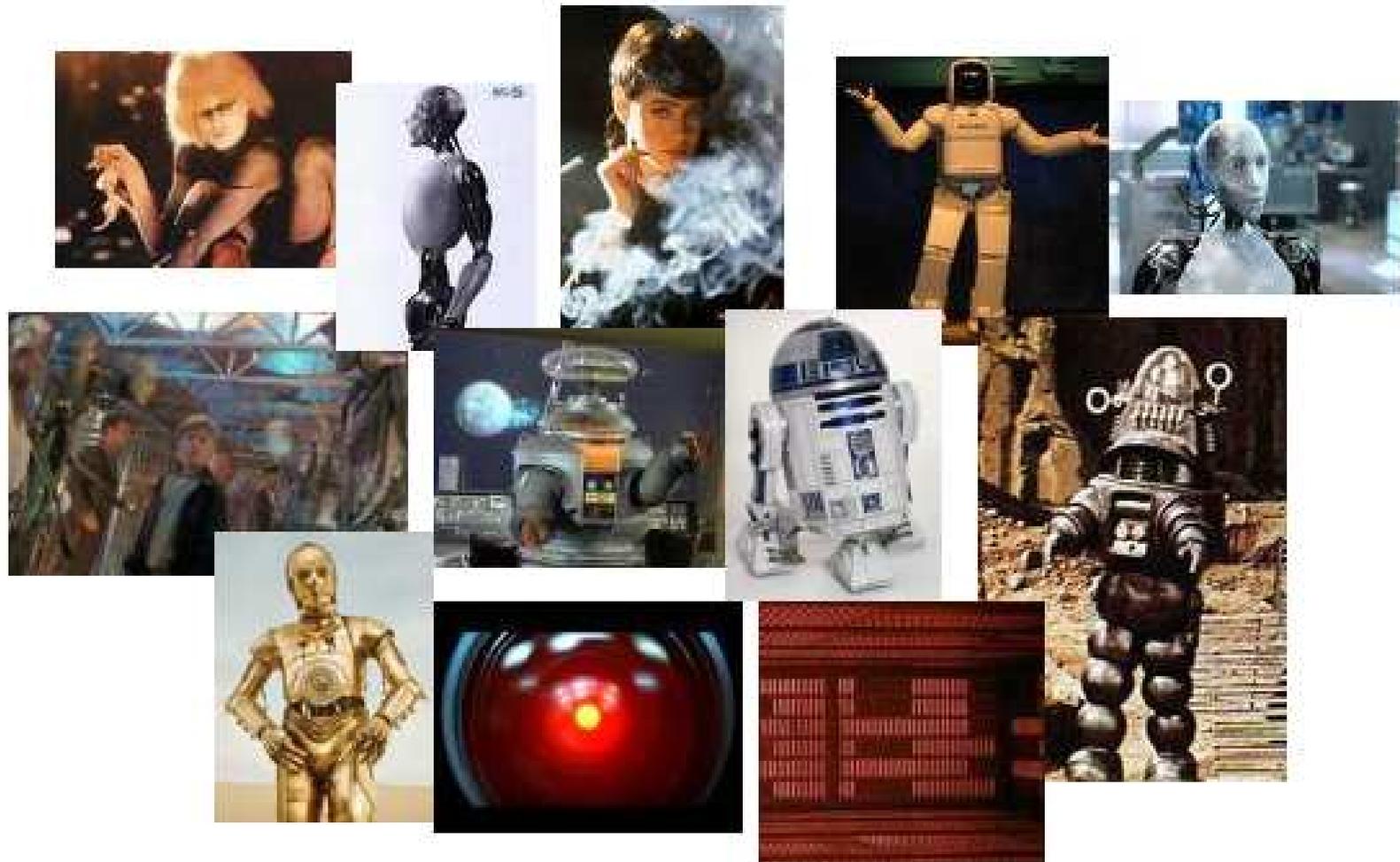
Pour le plaisir

- voitures, "réalité augmentée"
- jeux, robots de compagnie

Test de Turing...

1. Applications et enjeux du TALN

Robots mythiques du cinéma



1. Application et enjeux du TALN

Améliorer les ordinateurs et Internet

Aide à la rédaction/lecture

- dictée/synthèse vocale
- correction orthographique, grammaticale
- traduction automatique, TAO
- résumé automatique

Gestion de documents électroniques

- extraction de mots clés, des entités nommées
- classement assisté (mails, fichiers)
- recherche et extraction d'information, fouille de textes :
 - dans des documents scientifiques (biologie, droit...)
 - dans des documents techniques (documentation, veille)
 - sur Internet (mails clients, blogs, forums, recommandation)

1. Application et enjeux du TALN

Enjeux

Domaines concernés par le TALN

- politique : fichage, surveillance (Echelon), débats citoyens
- économique : les “industries de la langue”, Google !
- scientifique, technologique : domaine R & D très actif
- idéologique : préserver la diversité des langues
- culturel : numériser le patrimoine littéraire
- pédagogique : e-learning, apprentissage des langues étrangères
- social : accès à tous aux machines/ordinateurs (fracture numérique)
- philosophique : les machines peuvent-elles maîtriser le langage ?
- cognitif : aider les individus à trier l’information

2. Qu'est-ce que le langage naturel ?

Propriétés générales

Le langage est le propre de l'homme

- tous les groupes humains découverts à ce jour pratiquent au moins une langue
- il en existe entre 4 000 et 5 000 de par le monde
- l'essentiel d'une langue est acquise vers 5 ans
- les langues parlées par les humains sont les langues naturelles
- langage = capacité de langue

Objet de la linguistique ou "sciences du langage"

- étudier rigoureusement les langues telles qu'elles se parlent
- recherche de points communs entre différentes langues, universaux

2. Qu'est-ce que le langage naturel ?

Universaux du LN

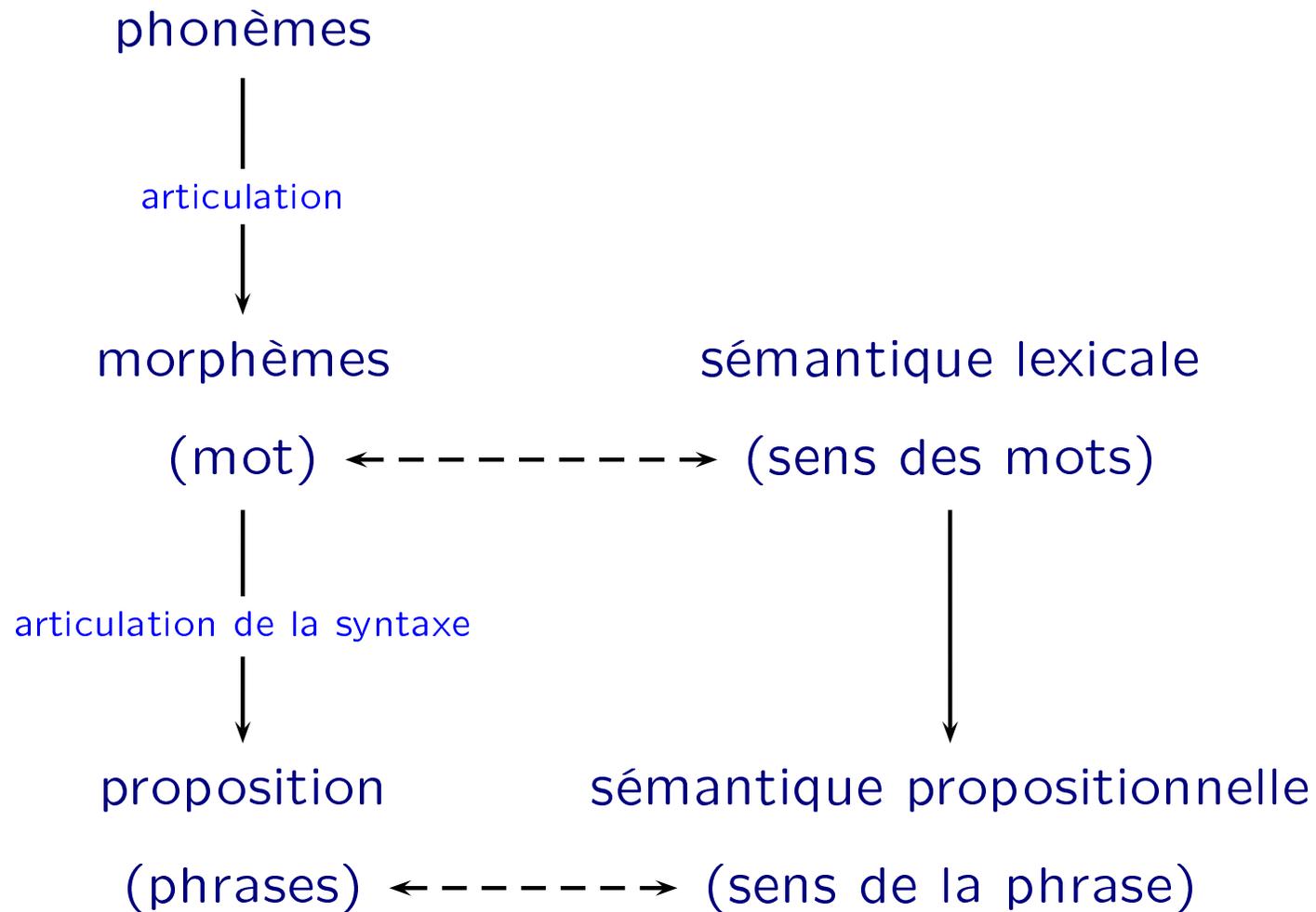
La double articulation

- les langues naturelles sont composées d'unités discrètes
 - phonèmes : sons élémentaires sans significations
 - morphèmes : unités qui ont du sens (mots)
 - propositions : énoncés vrais ou faux (phrases)
- pour passer d'un niveau à un autre : règles de combinaisons
- vrai pour toutes les langues naturelles (y compris langues des signes)
- aucun autre système naturel (communications animales) n'a l'ensemble de ces propriétés
- et en plus (par rapport aux langages de programmation) : grande expressivité du sens !

2. Qu'est-ce que le langage naturel ? Les niveaux d'analyse

niveaux d'analyse

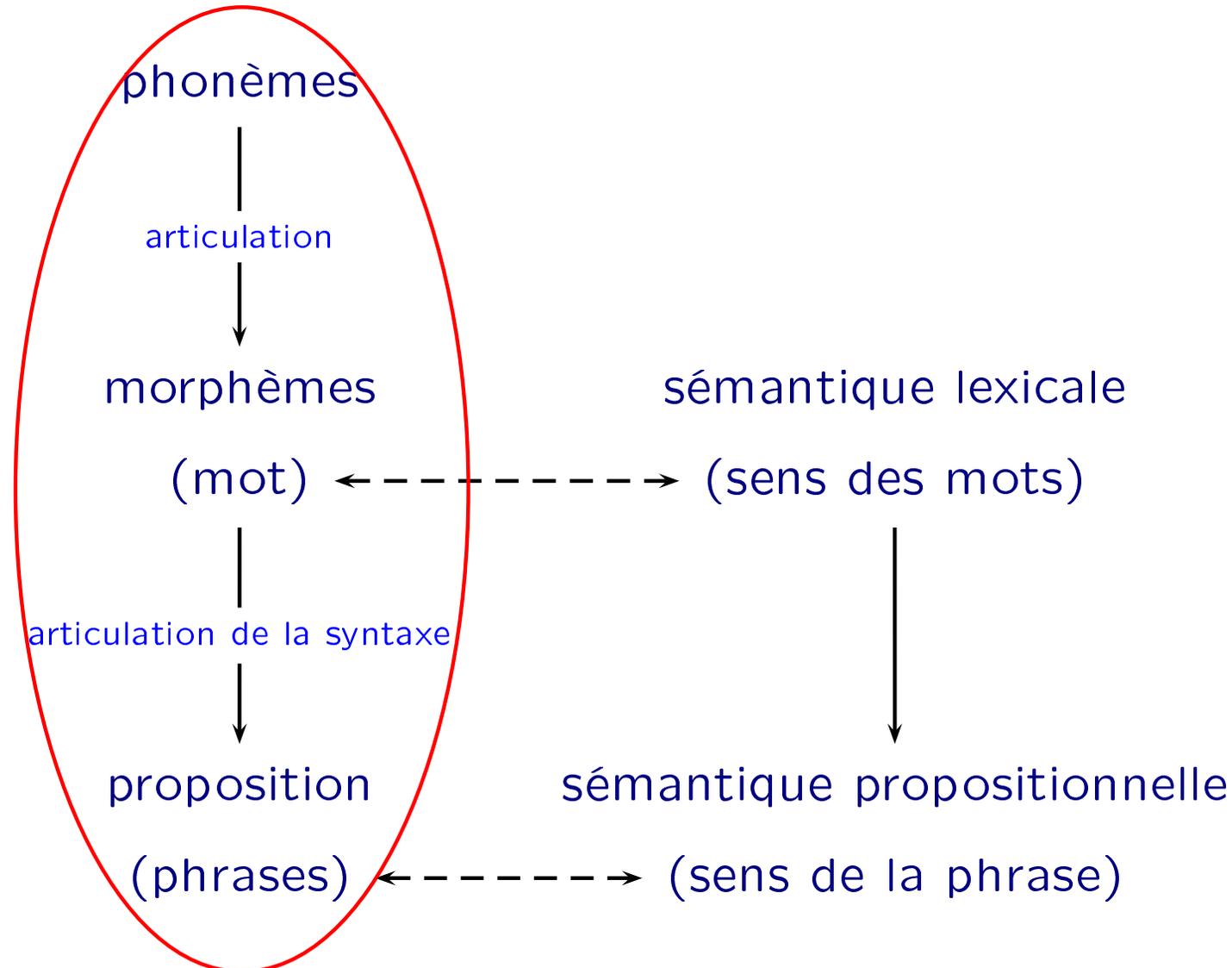
sémantique associée



2. Qu'est-ce que le langage naturel ? La double articulation

niveaux d'analyse

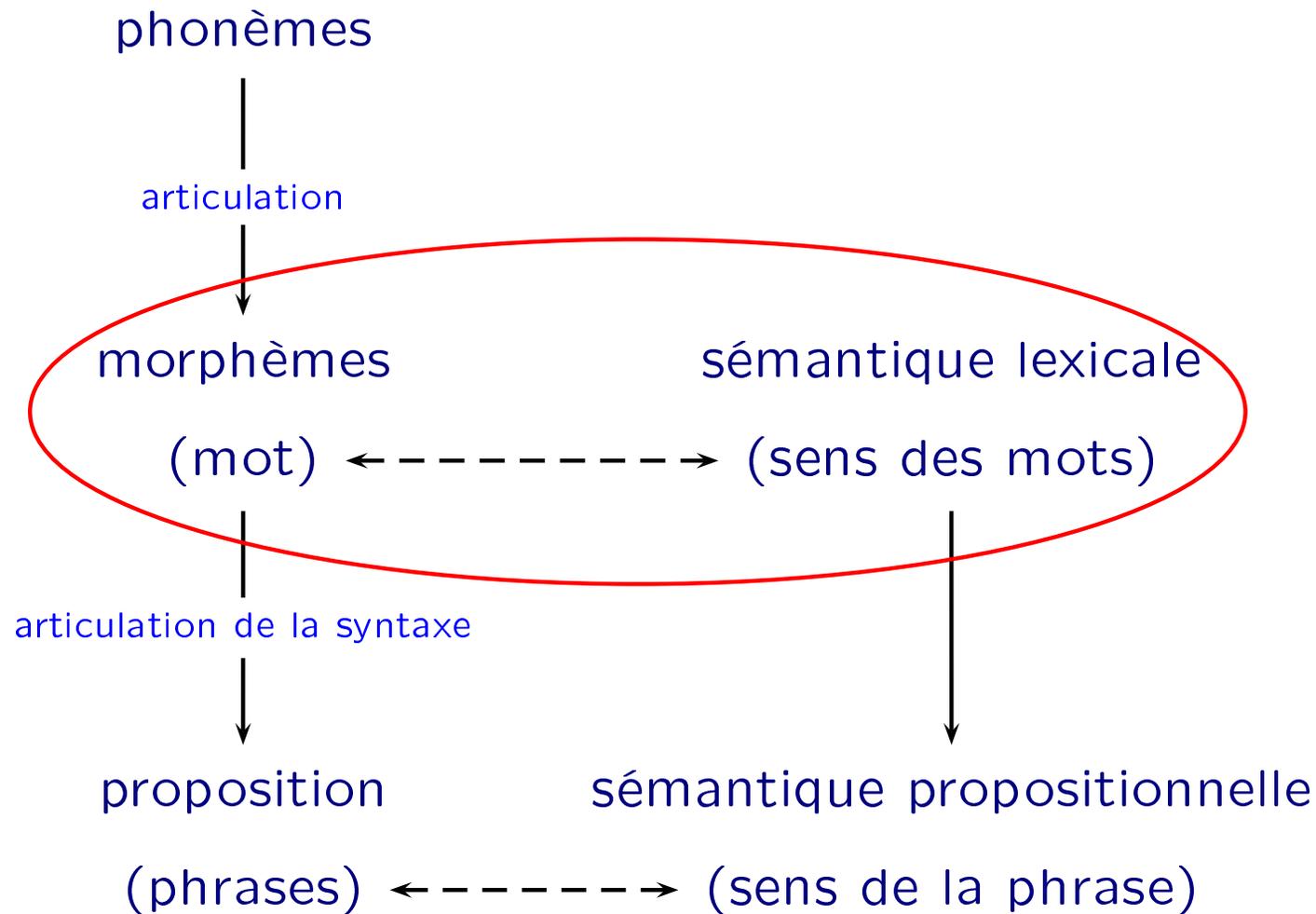
sémantique associée



2. Qu'est-ce que le langage naturel ? Le niveau des "protolangages"

niveaux d'analyse

sémantique associée



3. Du LN au TA

Modélisation informatique

Comment informatiser un domaine ?

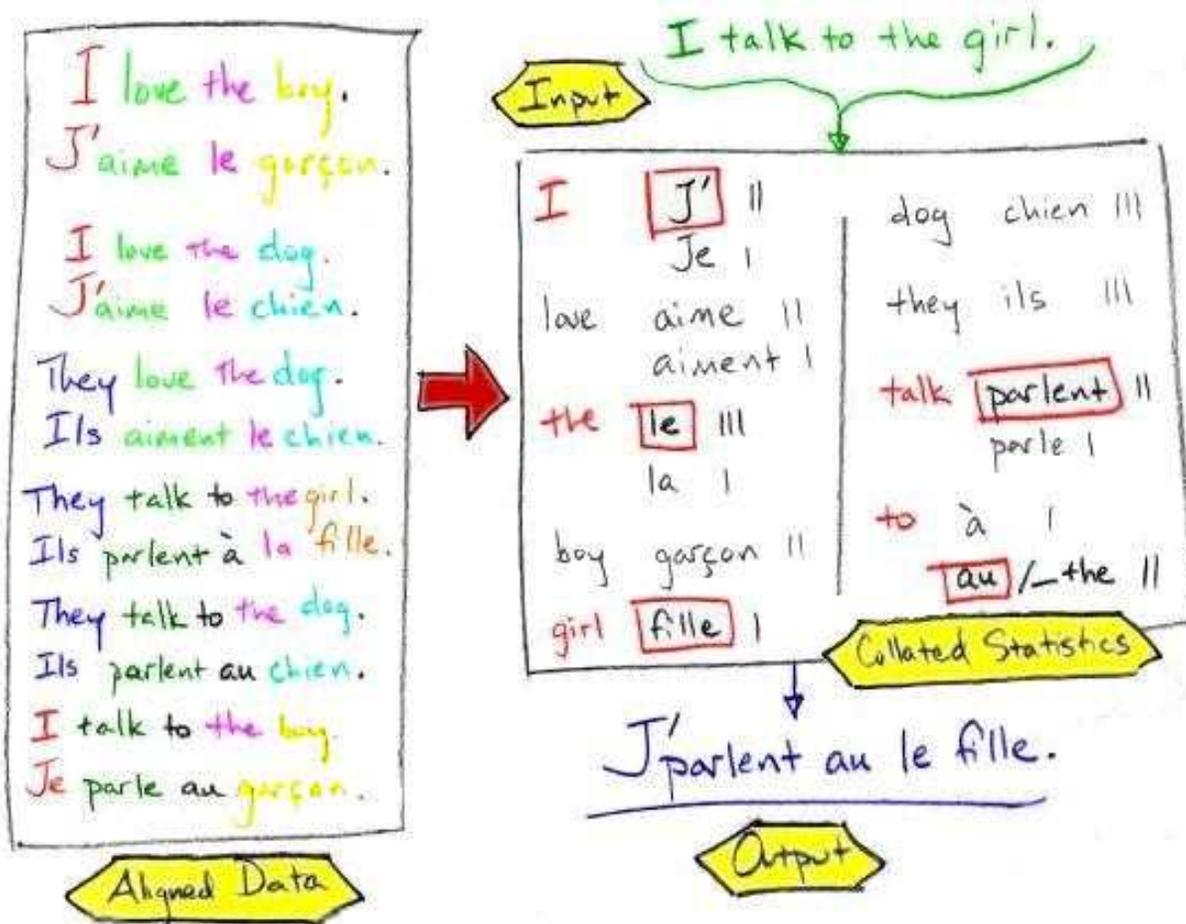
- les données du domaine doivent être codées dans l'ordinateur
- les opérations de manipulations, combinaisons... de ces données doivent être codées sous la forme de programmes

Ingénierie linguistique : deux approches possibles (comme en IA)

- une approche fondée sur le traitement statistique de données observées
- une approche fondée sur un modèle fonctionnel de l'esprit humain

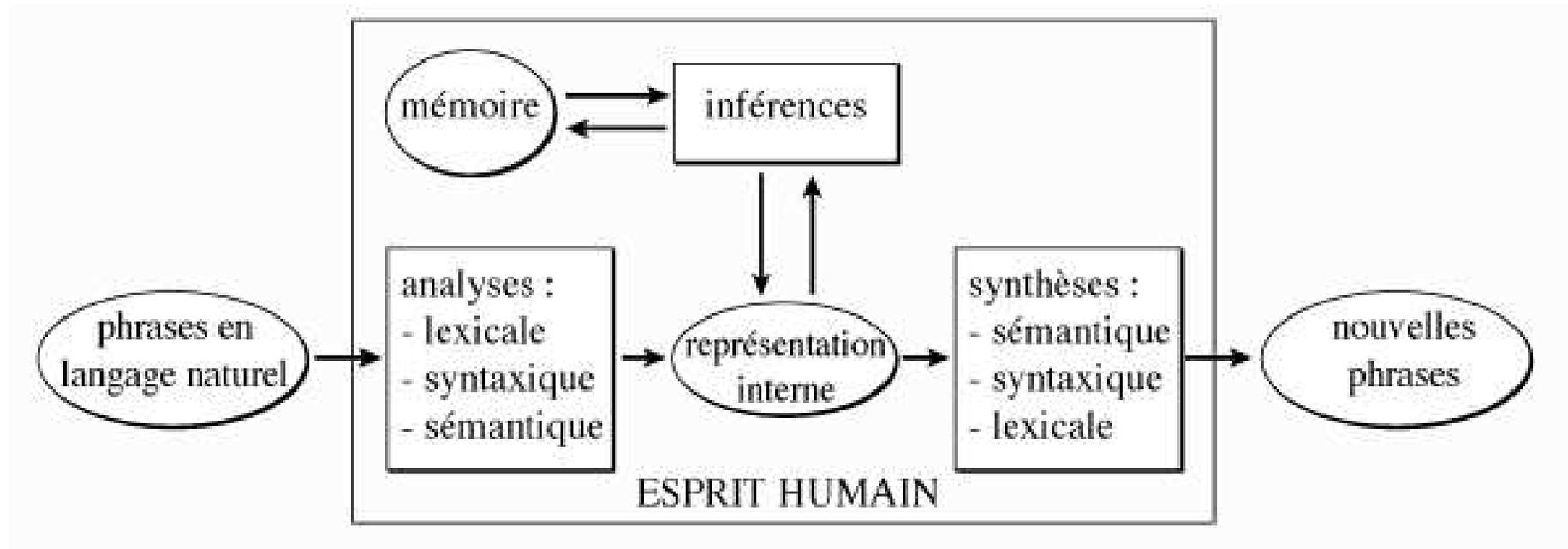
3. Du LN au TA

Traduction : approche fondée sur des données



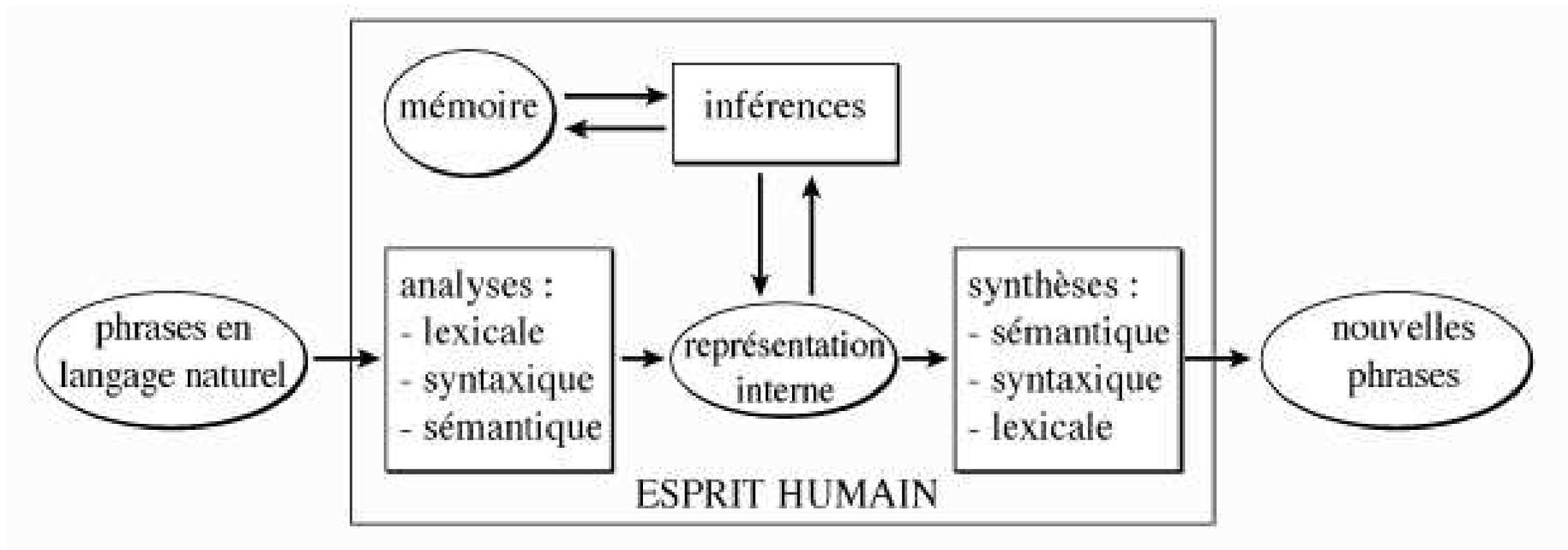
3. Du LN au TA

Traduction : approche fondée sur un modèle



3. Du LN au TA

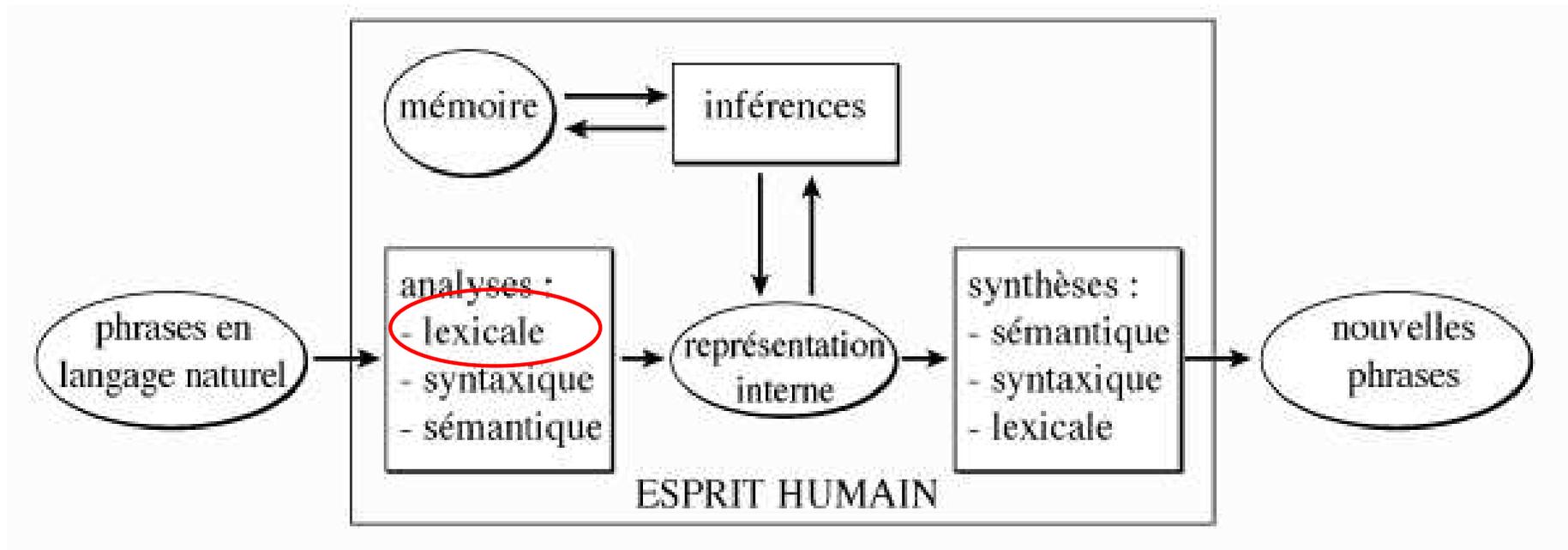
Traduction : approche fondée sur un modèle



un chat noir dort

3. Du LN au TA

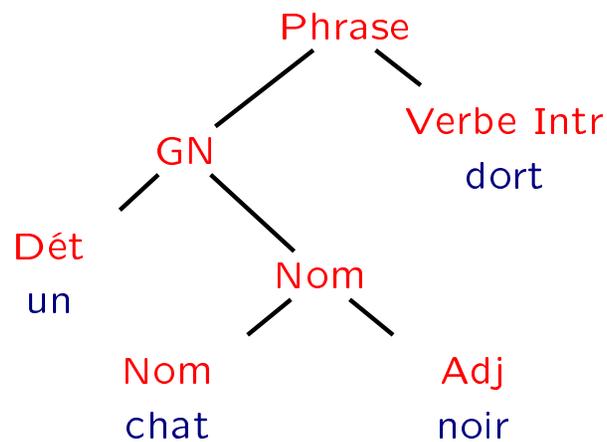
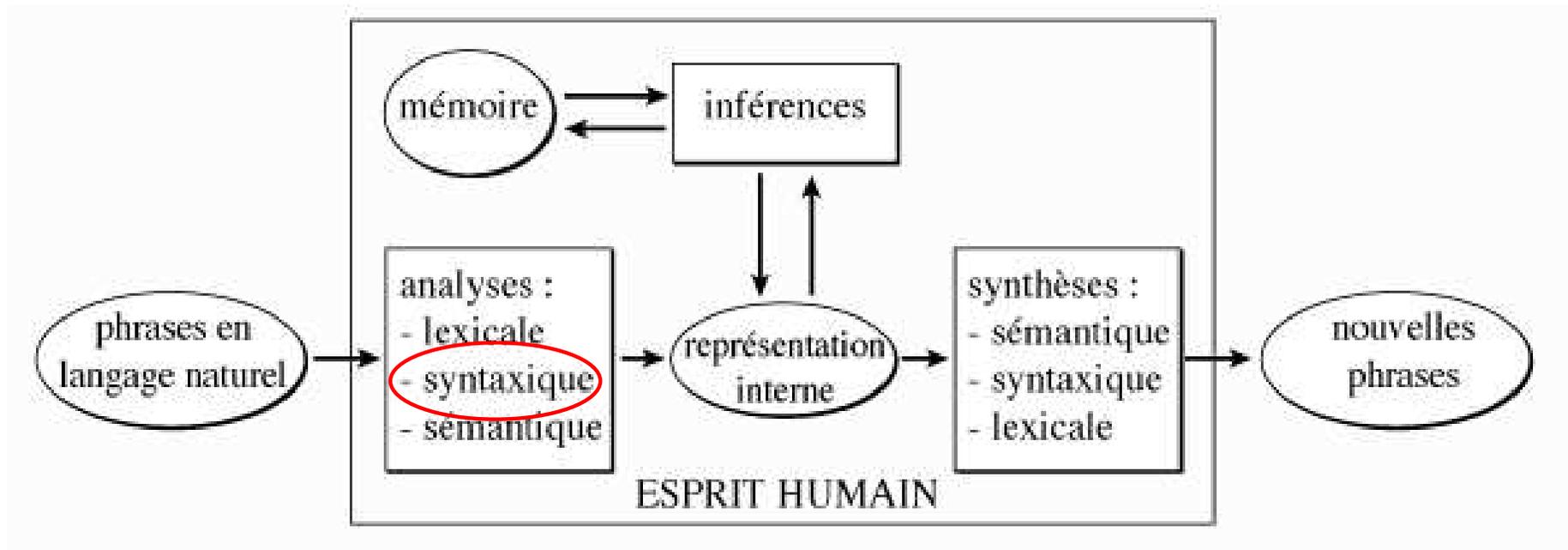
Traduction : approche fondée sur un modèle



Dét Nom Adj Verbe Intr
un chat noir dort

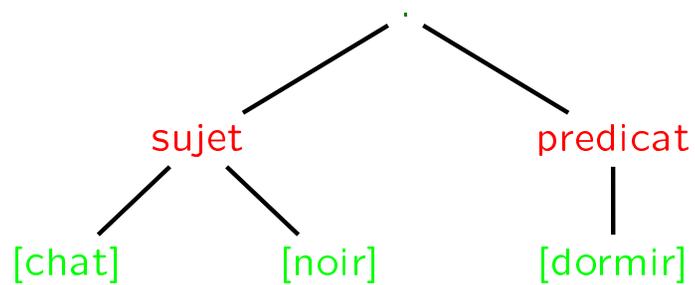
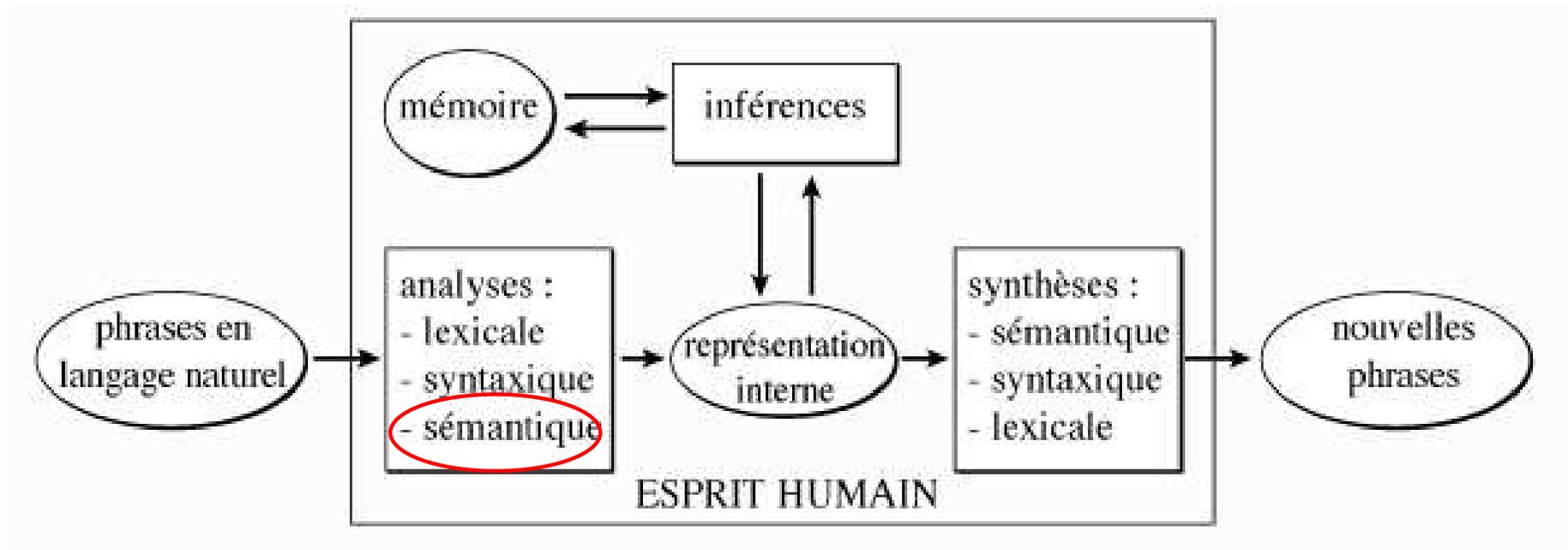
3. Du LN au TA

Traduction : approche fondée sur un modèle



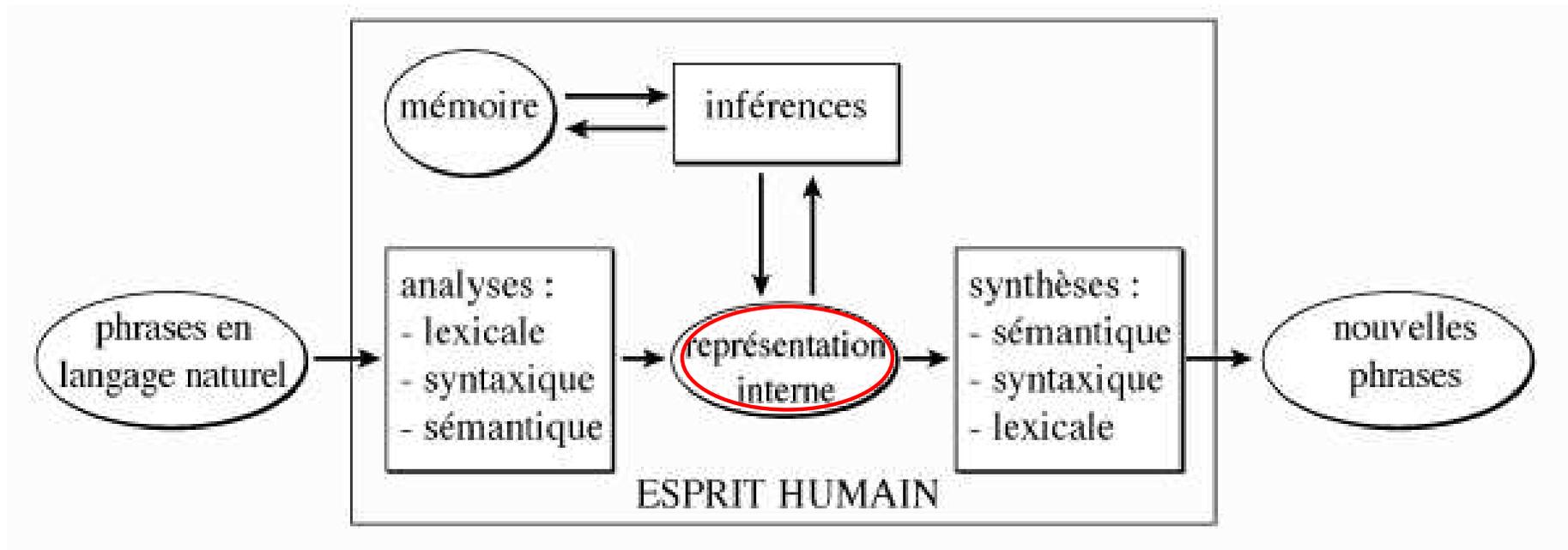
3. Du LN au TA

Traduction : approche fondée sur un modèle



3. Du LN au TA

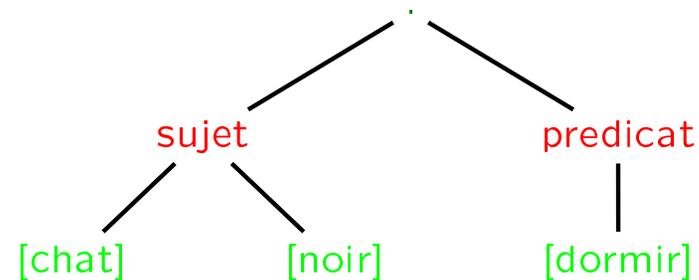
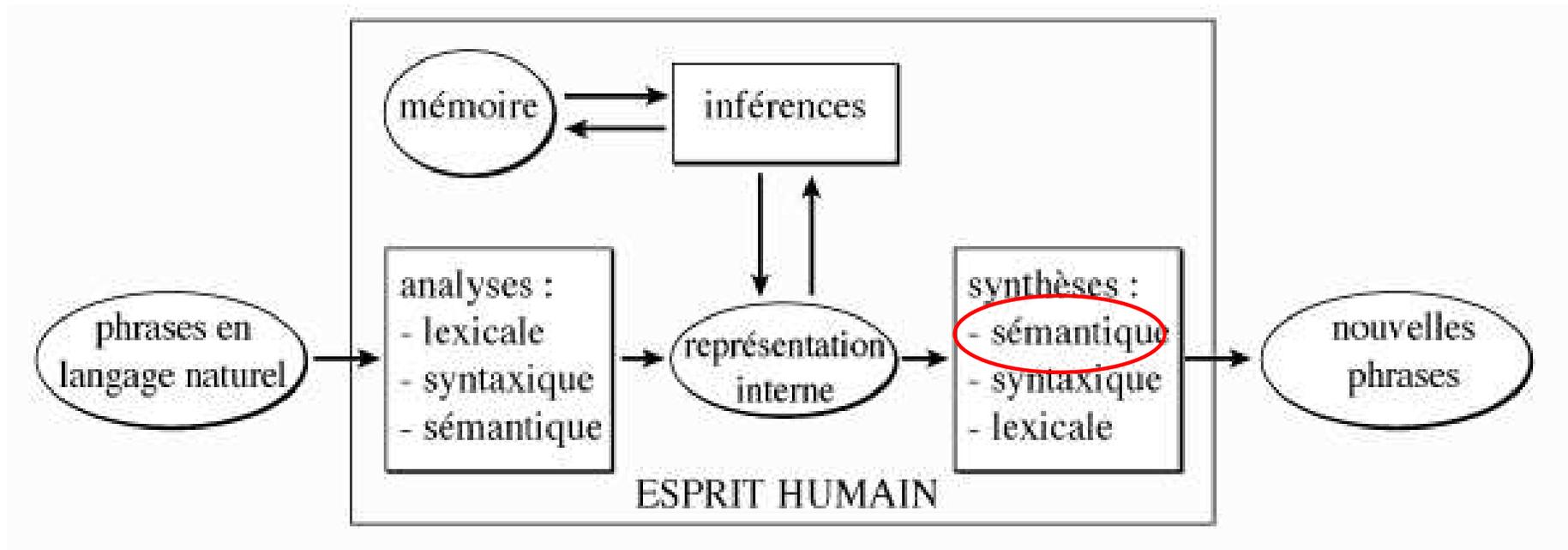
Traduction : approche fondée sur un modèle



$$\exists x[[chat(x) \wedge noir(x)] \wedge dormir(x)]$$

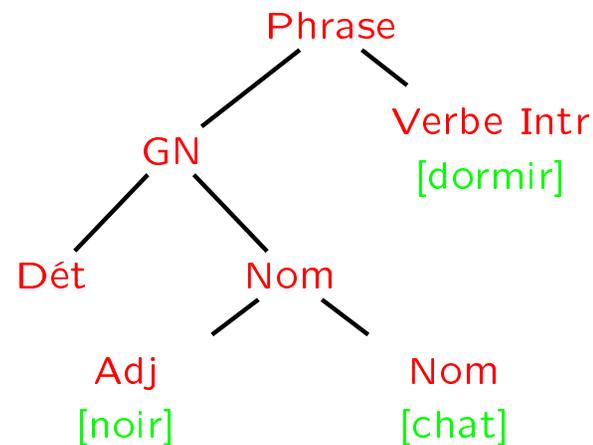
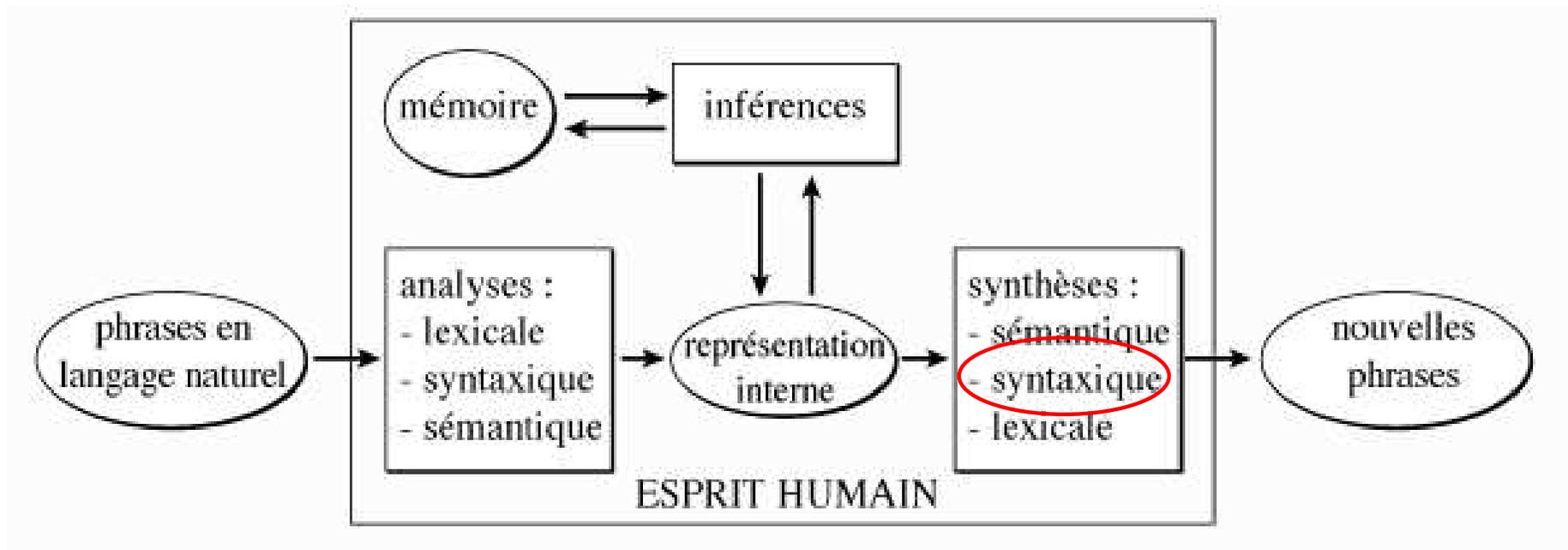
3. Du LN au TA

Traduction : approche fondée sur un modèle



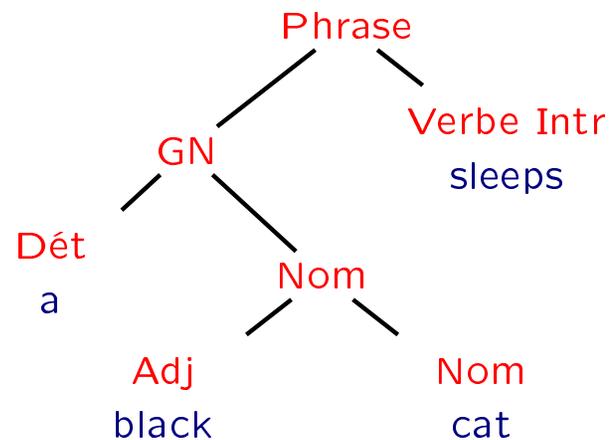
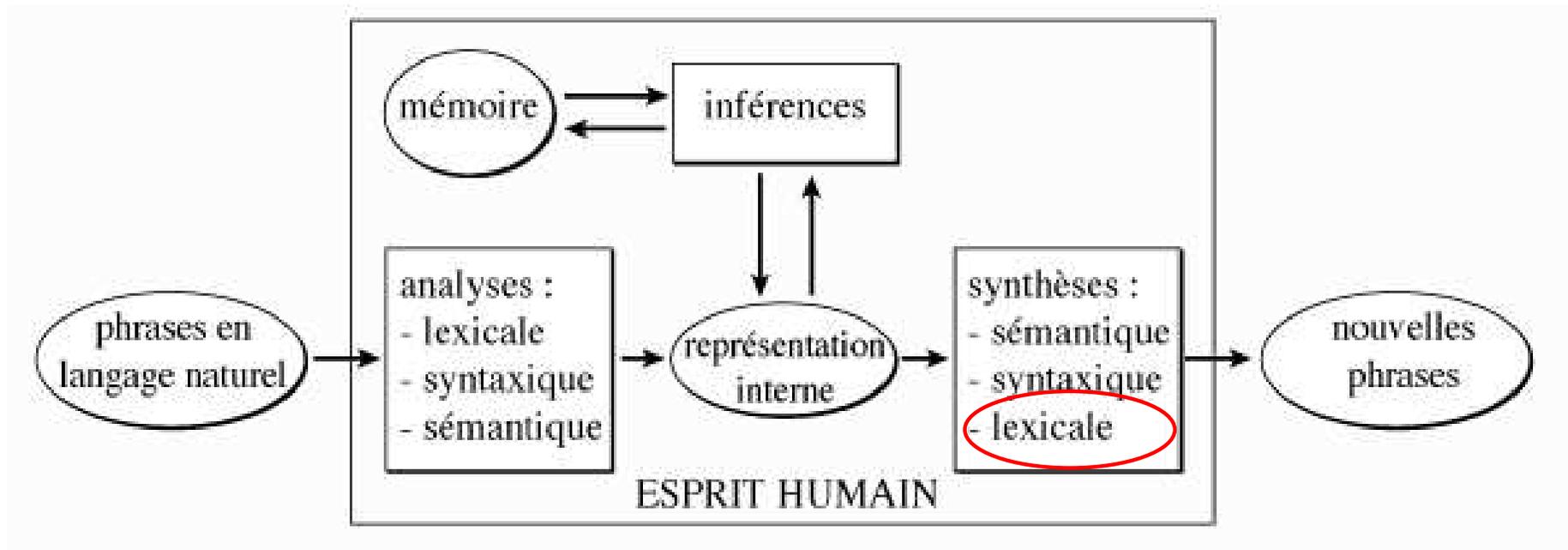
3. Du LN au TA

Traduction : approche fondée sur un modèle



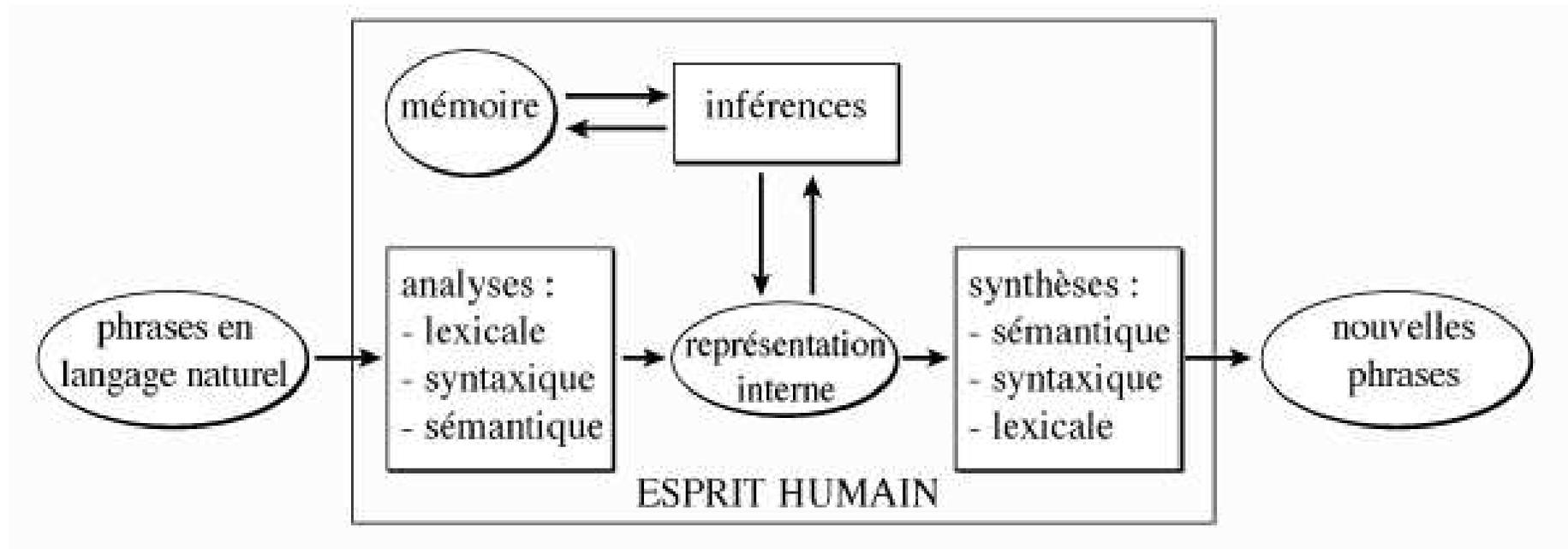
3. Du LN au TA

Traduction : approche fondée sur un modèle



3. Du LN au TA

Traduction : approche fondée sur un modèle



a black cat sleeps

3. Du LN au TA

Qu'est-ce qui marche le mieux ?

Approche par données

- pas de compréhension réelle
- très efficace pour gestion des documents électroniques (résumé, classement, recherche et extraction d'information...)
- nécessite beaucoup de données initiales
- rendue possible depuis 10-15 ans par l'augmentation des capacités de stockage/calcul des machines

Approche par modèle

- simule mieux la compréhension (pour les robots intelligents ?)
- mais aucun traitement intermédiaire réellement résolu

En pratique : mélange des deux...