

# LIDIA-1 : UNE PREMIÈRE MAQUETTE VERS LA TA INTERACTIVE "POUR TOUS"

**HERVÉ BLANCHON**

GETA – IMAG

(UJF & CNRS)

BP 53

38041 GRENOBLE Cedex 9

[Herve.Blanchon@imag.fr](mailto:Herve.Blanchon@imag.fr)

# Plan de l'exposé

- ✦ Rappel : le contexte du travail
- ✦ Visite guidée
- ✦ Zoom sur la clarification
- ✦ Architecture & implémentation
- ✦ Perspectives du passage à un prototype

# TAO du Rédacteur

## Objectifs & *problèmes*

- ✦ Rédiger dans sa propre langue et transmettre son texte
  - sans révision
  - sans compréhension explicite
    - pas de base de connaissances du domaine de référence
  - sans langage “contrôlé”
  - sans sous-langage déterminé
    - *toutes les ambiguïtés sont a priori possibles*
    - *pas d'heuristiques “sous-langagières” suffisantes*

# TAO du rédacteur

## Méthode : TAFD

(TAO Fondée sur le Dialogue avec l'auteur)

### ✦ Dialogue de standardisation

- styles d'énoncés
- genres de textes
- types de documents

### ✦ Dialogue de clarification

- ambiguïtés lexicales, syntaxiques, sémantiques, pragmatiques
- ambiguïtés propres à la langue de départ, ou contrastives
- réduction possible par guidage partiel a priori (marques)

### ✦ “Langage guidé” → collection de sous-langages

# Défis de la “TAFD pour tous”

## Aspects linguiciels

- ✦ **COUVERTURES GRAMMATICALE ET LEXICALE**  
1 ou 2 ordres de grandeur de plus qu’en TA du réviseur
  - Conception et ingénierie d’une très grande BDLM
  - Formalismes grammaticaux permettant de décrire une collection de sous-langages paramétrés
- ✦ **DÉCOUVRABILITÉ** des bases de connaissances linguistiques et grammaticales par l’utilisateur non spécialiste

# Défis de la “TAFD pour tous”

## Aspects ergonomiques

- ✦ ASYNCHRONIE pour permettre l'utilisation de micros, avec connexion à un serveur pour les tâches lourdes
  - analogie avec les serveurs de courrier-E (internet, minitel)
- ✦ Organisation des DIALOGUES (paraphrase, rephrasage)
- ✦ NON-PRÉEMPTIVITÉ des questions du système
- ✦ Utilisation du MULTIMODAL/MULTIMEDIA pour la désambiguïsation interactive

# Enjeux de la TAFD

## Les “niches” de la TAFD

### ✦ Critères de choix

- Qualité élevée, révision impossible
- Contexte fortement multilingue :
  - 1 → n      Dissémination de documents
  - n → n      Projets internationaux
- Domaines et types de texte ouverts et variés
  - non contraints
  - non contrôlés
- Utilisateur désireux de participer

# Enjeux de la TAFD

## Les “niches” de la TAFD

### ✦ Situations traditionnelles

- Traduction de volumes relativement faibles de doc. technique en plusieurs langues (5000 à 8000 p. sur DON)
- Diffusion d'information dans plusieurs langues
  - (circulation routière, météo, congrès, manifestations sportives, situations d'urgence...)
- Échange télématique de notes et documents de travail dans des contextes internationaux
  - multinationales, projets européens
- Préparation de bandes vidéo multilingues

# La maquette LIDIA-1

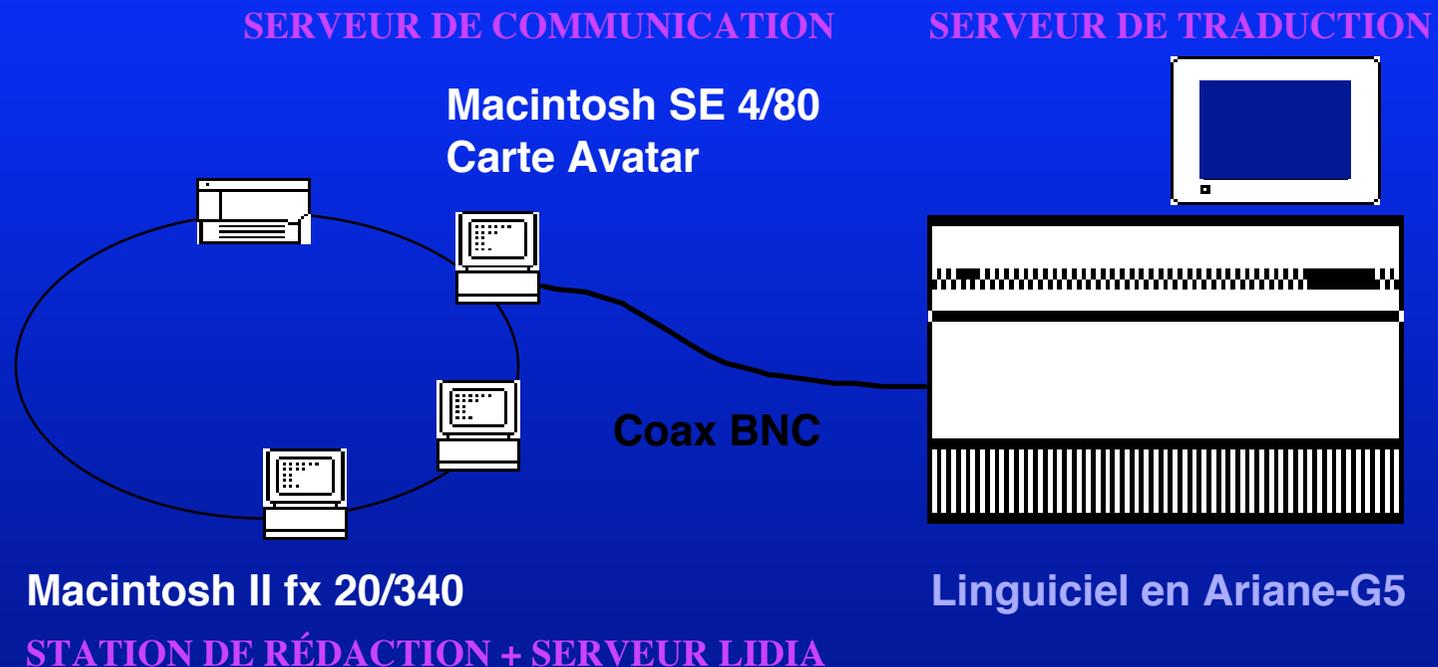
## Situation choisie pour cette maquette (1/2)

- ✦ Une langue source, plusieurs langues cibles
  - français → allemand  
russe  
anglais
- ✦ Traduction d'une pile HyperCard qui présente
  - des ambiguïtés en contexte
  - l'utilité de la désambiguïstation interactive

# La maquette LIDIA-1

## Situation choisie pour cette maquette (2/2)

### ✦ Architecture distribuée



### ✦ Traitements asynchrones

# LIDIA-1 : démonstration

- ✦ Session de travail de l'utilisateur
  - Appel aux traitements
  - Étape de clarification
  - Voir les annotations
  - Voir la rétrotraduction
  - Construire une pile en langue cible

# Démonstration

- ◆ Intégration à l'environnement de l'auteur

Preferences Home

Your Name: Hervé Blanchon

Click the user level you want:

<b>5 Scripting</b>	Edit scripts of buttons, fields, cards, backgrounds, and stacks.
<b>4 Authoring</b>	Create buttons and fields. Link buttons to cards and stacks.
<b>3 Painting</b>	Use the Paint tools to change the appearance of cards and backgrounds.
<b>2 Typing</b>	Enter and edit text in fields.
<b>1 Browsing</b>	Explore stacks but make no changes.

Other settings:

Blind Typing

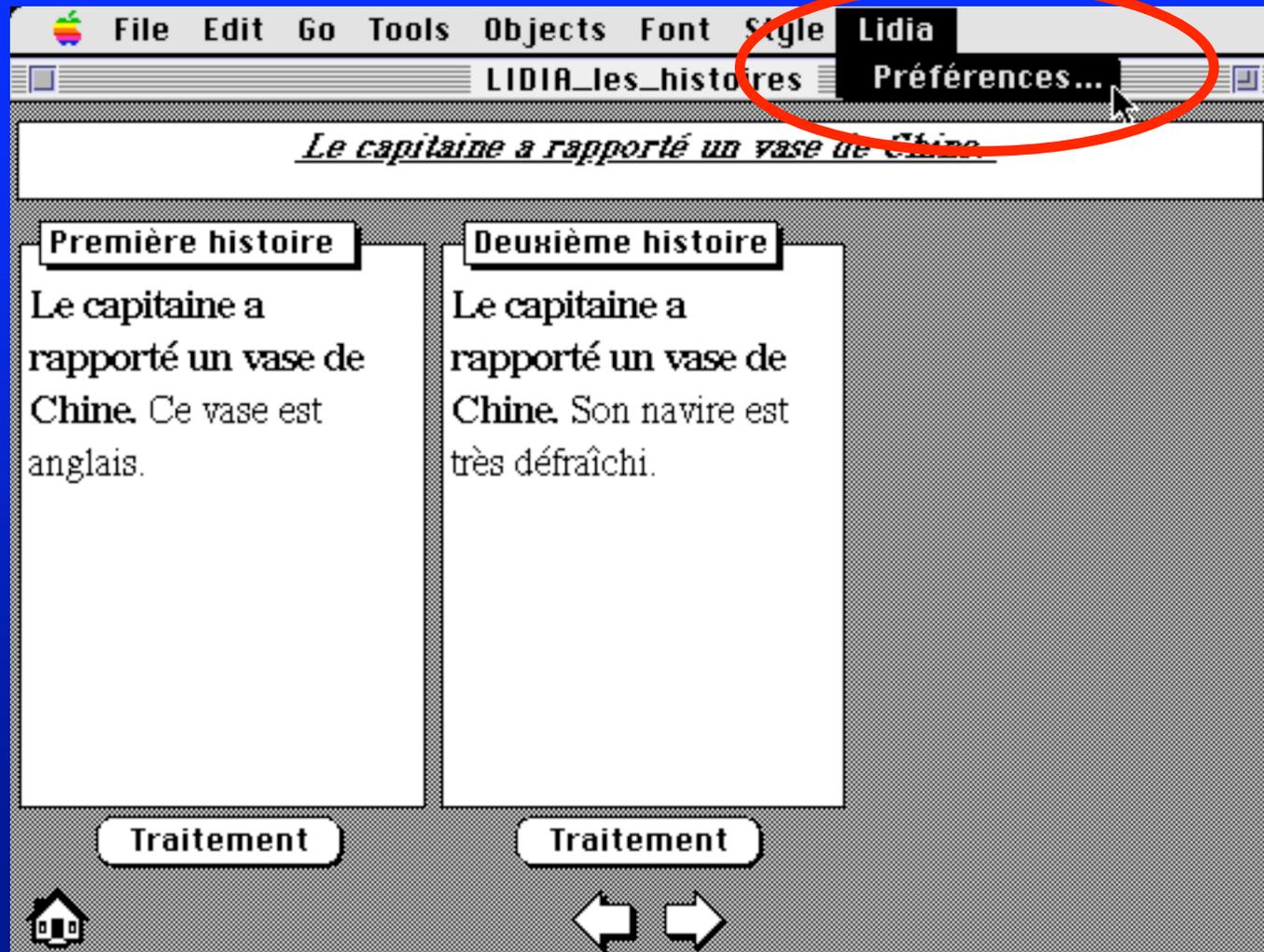
Power Keys

Arrow Keys in Text

Utiliser LIDIA

# Démonstration

- ✦ Une carte d'histoires de la pile de démonstration



# Démonstration

- ✦ Carte de traitement & palette LIDIA, sélection d'un objet



# Démonstration

- ✦ Objet en cours de traitement, voir l'état d'avancement



# Démonstration

## ✦ État d'avancement des traitements



# Démonstration

- ✦ Objet en cours de traitement, demande de clarification



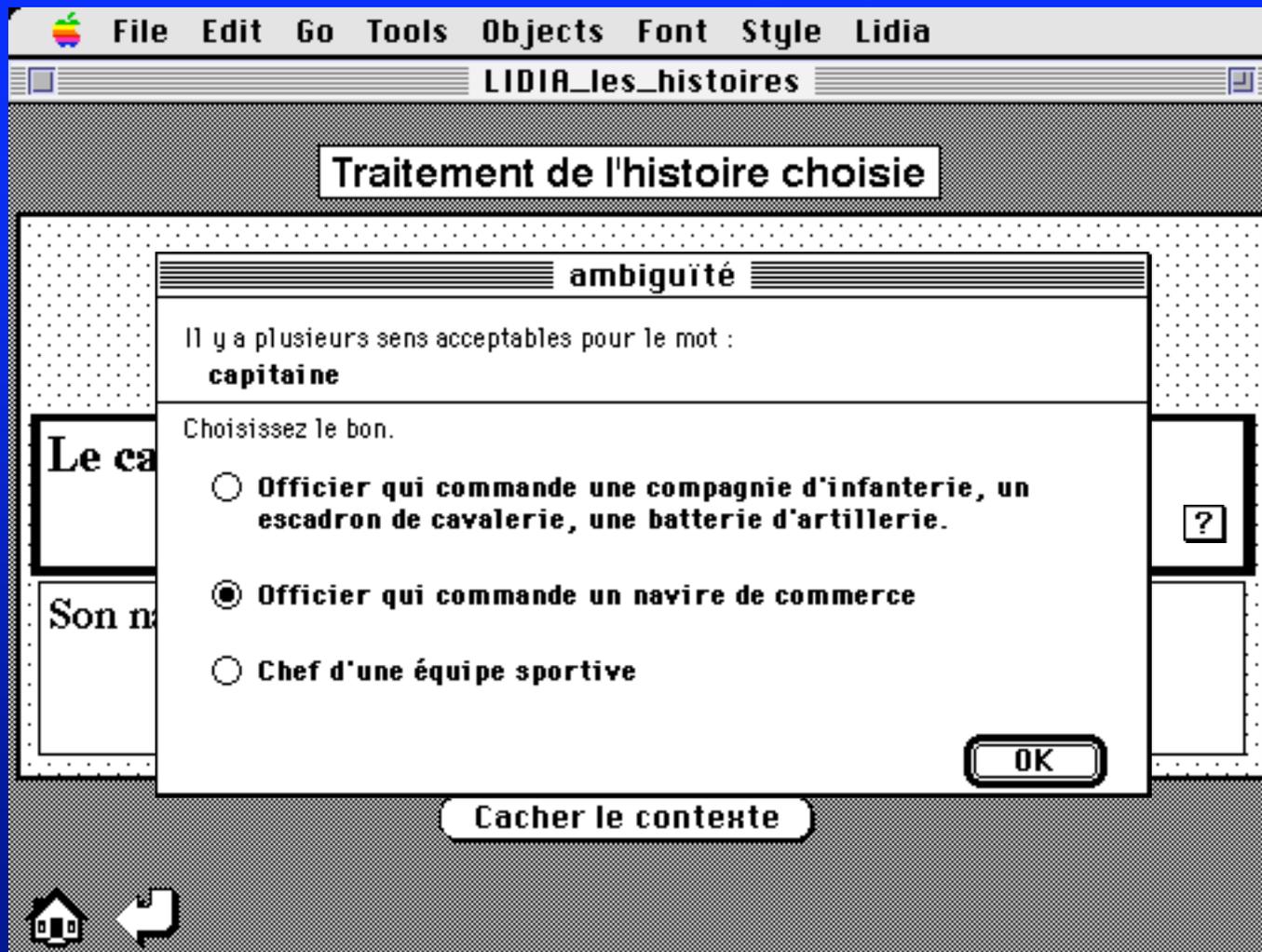
# Démonstration

- ✦ Session de clarification : ambiguïté de géométrie



# Démonstration

## ✦ Session de clarification : polysémie



# Démonstration

- ✦ Objet poursuivant son traitement, voir les annotations



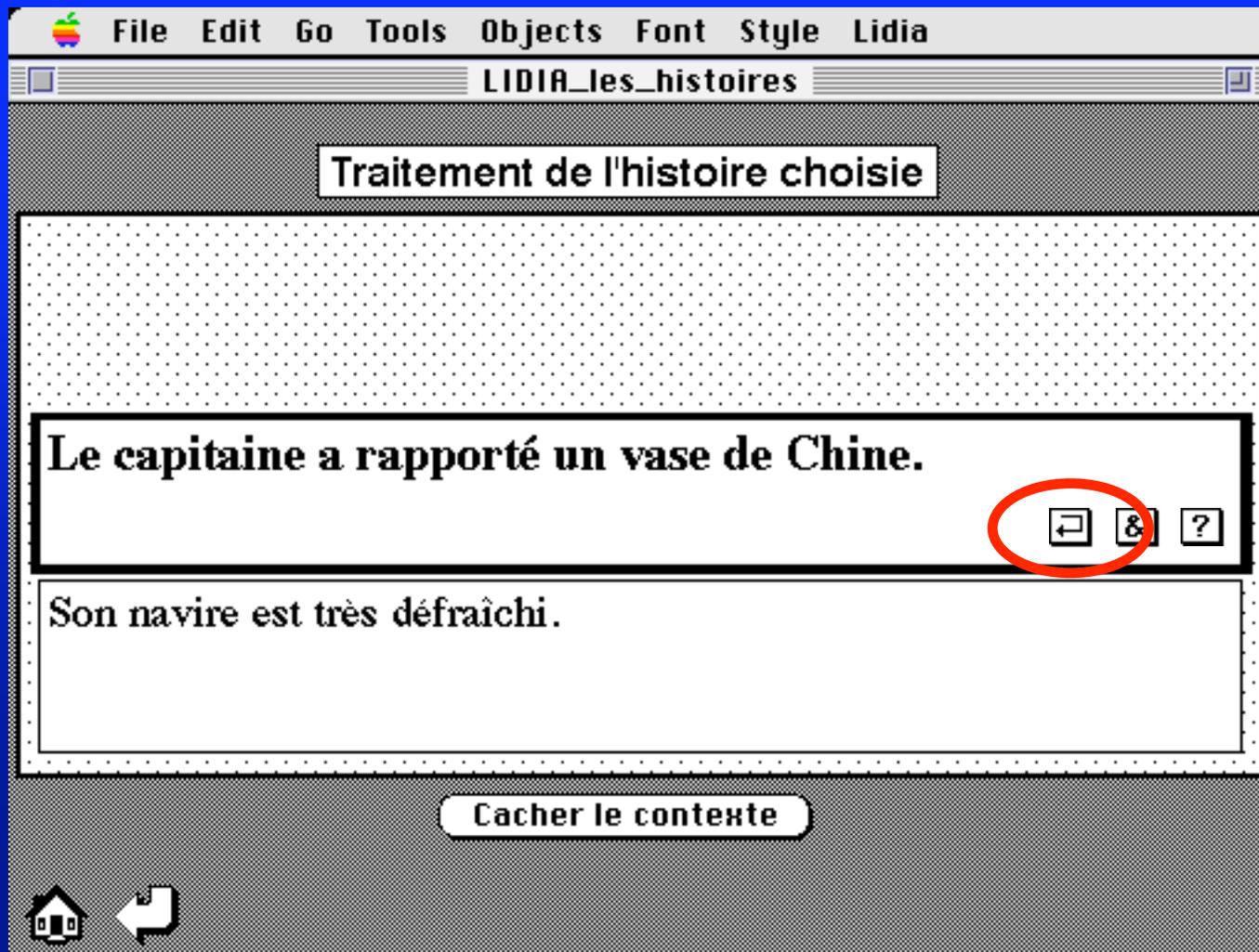
# Démonstration

- ◆ Annotations : classes syntaxiques, groupes syntagmatiques



# Démonstration

- ✦ Objet en fin de traitement, voir la rétrotraduction



# Démonstration

## ✦ Rétrotraduction



# Démonstration

## ✦ Production de la pile traduite

The screenshot shows a window titled "LIDIA\_les\_histoires.All". At the top, the text "*Le capitaine a rapporté un vase de Chine.*" is displayed. Below this, there are two text boxes. The left box is titled "Erste Geschichte" and contains the text: "DerHauptmann hat eine Vase aus China mitgebracht. Die Vase ist englisch." The right box is titled "Zweite Geschichte" and contains the text: "Der Kapitän hat eine chinesische Vase mitgebracht. Sein Boot ist sehr verblasst." The right box is highlighted with a red border. At the bottom of the window, there are two buttons labeled "Behandlung" and a set of navigation arrows (left and right) with a home icon on the far left.

LIDIA\_les\_histoires.All

*Le capitaine a rapporté un vase de Chine.*

**Erste Geschichte**

DerHauptmann hat eine Vase aus China mitgebracht. Die Vase ist englisch.

**Zweite Geschichte**

Der Kapitän hat eine chinesische Vase mitgebracht. Sein Boot ist sehr verblasst.

Behandlung

Behandlung

Home icon, Left arrow, Right arrow

# LIDIA-1 : clarification

- ✦ Organisation du processus de clarification
  - Classes d'ambiguïtés retenues, définies
  - Stratégie
  - Méthodologie
  - Exemple

# Clarification

## Types d'ambiguïtés (1/5)

### ✦ Ambiguïté de classe syntaxique

#### – Coordination verbale

– Il observe la photo et **la classe**.

- Il observe **la classe**.
- Il **classe la photo**.

#### – Classe simple

– Cette petite **brise la glace**.

- Ce petit **vent la frigorifie**.
- Cette petite (fille) **casse la glace**.

# Clarification

## Types d'ambiguïtés (2/5)

### ✦ Ambiguïté de géométrie

#### – Coordination nominale

- L'évolution de la structure du réseau **et des investissements** est étudiée.
  - évolution de la structure des investissements
  - évolution des investissements

#### – Coordination adjectivale

- Il prend des classeurs et des cahiers **noirs**.
  - des classeurs **noirs** et des cahiers **noirs**.
  - des cahiers **noirs** et des classeurs.

# Clarification

## Types d'ambiguïtés (3/5)

### ✦ Ambiguïté de géométrie

#### – Subordination

- Le capitaine a rapporté un vase **de Chine**.
  - **De Chine**, la capitaine a rapporté un vase.
  - Le capitaine a rapporté un vase **chinois**.

#### – Structure argumentaire du verbe

- Il parle **de** l'école **de** cuisine.
  - Il parle **depuis** l'école **à propos de** cuisine.
  - Il parle **depuis** (l'école **de** cuisine).
  - Il parle **à propos de** (l'école **de** cuisine).

# Clarification

## Types d'ambiguïtés (4/5)

### ✦ Ambiguïté de décoration

#### – Étiquetage syntaxique

- Il parle **de** la Tour Eiffel.
  - Il parle **depuis** la Tour Eiffel.
  - Il parle **à propos de** la Tour Eiffel.

#### – Étiquetage logico-sémantique

- Pierre a fait porter des chocolats **à** Lucie.
  - Pierre a fait porter des chocolats **pour** Lucie.
  - Pierre a fait porter des chocolats **par** Lucie.

# Clarification

## Types d'ambiguïtés (5/5)

### ✦ Ambiguïté de décoration

- Ordre des arguments d'un verbe transitif
  - Quel auteur cite ce conférencier ?
    - Un auteur cite un conférencier.
    - Un conférencier cite un auteur.

# Clarification

## Stratégie

- ✦ 1 Trouver les bons groupes simples
  - Ambiguïté de classe syntaxique
- ✦ 2 Trouver la construction des verbes
  - Structure argumentaire du verbe
  - Ambiguïtés de décoration
- ✦ 3 Trouver la bonne géométrie
  - Ambiguïté de subordination
  - Ambiguïté de coordination nominale, adjectivale
- ✦ 4 Trouver le sens des occurrences

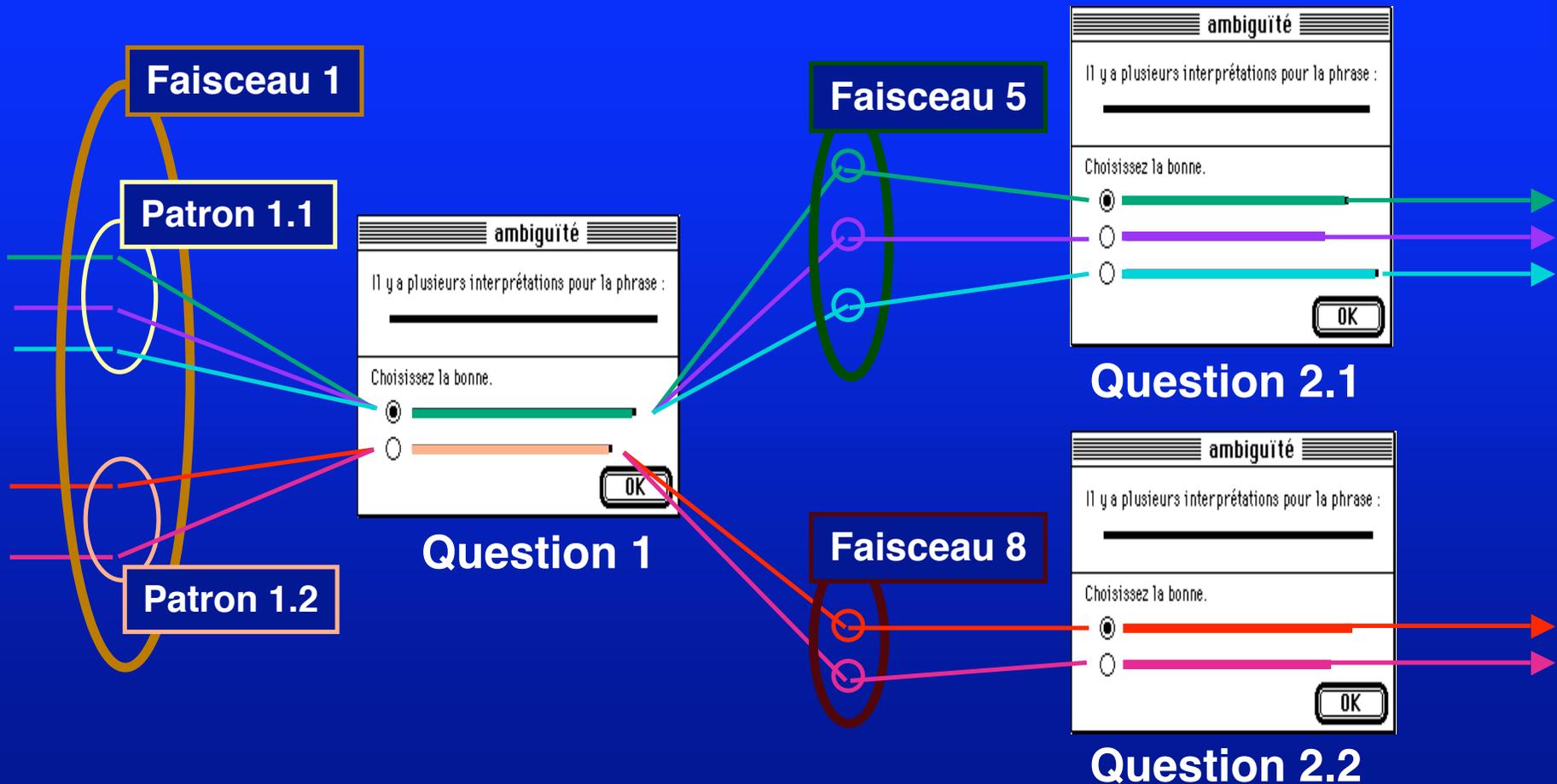
# Clarification

## Méthodologie

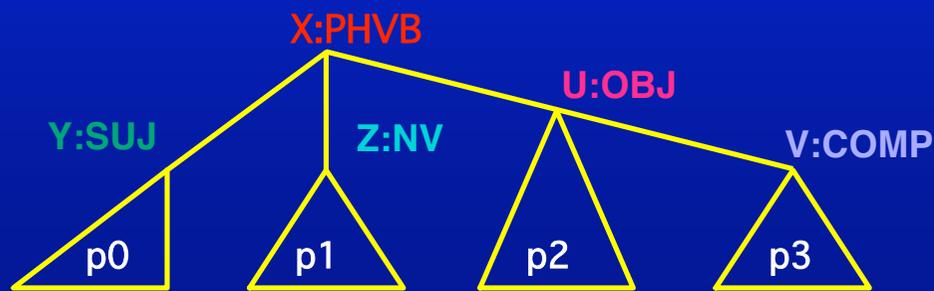
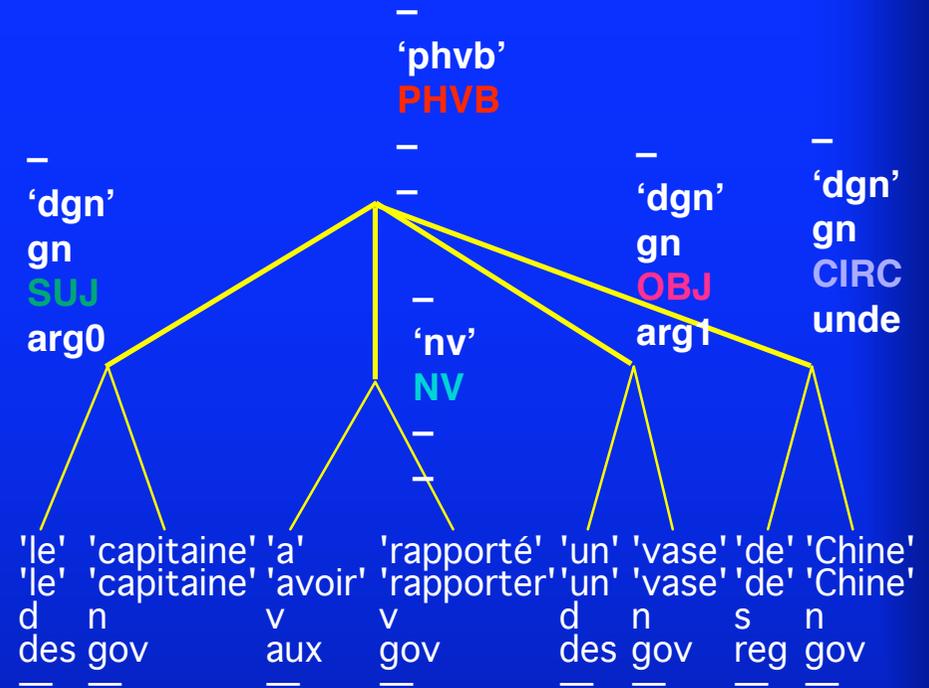
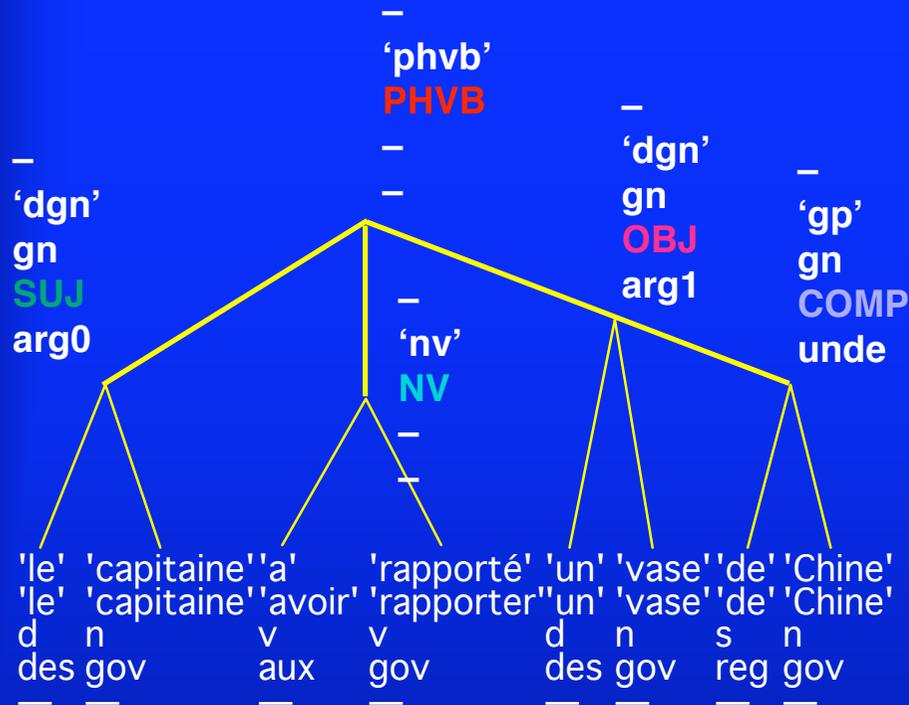
- ✦ Reconnaissance de patrons
  - Plusieurs ensembles (faisceaux) de patrons associés à chaque classe d'ambiguïté
  - Un patron décrit un schéma d'arbre
- ✦ Production du dialogue
  - Une méthode de production associée à chacun des patrons
  - Chaque méthode est décrite au moyen d'un jeu d'opérateurs de base
- ✦ Production d'un arbre de questions
  - Le processus de désambiguïsation produit un arbre de questions

# Clarification

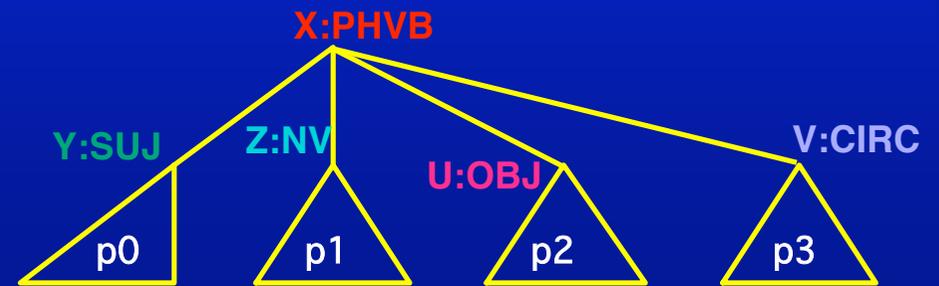
## Méthodologie



# Exemple: analyses & patrons



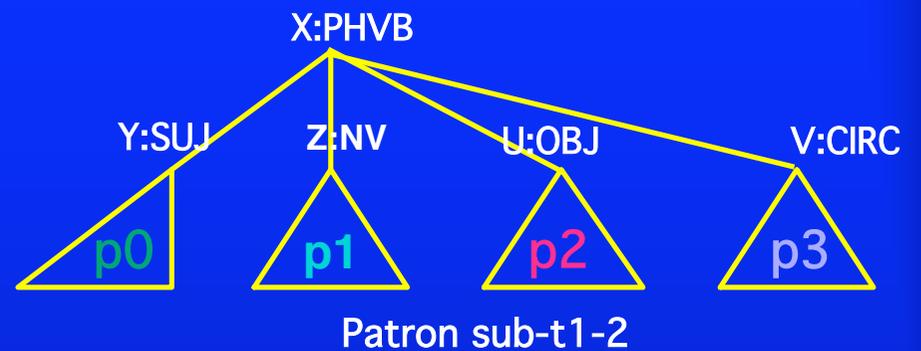
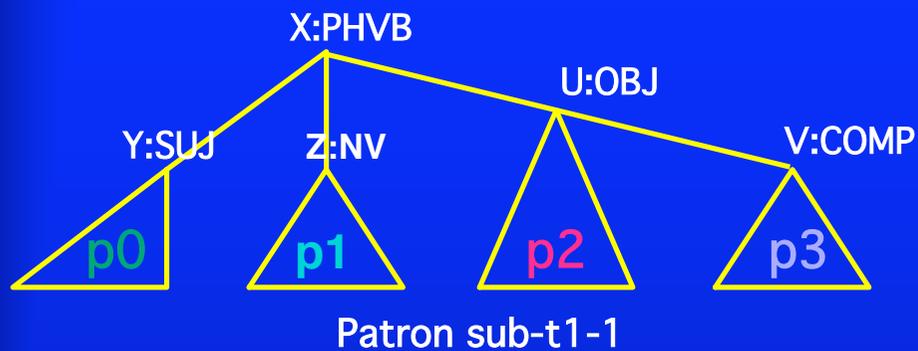
Patron sub-t1-1



Patron sub-t1-2

# Exemple

## ✦ Patrons



## ✦ Méthodes

- Texte(p0) Texte(p1) Parenthèse(Texte(p2), Texte(p3))
  - Le capitaine a rapporté (un vase de Chine).
- Texte(p3) , Texte(p0) Texte(p1) Texte(p2)
  - De Chine, le capitaine a rapporté un vase.

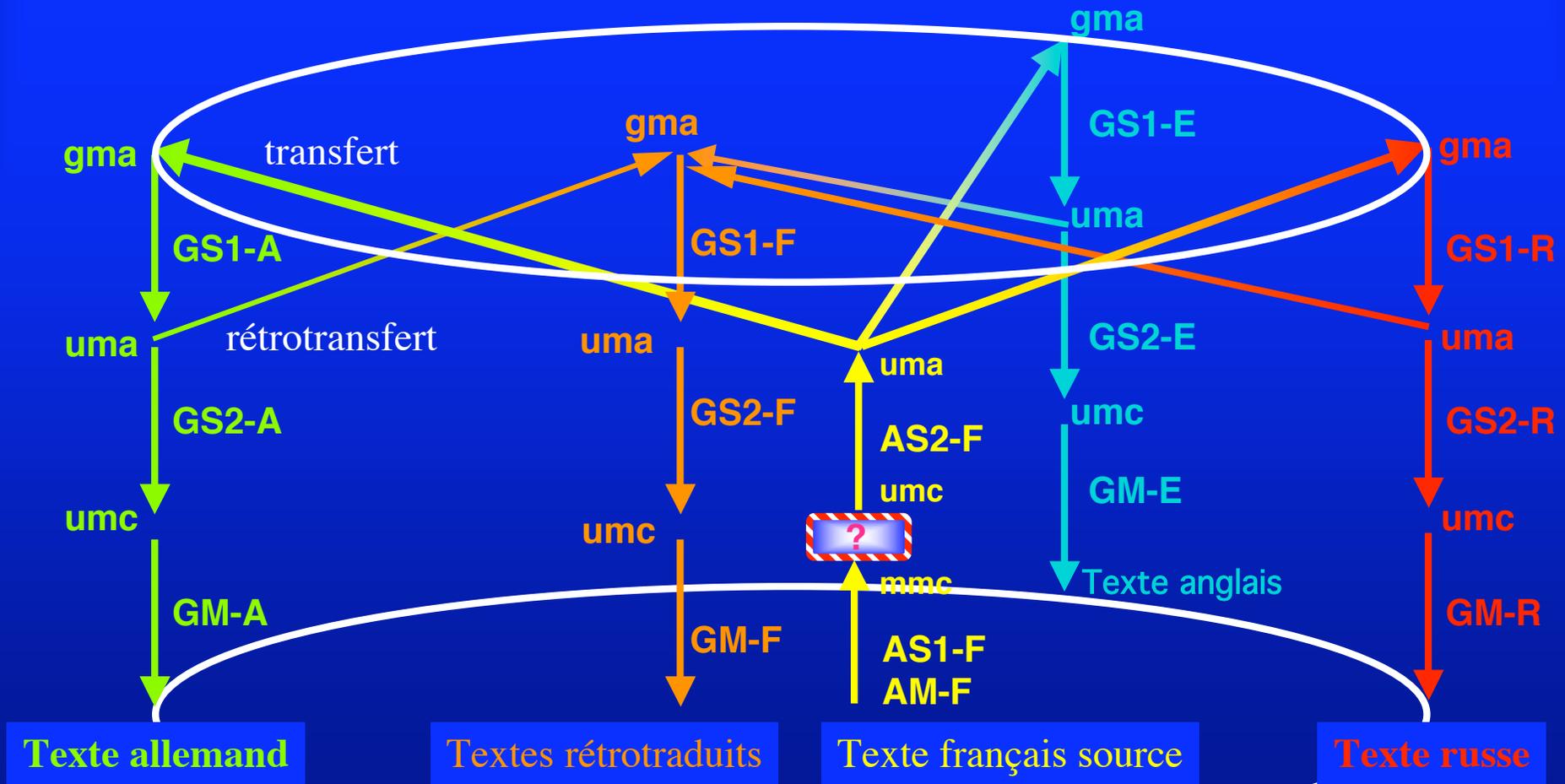
# LIDIA-1

## Architecture & implémentation

- ✦ Architecture linguistique
  - Linguiciel
  - Structure mmc
- ✦ Architecture informatique
  - Architecture distribuée
  - Modèle de “tableau blanc”
  - Outils de programmation
- ✦ Bilan du maquettage

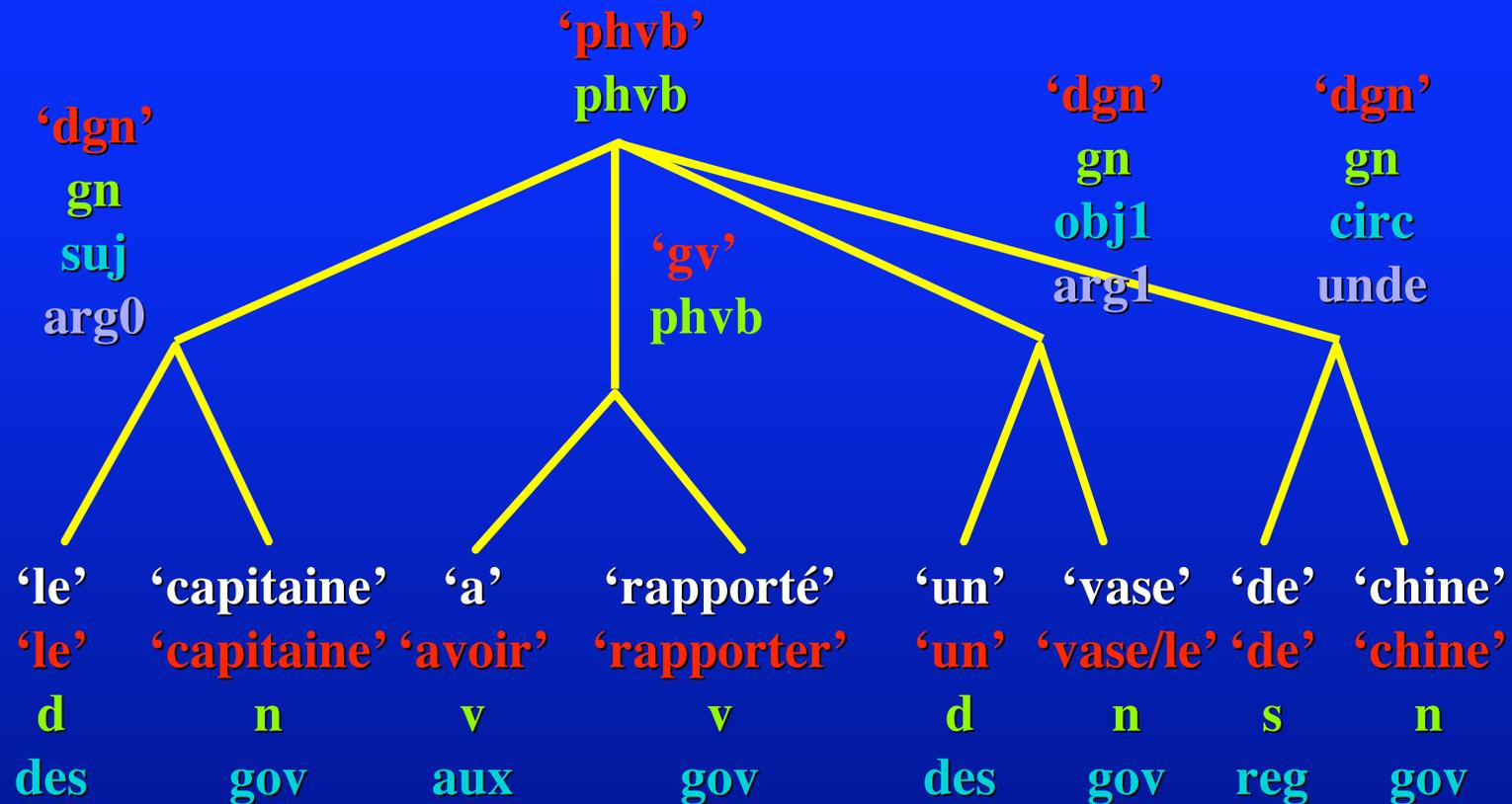
# Architecture linguistique

◆ Le logiciel Ariane-G5 mis en jeu



# Architecture linguistique

- ◆ Structure Multirésolution, Multiniveau, Concrète (MMC)



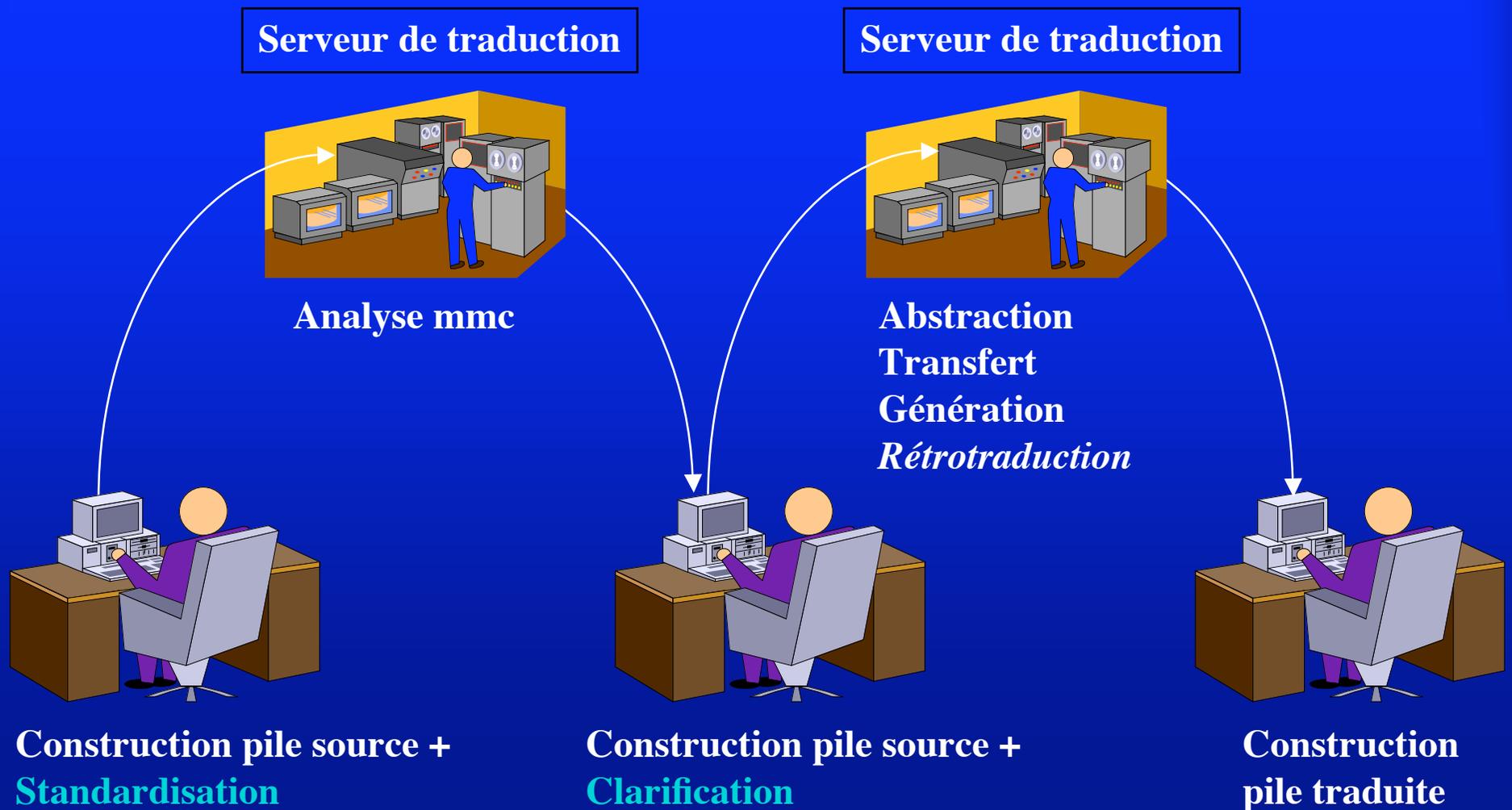
occurrence  
**unité lexicale**

classe syntagmatique  
 classe syntaxique

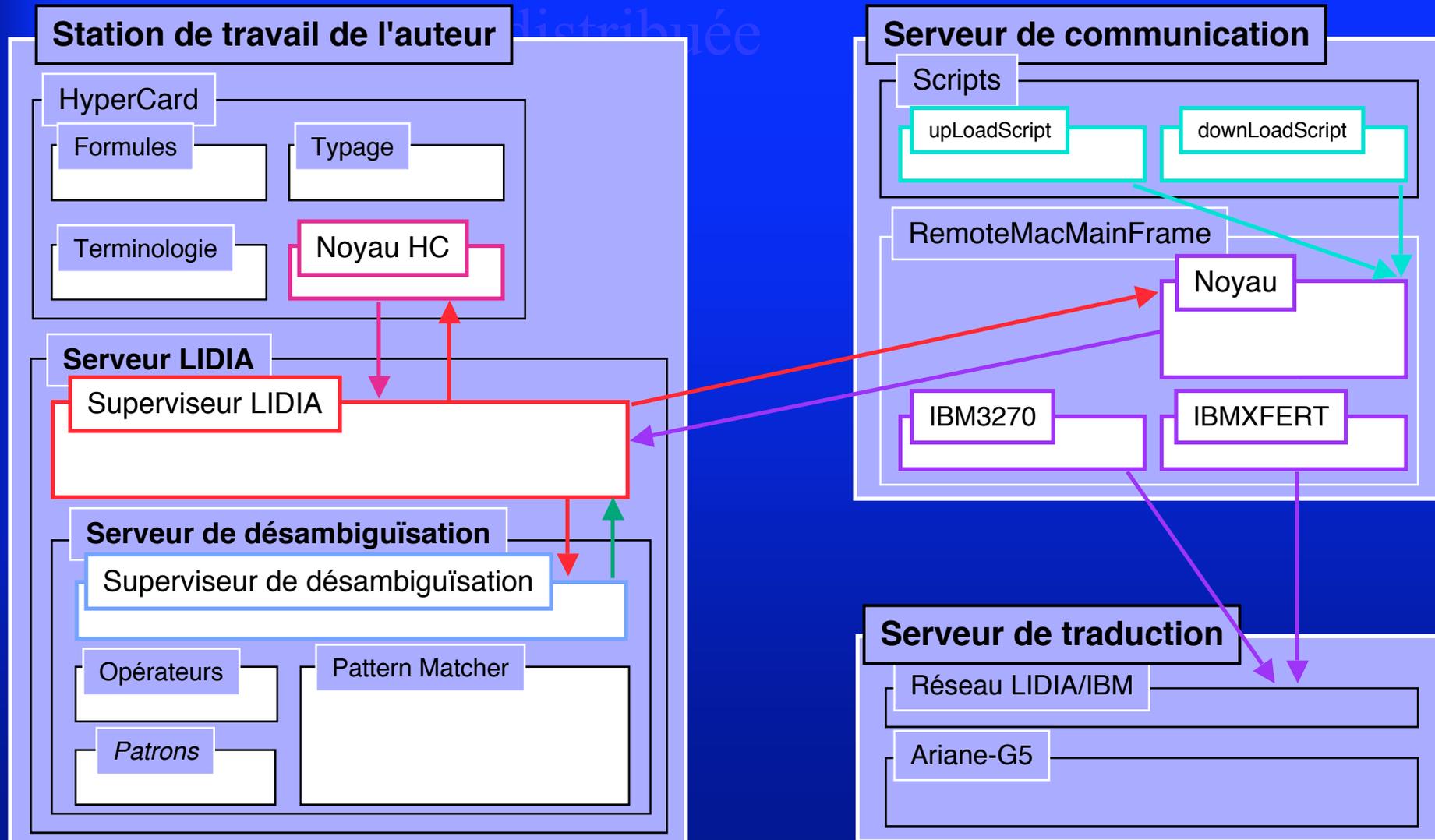
fonction syntaxique  
 relation logique et sémantique

# Architecture

## ✦ Vue générale du processus



# Architecture informatique



# Architecture informatique

## ✦ Vers un modèle de "tableau blanc"

### – Modèle

Le coordinateur gère le tableau blanc et soumet des tâches à différents composants

→ intégration de nouveaux composants sans les modifier

### – Le tableau blanc : un objet miroir

- Un objet miroir par objet HyperCard en cours de traitement

### – Le coordinateur : le superviseur LIDIA

### – Communication par messages et par boîtes aux lettres

- À chaque message est associé un fichier qui contient soit
  - une donnée de traitement
  - un résultat de traitement

# Architecture informatique

## ✦ Outils de programmation

- Langages à objets
  - CLOS (MCL)
  - HyperTalk
  - AppleScript
- Messages (Communication Inter Applications)
  - AppleEvents
  - AppleScripts
  - Vers un modèle d'acteurs
- Représentation des données
  - Fichiers textes (objets miroirs)
  - Listes d'associations (facilement exploitables en Lisp)

# Bilan du maquetage

- ◆ En TAO, même une maquette demande un travail énorme
  - *linguiciel, environnement, algorithmique interne, intégration des composants*
- ◆ Cette maquette n'est pas encore totalement figée :
  - si certains aspects sont suffisants pour le but recherché,
    - principes architecturaux
    - communication
    - interface
  - d'autres sont trop ad hoc ou inefficaces, et doivent être améliorés
    - l'accès à la base lexicale PARAX *est trop lent*
    - la construction de l'arbre des questions *suppose des propriétés fortes des patrons*
    - le passage arbre-ariane → forme-lisp *est écrit en Ariane, ce qui le rend coûteux et dépendant du linguiciel*

# Perspectives (1/3)

LIDIA-2 : vers un prototype expérimentable en interne

## ◆ Refonte de l'architecture de LIDIA-1

- Outil spécifique pour l'interaction avec le texte
  - HyperCard ne permet pas de structurer finement
  - Outil inspiré de SGML/TEI — coopération avec OPERA ?
  - Interfaçage avec HyperCard, Word, Grif
- Mise en œuvre de nouveaux outils
  - Outils Sambre en cours de développement
  - Intégration de techniques numériques
- Processus de clarification modifiable
  - par le développeur linguiste
  - par l'utilisateur ?

# Perspectives (2/3)

LIDIA-2 : vers un prototype expérimentable en interne

- ✦ Développement linguistique et ergonomique
  - connaissances lexicales et grammaticales restant limitées au domaine d'expérience choisi
  - sous-ensemble des ambiguïtés, le reste étant traité sans dialogue (meilleur score par défaut)
- ✦ Volonté de progresser dans les aspects non encore traités
  - Ambiguïtés contrastives difficiles à rendre sensibles
    - modalité, aspect
  - Définition de familles de sous-langages
    - styles d'énoncés, genres de textes, types de document

# Perspectives (3/3)

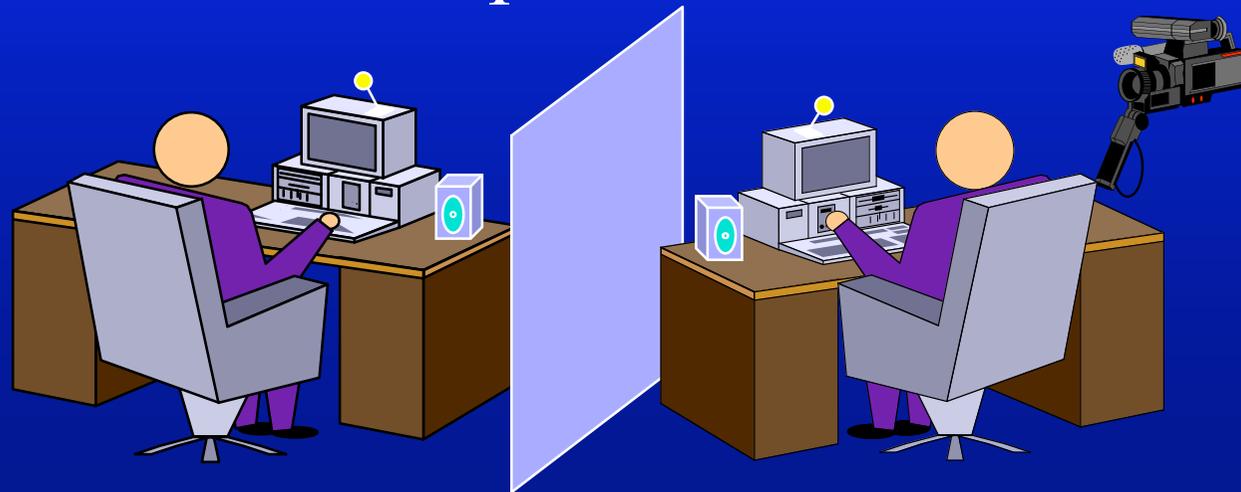
## Simulateur pour la TAFD et la désamb. interactive

### ◆ Objectifs

- Expérimentation possible avant d'avoir un logiciel complet
- Expérimentation d'autres modalités
- Collecte d'informations

### ◆ Configuration minimale pour la TAFD

MAGICIEN D'OZ



AUTEUR