

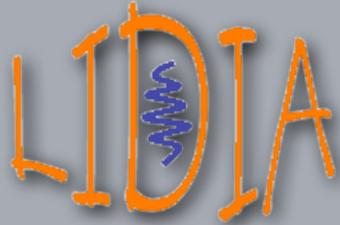
LIDIA & les Documents Auto-Explicatifs



Hervé Blanchon

herve.blanchon@imag.fr

Plan

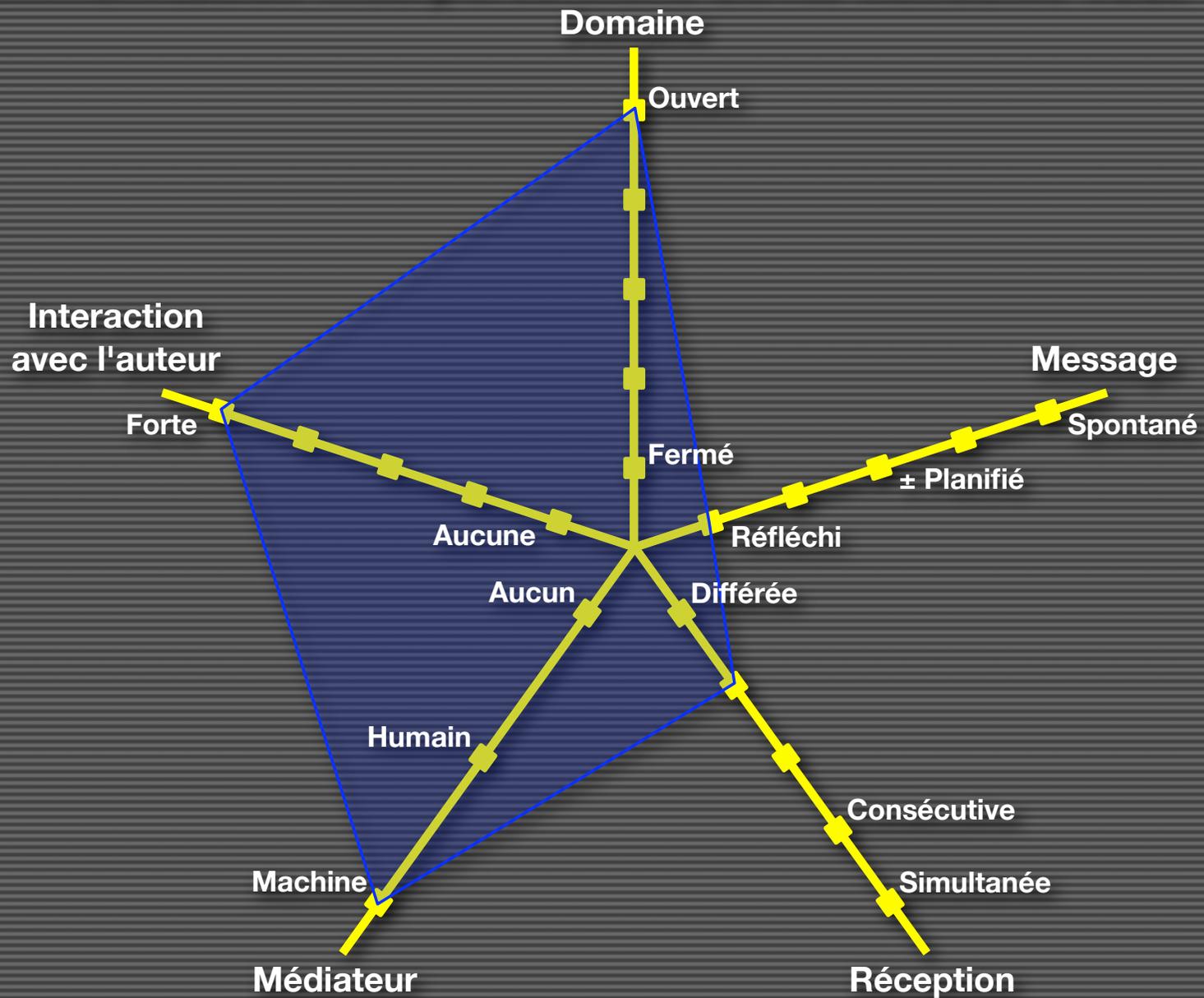


- Introduction
- Désambiguïsation interactive
- LIDIA-1 : démonstration
- Vers les Documents Auto-Explicatifs
 - LIDIA-3
- Verrous et pistes pour le futur

Introduction

- Projet LIDIA ⇒ la TAO de l'auteur monolingue
 - Rédiger dans sa propre langue et faire traduire
 - sans révision
 - sans compréhension explicite
 - sans langage “contrôlé” (lexique, syntaxe)
 - sans heuristique
 - ⚠ Toutes les ambiguïtés sont *a priori* possibles
 - Comment lutter ?
 - 💡 **La désambiguïstation interactive**
 - L'auteur est le mieux placé pour savoir ce qu'il veut dire !

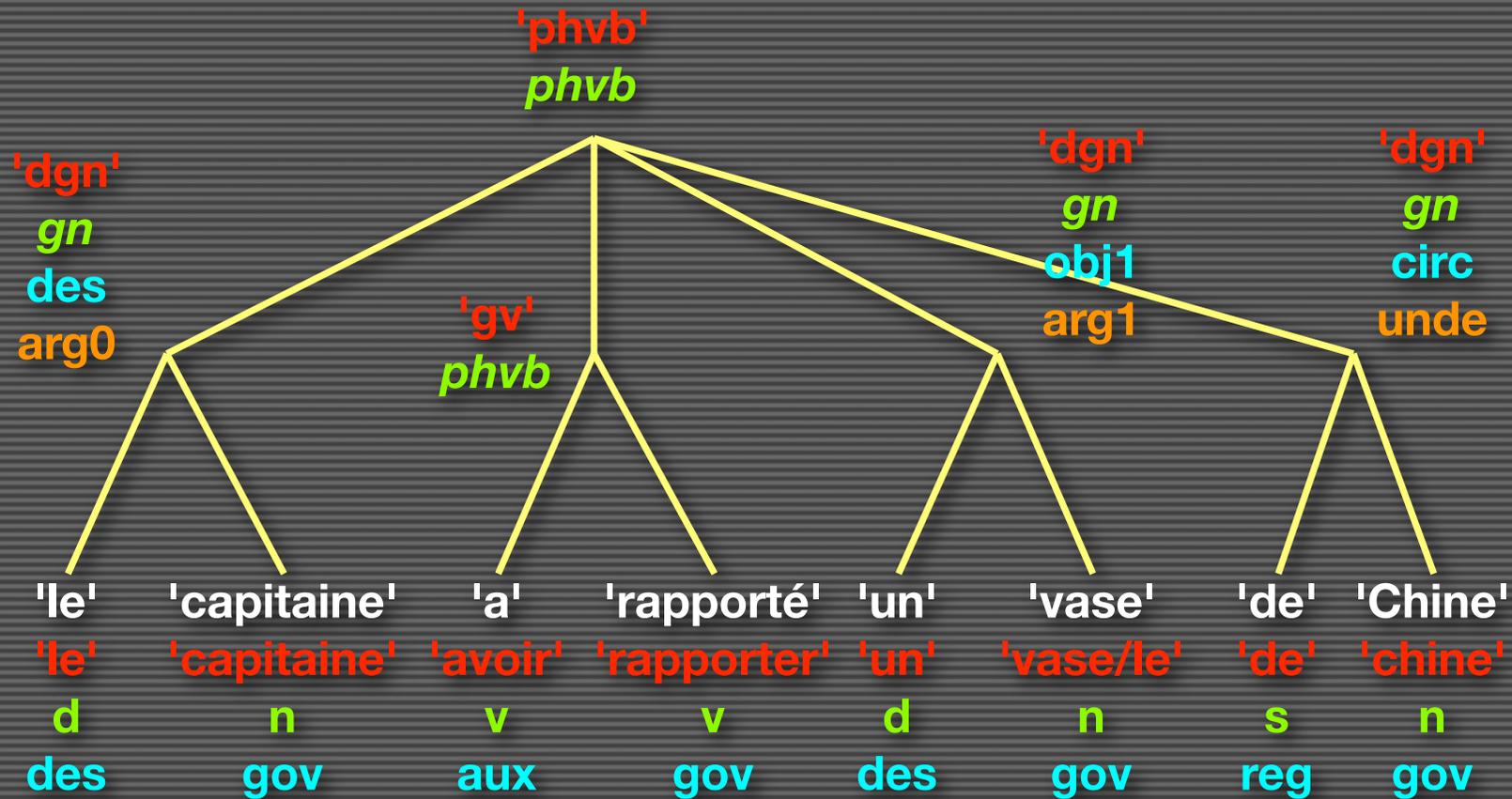
Communication Interpersonnelle Écrite Réfléchi Tout Terrain



Introduction

- Architecture linguistique
 - Transfert multiniveau à acceptions, propriétés et relations interlingues
- Structures manipulées
 - Des arbres
 - Structure Multisolution, Multiniveau, Concrète (MMC)

Structure Multiniveau, Concrète



occurrence

unité lexicale

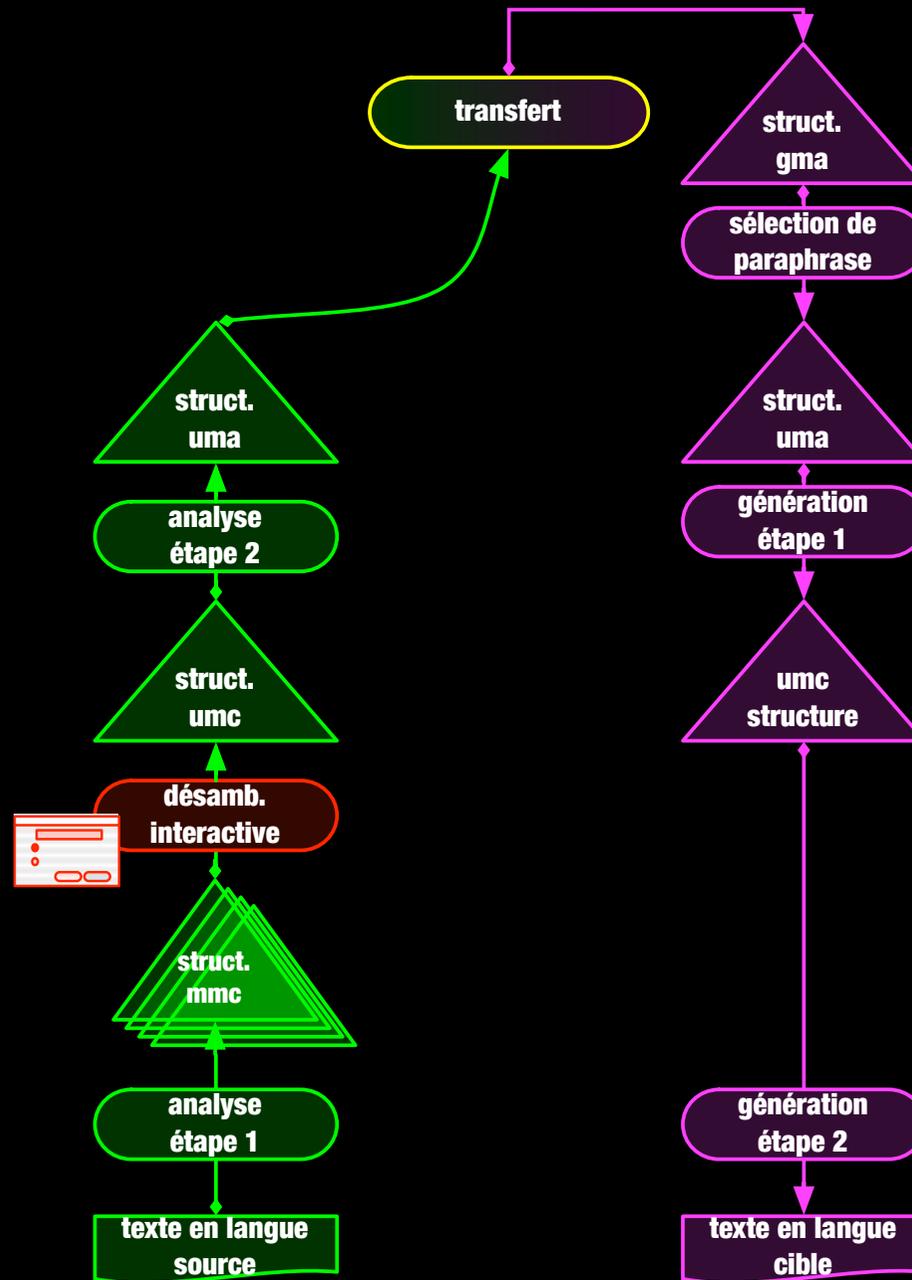
classe syntagmatique

classe syntaxique

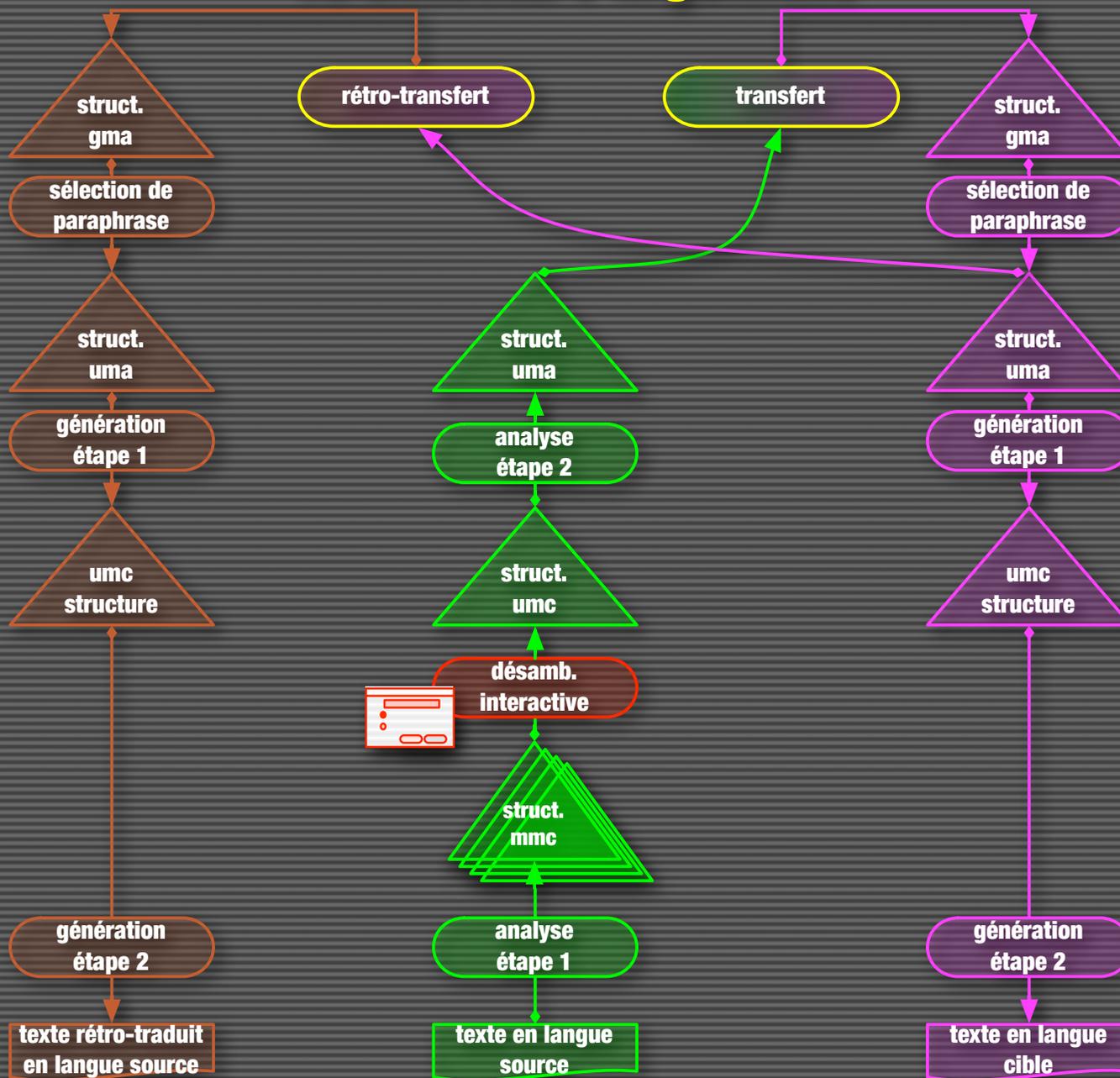
fonction syntaxique

relation logique et sémantique

Architecture linguicielle



Architecture linguistique



Introduction

- Architectures informatiques
 - Vers une meilleure architecture client-serveur
 - un environnement de rédaction indépendant de la plate-forme
 - le même protocole de communication entre les différents composants

LIDIA-I.mail

Rédaction

HyperCard (Mac OS 9)

Communication

Via un **Coordinateur**

AppleEvents (Mac OS)

Protocole NMTP

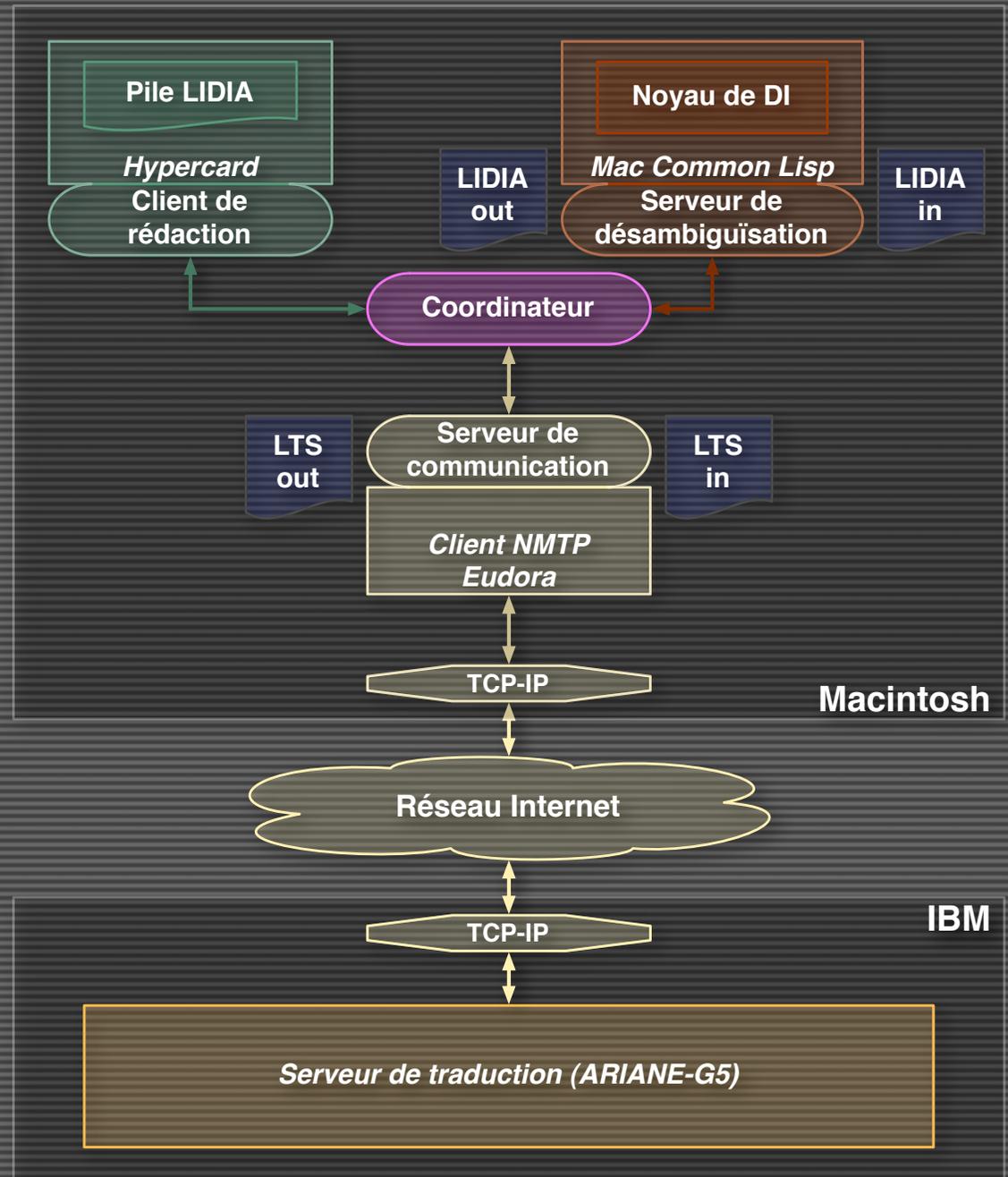
Données

Fichiers HyperCard

Fichiers de texte

Messages

[Planas, Lafourcade
& Sérasset 94]



LIDIA-2 & 3

Rédaction

Client - Java

Client - Amaya

Communication

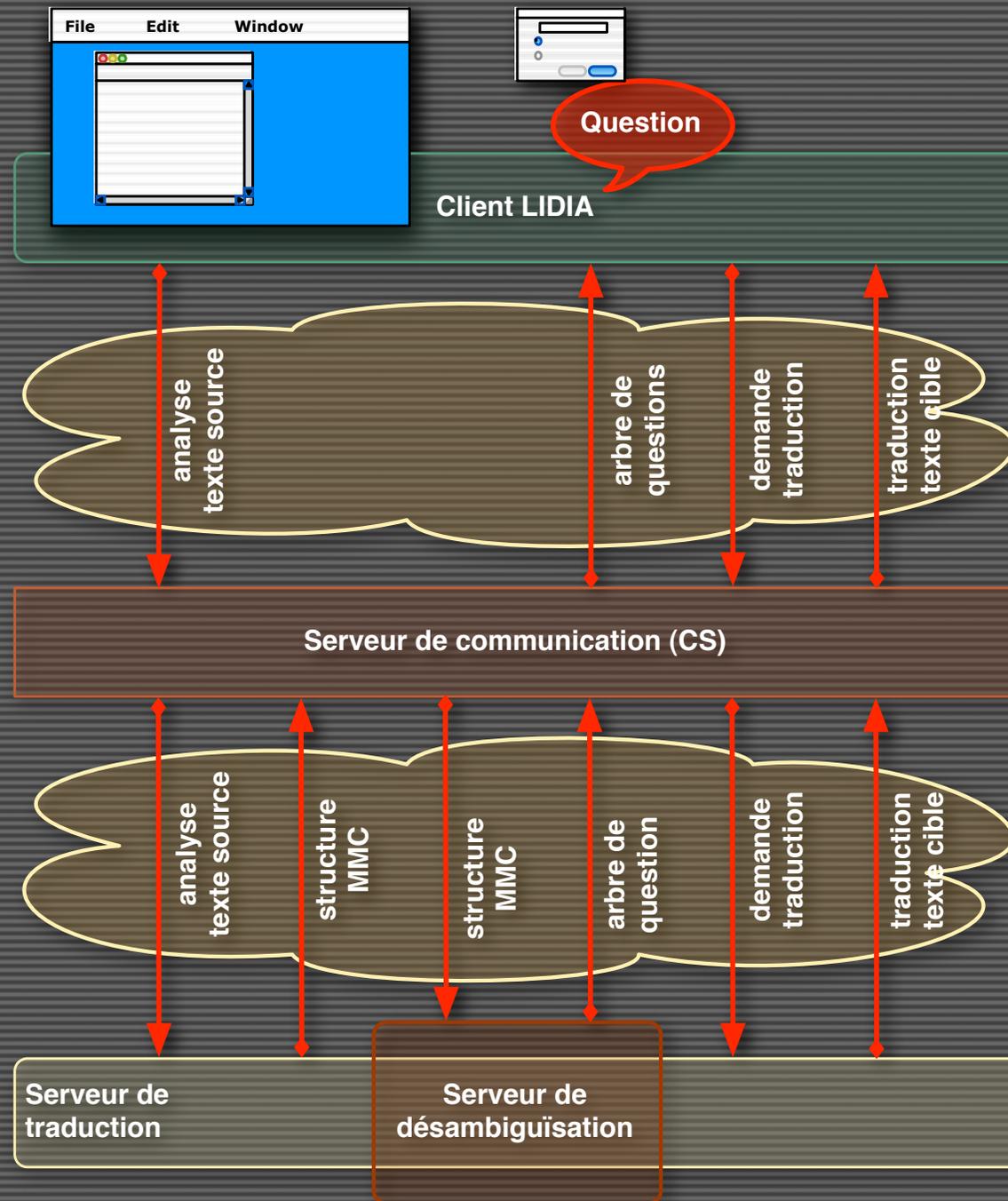
Via un CS

Protocole Telnet

Données

Fichiers XML

Messages



Désambiguïsation interactive

⚠️ Qu'est-ce qu'une ambiguïté ?

- Définition linguistique non opératoire

[Fuchs 96]

💡 Définition formelle

[Blanchon & Boitet 04]

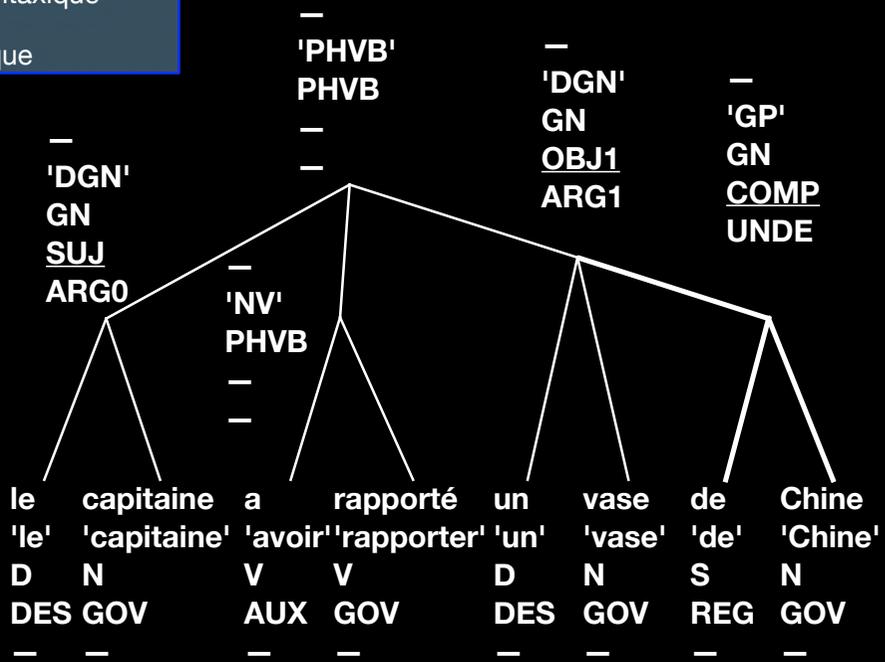
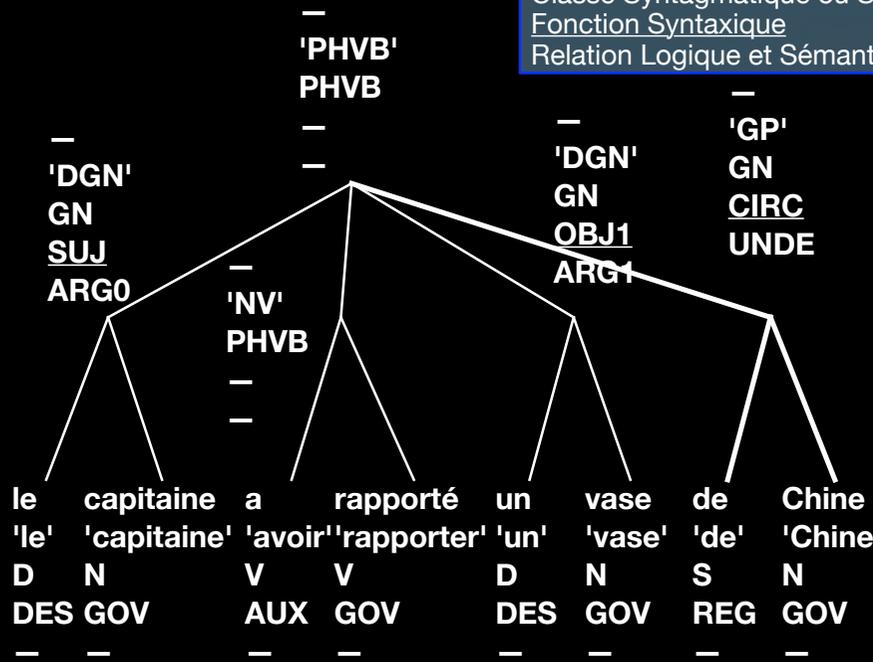
⚠️ Comment représenter une ambiguïté ?

[Blanchon 95]

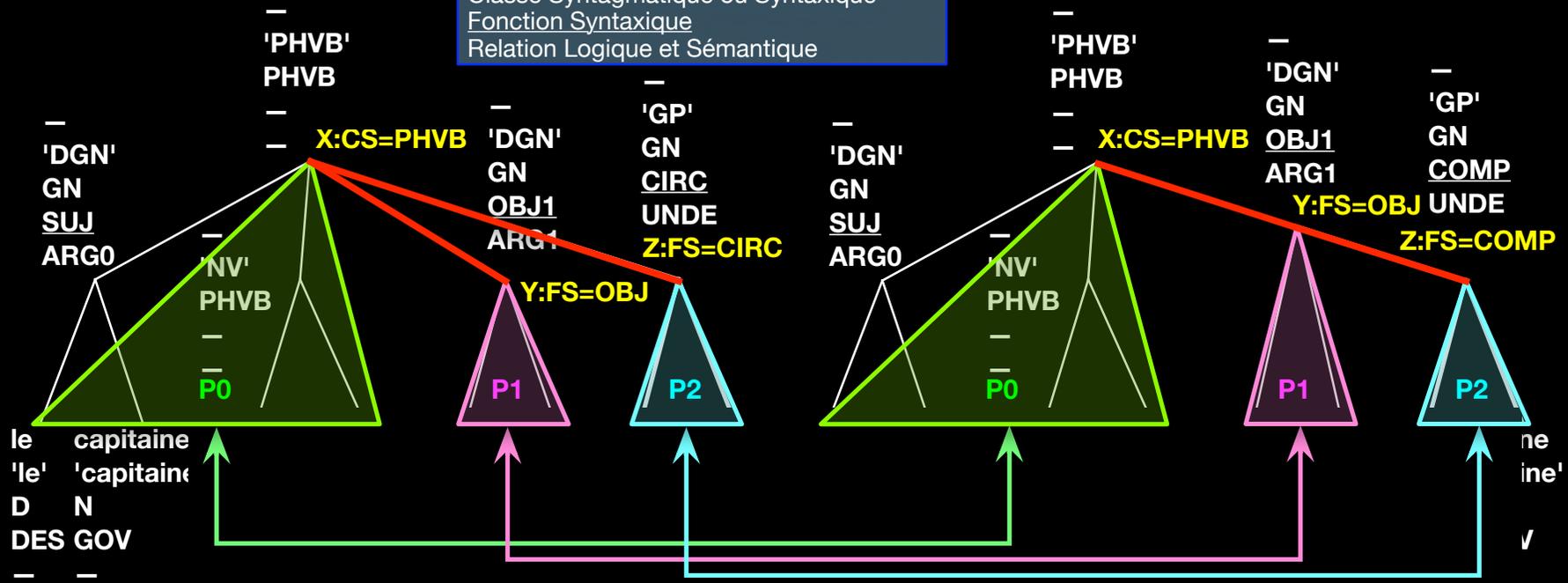
💡 Ambiguïté = co-occurrence de plusieurs schémas d'arbres *synchronisés* groupés en faisceaux

- schéma d'arbre \ni variables de nœud + variables de forêt
- *synchronisation* sur les variables de forêts
- Représentation non liée à un analyseur donné

occurrence
 Unité Lexicale
 Classe Syntagmatique ou Syntaxique
 Fonction Syntaxique
 Relation Logique et Sémantique



occurrence
 Unité Lexicale
 Classe Syntagmatique ou Syntaxique
 Fonction Syntaxique
 Relation Logique et Sémantique



Désambiguïsation interactive

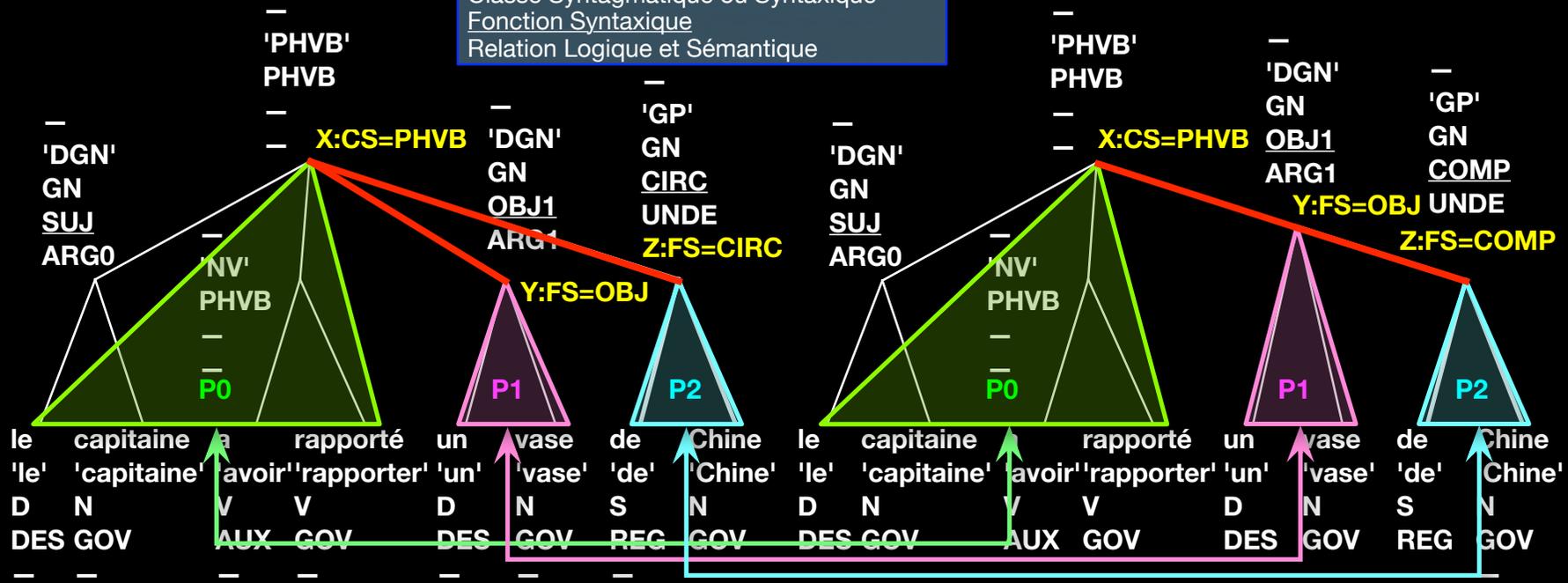
⚠ Comment poser une question ?

⚠ Il ne faut pas que la question soit ambiguë !

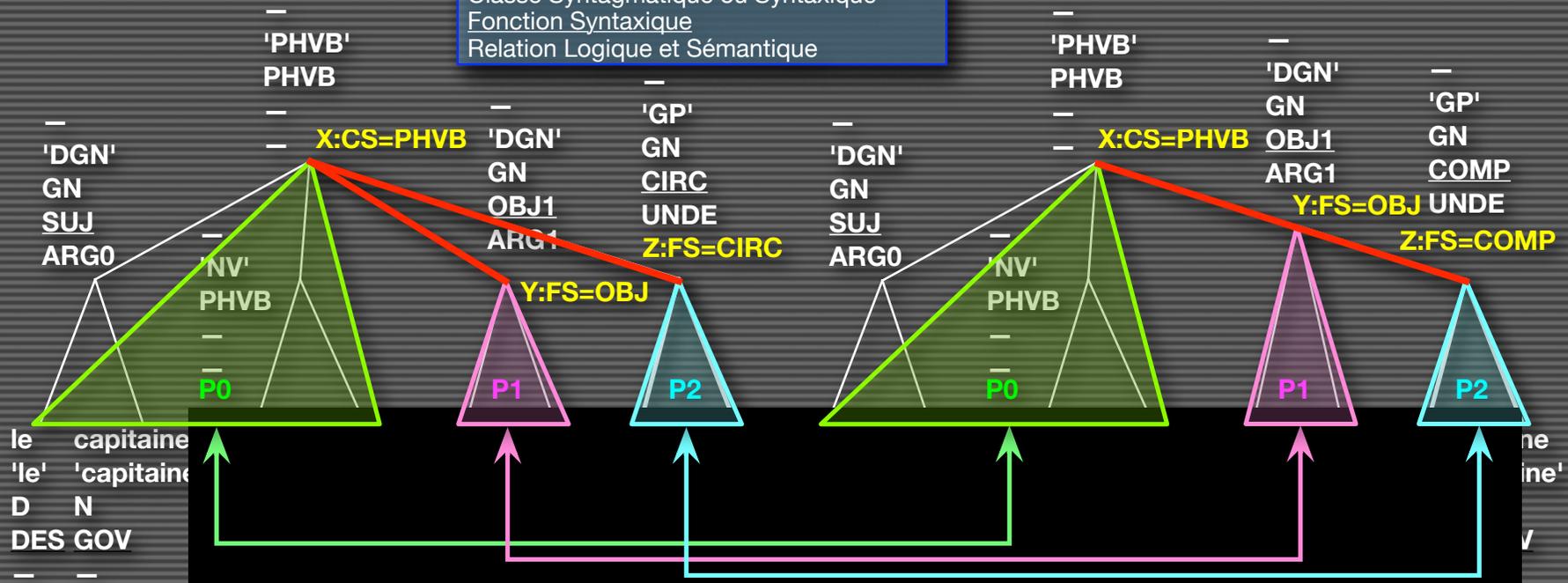
💡 **Faire un rephrasage contrôlé**

- Différentes projections des variables de forêt
 - Utilisation de marques (ponctuation, parenthèses)
 - Remplacement d'occurrences ambiguës
 - Génération morphologique au niveau des lemmes
- ⇒ **une méthode de rephrasage est associée à chaque patron**

occurrence
 Unité Lexicale
 Classe Syntagmatique ou Syntaxique
 Fonction Syntaxique
 Relation Logique et Sémantique



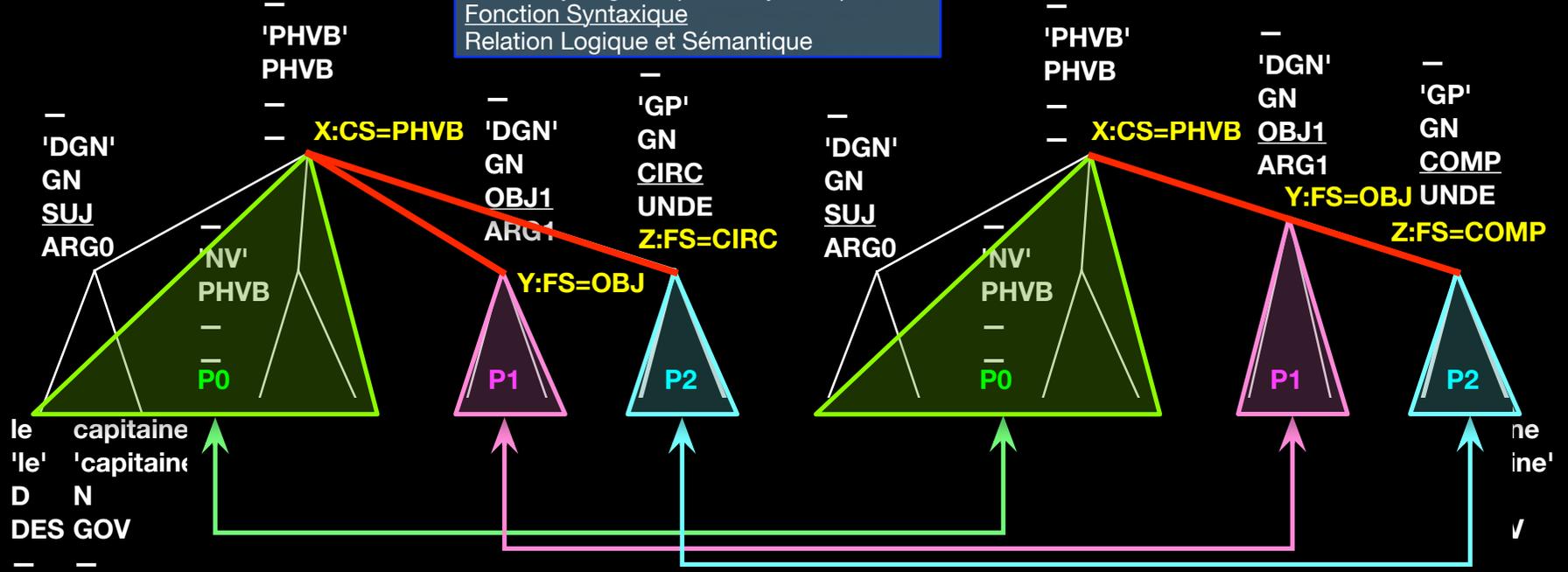
occurrence
 Unité Lexicale
 Classe Syntagmatique ou Syntaxique
 Fonction Syntaxique
 Relation Logique et Sémantique



Text(P2), Text(P0) Text(P1)

Text(P0) Bracket(Text(P1), Text(P2))

occurrence
 Unité Lexicale
 Classe Syntagmatique ou Syntaxique
 Fonction Syntaxique
 Relation Logique et Sémantique



Text(P2), Text(P0) Text(P1)

Text(P0) Bracket(Text(P1), Text(P2))

ambiguïté

Il y a plusieurs interprétations pour la phrase :
 Le capitaine a rapporté un vase de Chine.

Choisissez la bonne.

De Chine, le capitaine a rapporté un vase.

Le capitaine a rapporté (un vase de Chine).

OK

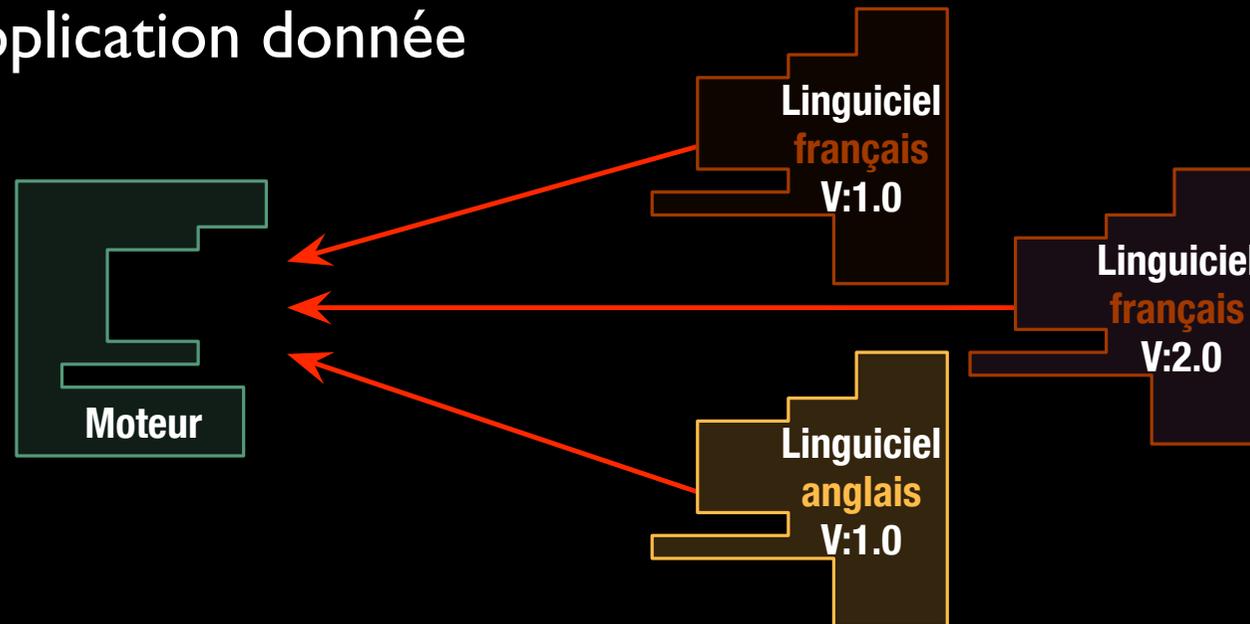
Désambiguïsation interactive

💡 Un module de DI est l'association

[Blanchon 95]

[Grasson 96]

- d'un **moteur** réutilisable fournissant les services de base
- d'un **linguiciel** particulier adapté à une langue et une application donnée



Désambiguïsation interactive

-  **Un module de DI** est l'association *[Blanchon 95]*
[Grasson 96]
- d'un **moteur** réutilisable fournissant les services de base
 - d'un **linguiciel** particulier adapté à une langue et une application donnée

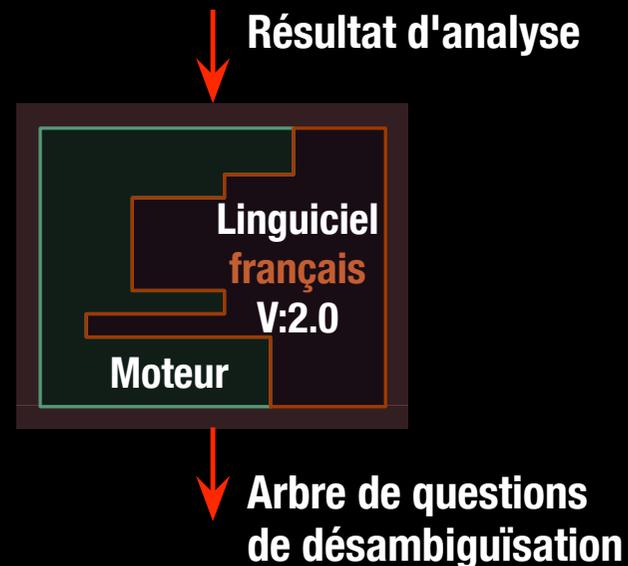
Désambiguïsation interactive

💡 **Un module de DI** est l'association

[Blanchon 95]

[Grasson 96]

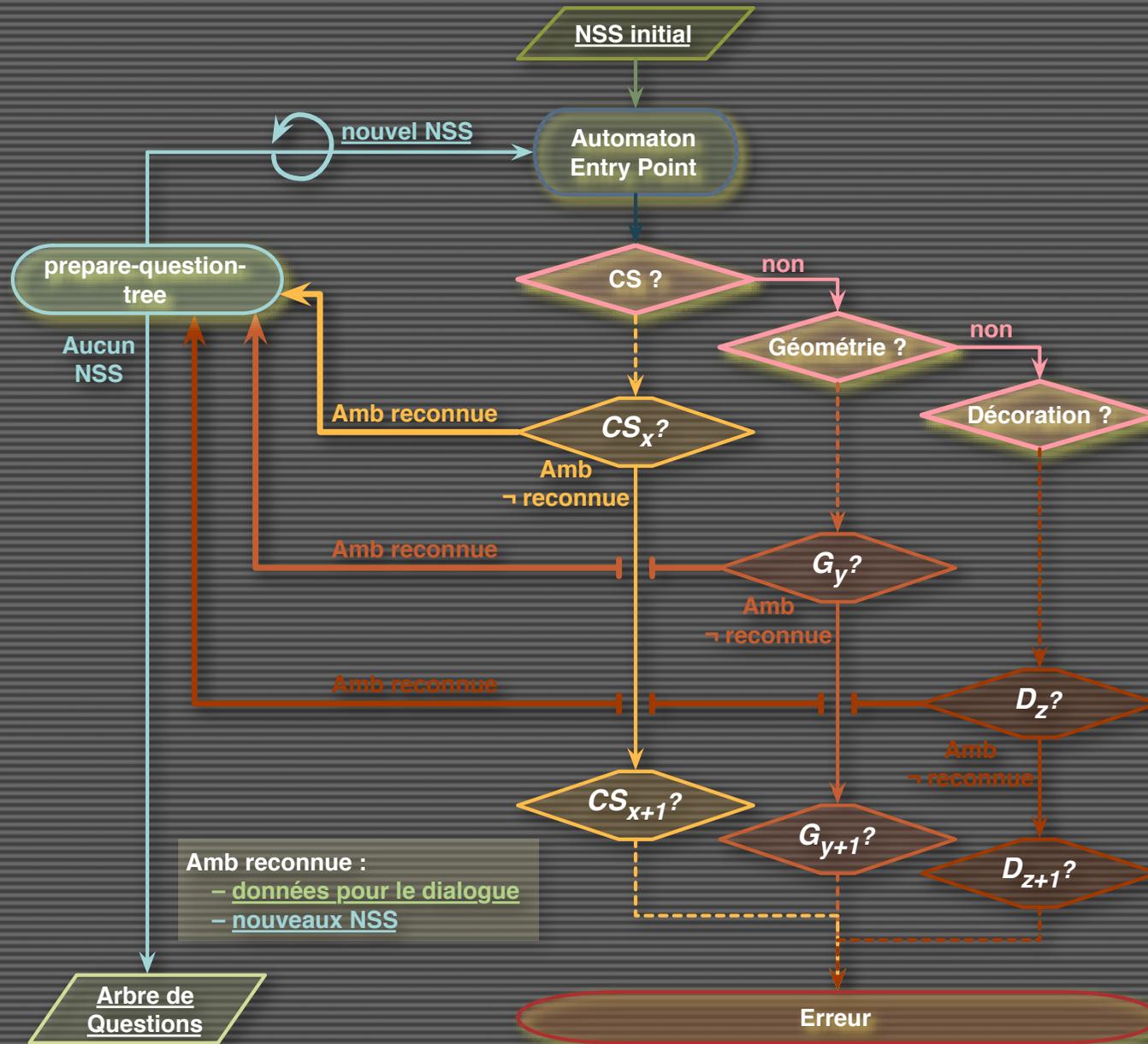
- d'un **moteur** réutilisable fournissant les services de base
- d'un **linguiciel** particulier adapté à une langue et une application donnée



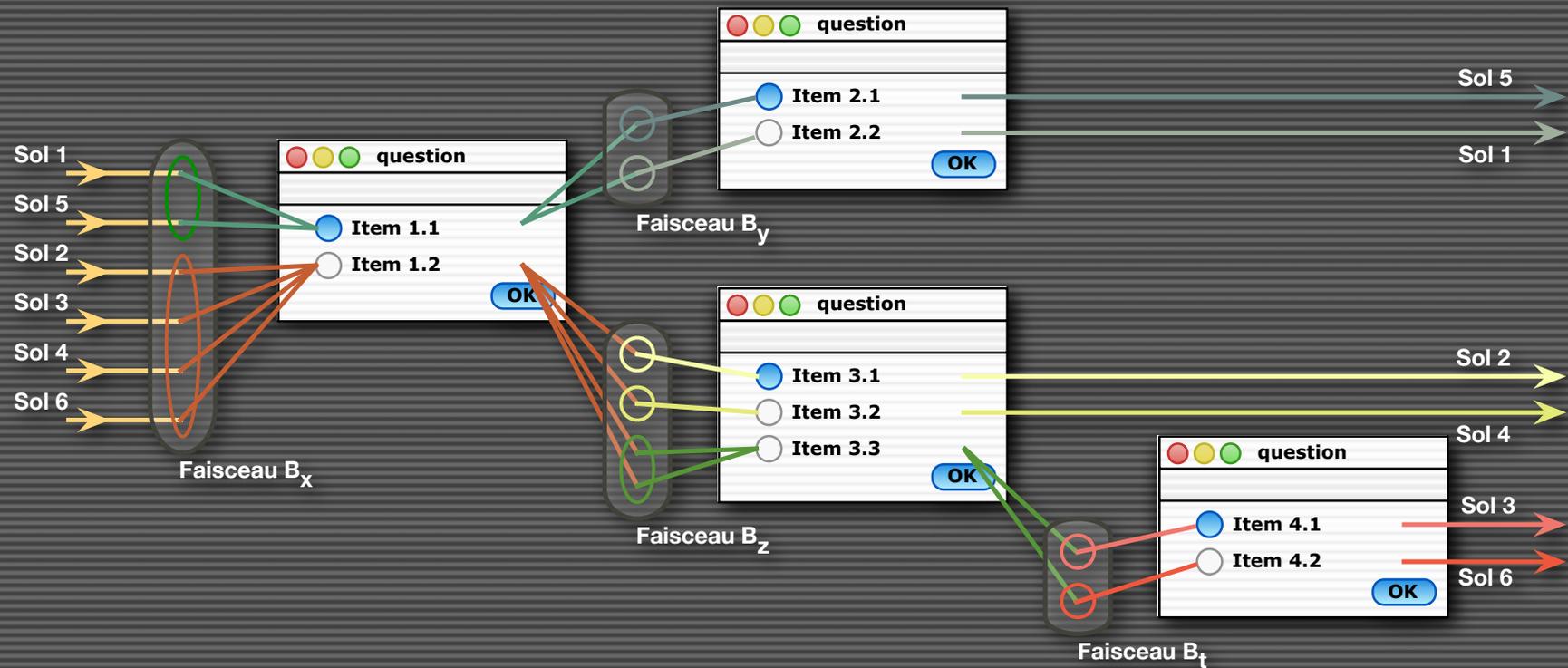
Désambiguïsation interactive

- **Linguiciel**
 - description des ambiguïtés
 - description de la construction des items de question
 - stratégie de désambiguïsation
 - ordre des questions
- **Moteur**
 - reconnaissance des ambiguïtés du linguiciel
 - reconnaissance de méta-classes d'ambiguïtés
 - classe syntaxique, géométrie, décoration
 - opérateurs de production des items de question

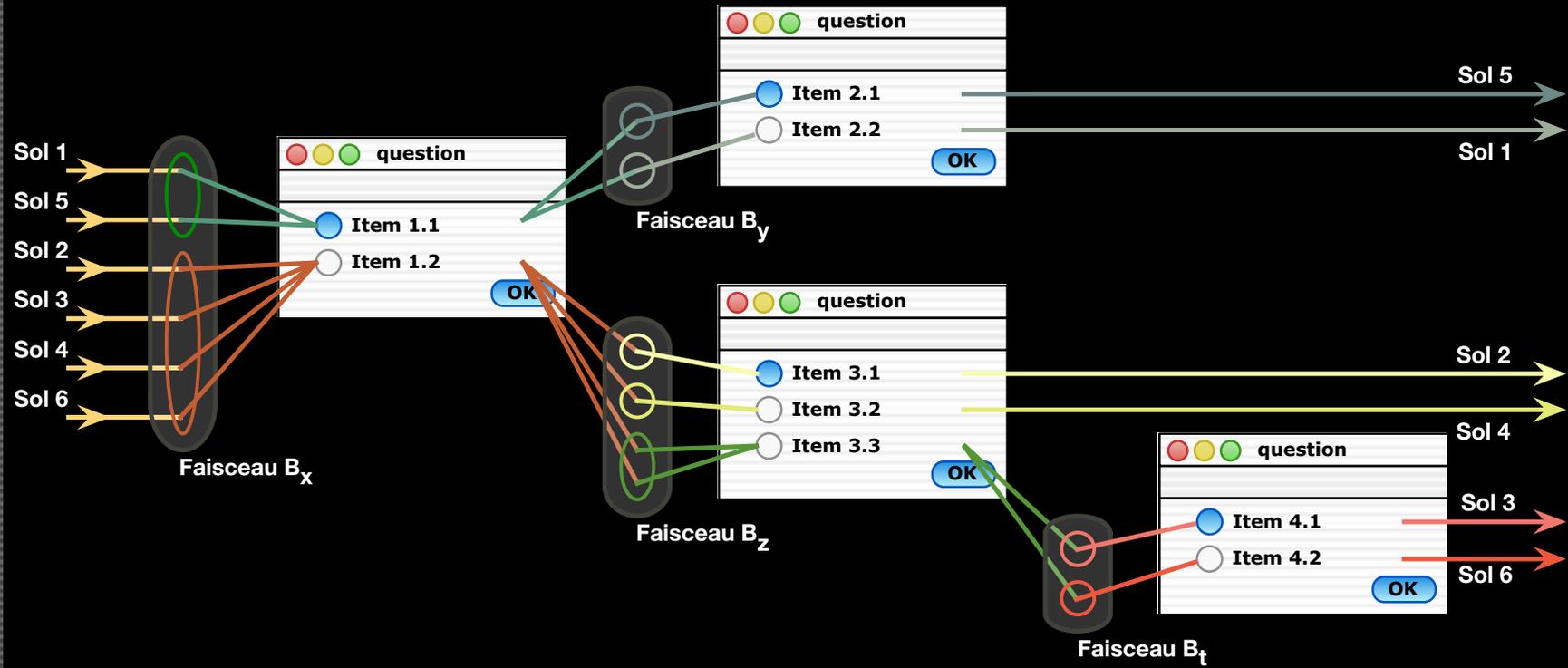
Automate de désambiguïsation



Arbre de questions



Arbre de questions



Nombre de questions

Facteur de branchement	2	3	4
Nombre d'analyses	4	3	2
10	7	5	4
100	9	6	5
500	10	7	5
1000	15	10	8
32 728 (30 mots • 2 ^{30/2} analyses)			

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

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**.

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).

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.

Types d'ambiguïtés (5/5)

- Ambiguïté de décoration (suite)
 - Ordre des arguments d'un verbe transitif
 - Quel **auteur** cite ce **conférencier** ?
 - Un **auteur** cite un **conférencier**.
 - Un **conférencier** cite un **auteur**.

Évaluation en utilisabilité et utilité

[Blanchon & Fais 96]
ATR

Évaluation en utilisabilité et utilité

- Seconde expérience

[Blanchon & Fais 96]
ATR

- 60 sujets en 4 groupes
 - 2 modalités : orales (O) & écrites (E)
 - 2 types de questions : humaines (H) & machines (M)
- 35 phrases ambiguës (5 classes d'ambiguïtés)

Évaluation en utilisabilité et utilité

- Seconde expérience

[Blanchon & Fais 96]
ATR

- 60 sujets en 4 groupes

- 2 modalités : orales (O) & écrites (E)

- 2 types de questions : humaines (H) & machines (M)

- 35 phrases ambiguës (5 classes d'ambiguïtés)

- Résultats

- Pas de \neq significative entre les groupes Mx et Hx

- Pas de \neq significative entre les groupes xE et xO

Évaluation en utilisabilité et utilité

● Seconde expérience

- 60 sujets

- 2 m

- 2 ty

- 35 ph

● Résultats

- Pas de \neq

- Pas de \neq significative entre les groupes xE et xO

[Blanchon & Fais 96]
ATR

Le lecteur comprend les questions !

L'auteur devrait aussi les comprendre !

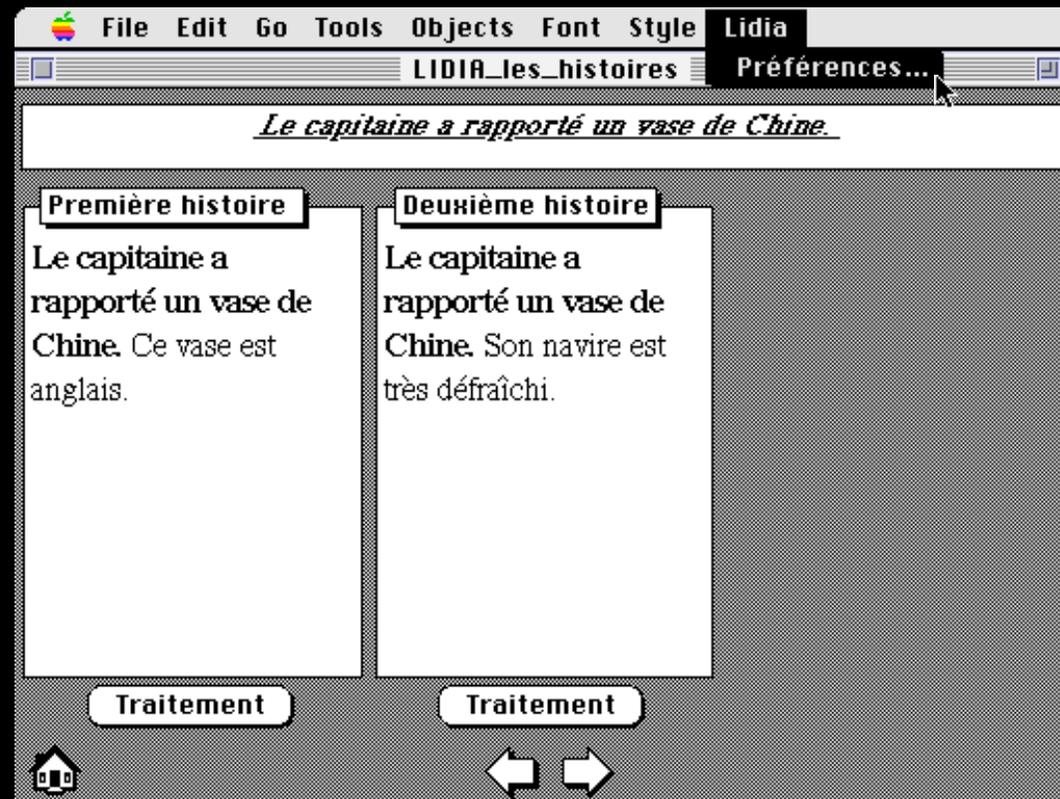
machines (M)

quités)

les groupes Mx et Hx

LIDIA-I : Démonstration (1/12)

- Une carte d'histoire de la pile de démonstration



LIDIA-I : Démonstration (2/12)

- Sélection de l'objet à traduire



LIDIA-I : Démonstration (3/12)

- **Objet en cours de traitement, voire l'état d'avancement**



LIDIA-I : Démonstration (4/12)

- État d'avancement des traitements



LIDIA-I : Démonstration (5/12)

- **Objet en cours de traitement, demande de clarification**



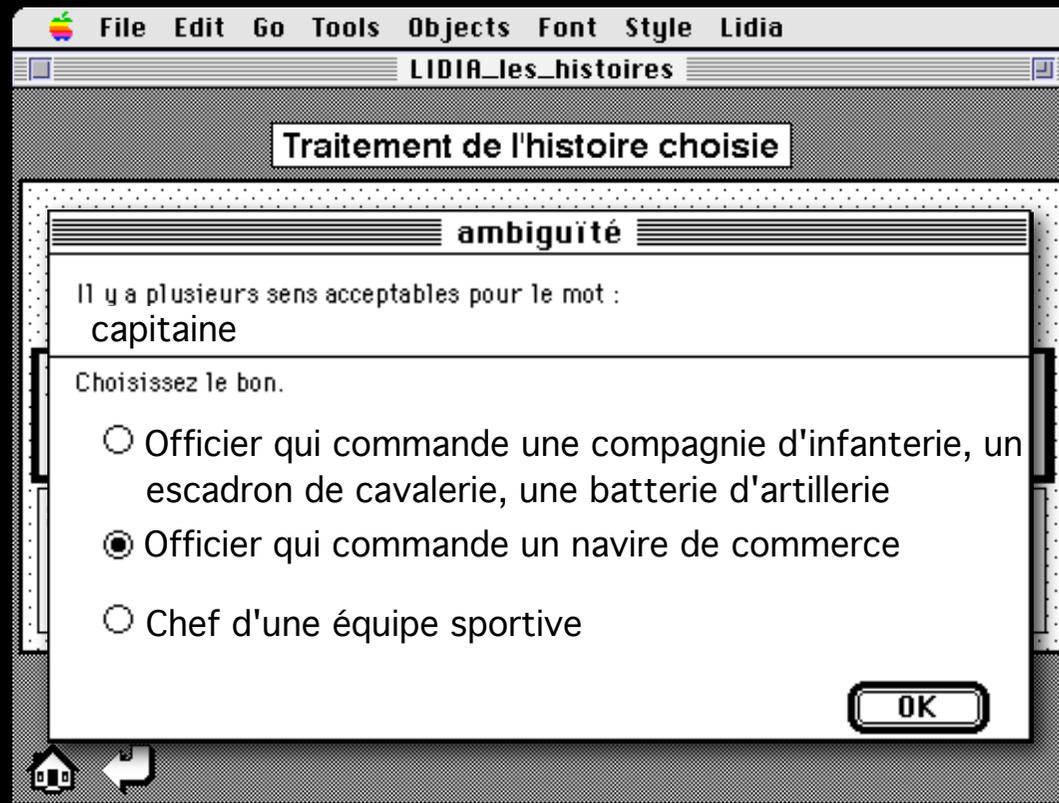
LIDIA-I : Démonstration (6/12)

- Réponse aux questions : Structure argumentaire du verbe



LIDIA-I : Démonstration (7/12)

- Réponse aux questions : polysémie



induite par l'allemand (acceptions interlingues)

LIDIA-I : Démonstration (8/12)

- **Objet poursuivant son traitement, voir les annotations**



LIDIA-I : Démonstration (9/12)

- Annotations : classes syntaxiques, groupes syntagmatiques



LIDIA-I : Démonstration (10/12)

- **Objet en fin de traitement, voir la rétrotraduction**



LIDIA-I : Démonstration (1/1/2)

- **Rétrotraduction**



LIDIA-I : Démonstration (12/12)

- **Obtention d'un document traduit**

Erste Geschichte

Der Hauptmann hat eine Vase aus China mitgebracht. Die Vase ist englisch.

Behandlung

Zweite Geschichte

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

Behandlung

Documents Auto-Explicatifs

⚠ Pourquoi “jeter” les informations de DI ?!

- Elles représentent le sens choisi par l’auteur !

⚠ Ne peut-on pas en faire profiter les lecteurs ?

Documents Auto-Explicatifs

Documents Auto-Explicatifs

- Enrichir le document avec les réponses de la DI
- **Présenter au lecteur**, à la demande, ...
 - ... **le sens choisi par l'auteur** parmi les interprétations potentielles
- Un visualiseur de DAE (un format de document portable ?)

[Blanchon & Boitet --]

[Vigli 95]

[Ghorayeb 02]

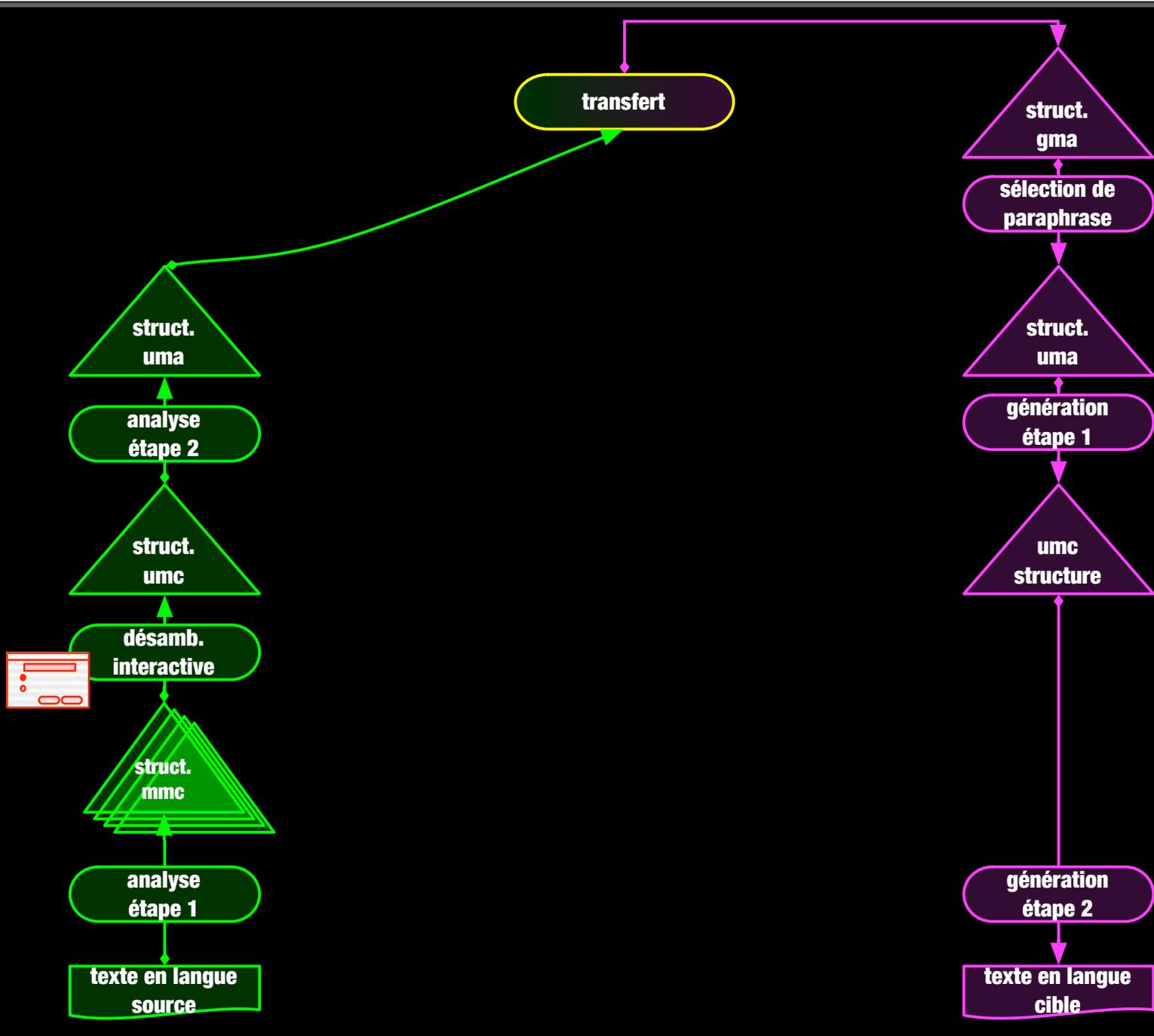
[Belkhatir 01]

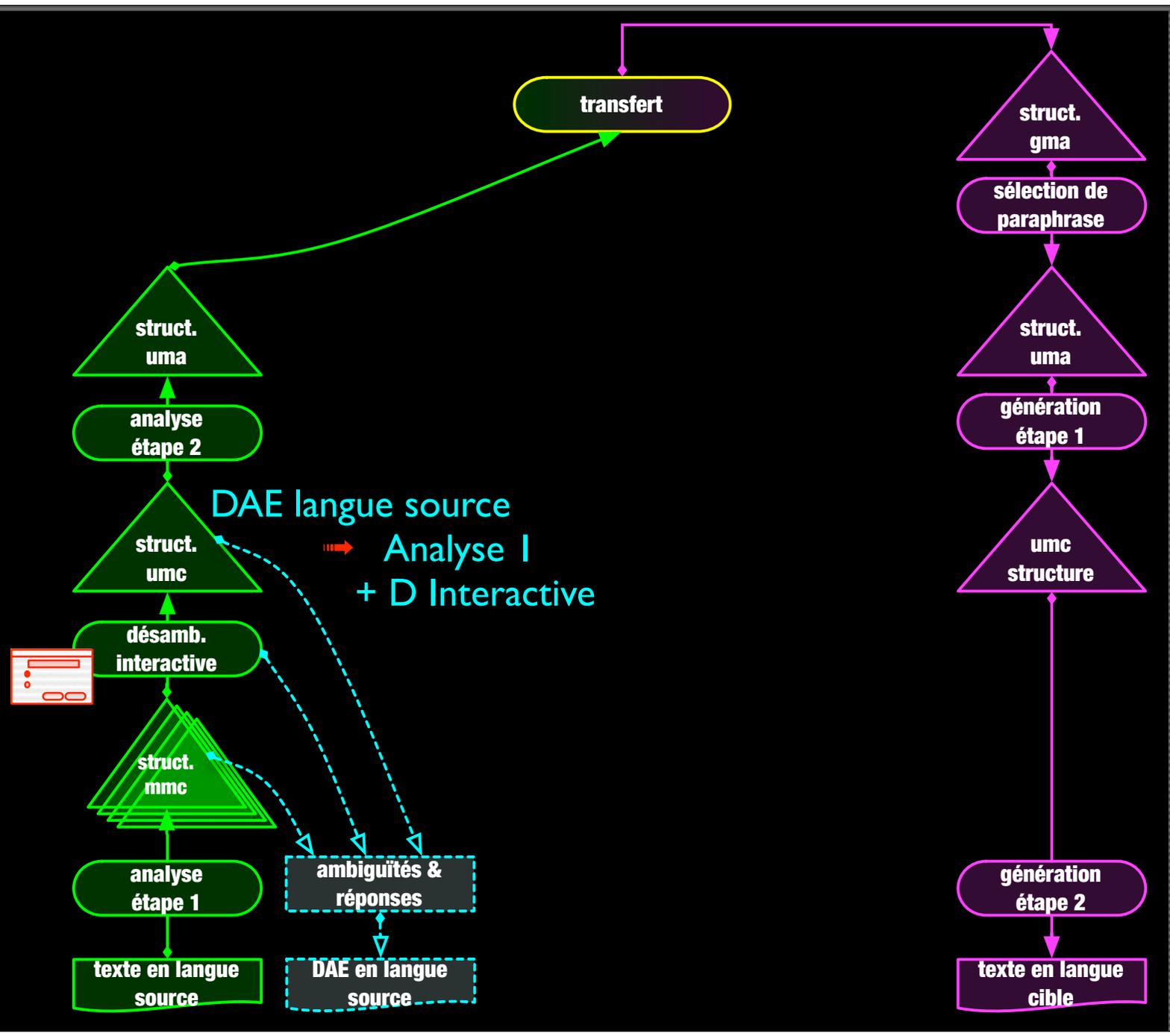
[Gressard 01-03]

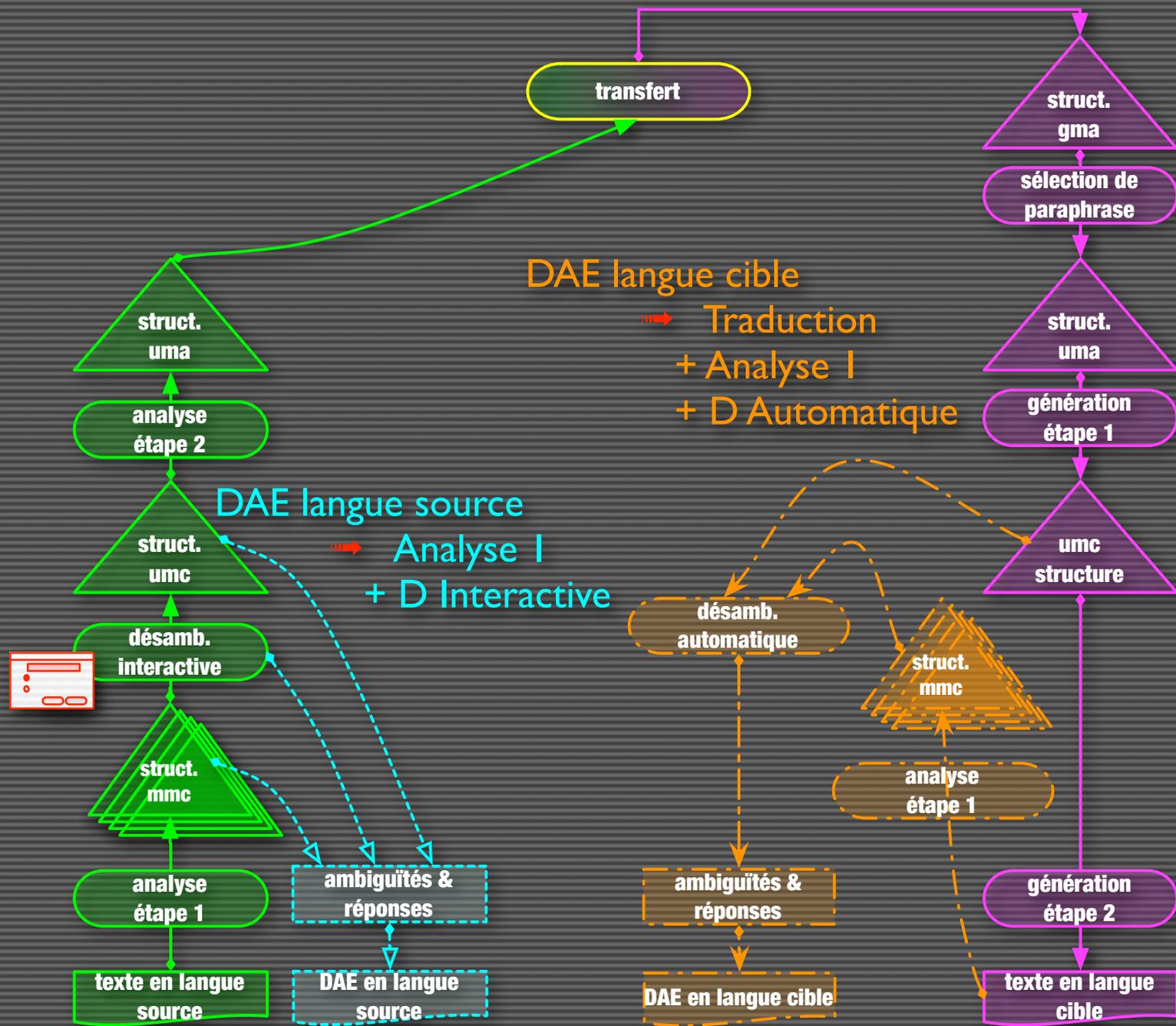
[Schonek 03]

[Romier 04]

[Choumane 05]

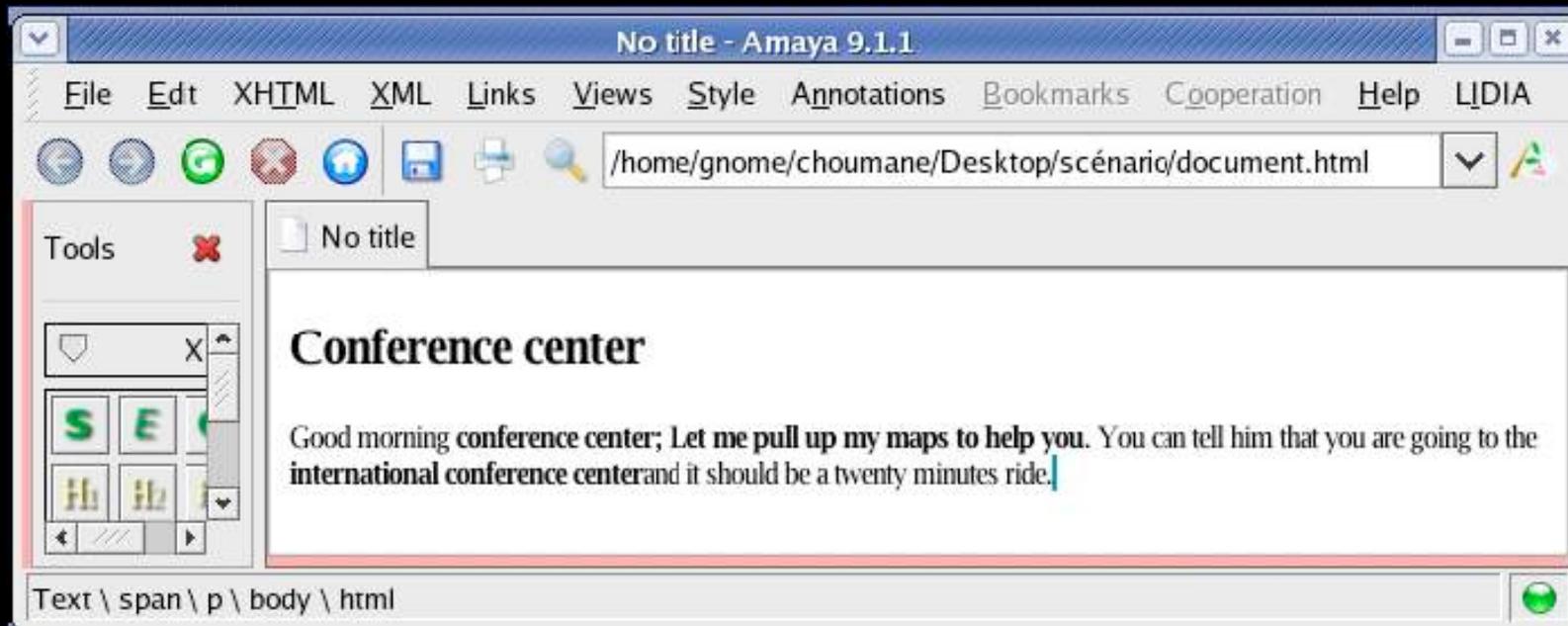






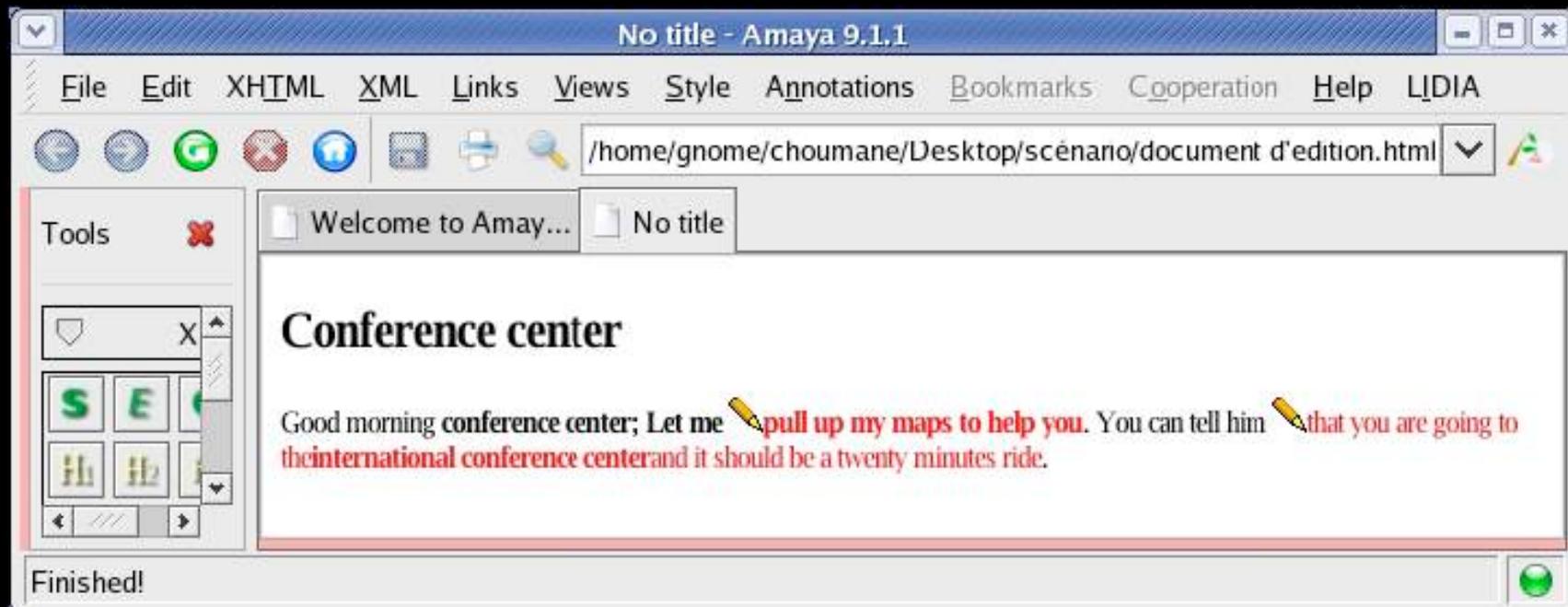
LIDIA-3 : Démonstration (1/6)

- Document édité

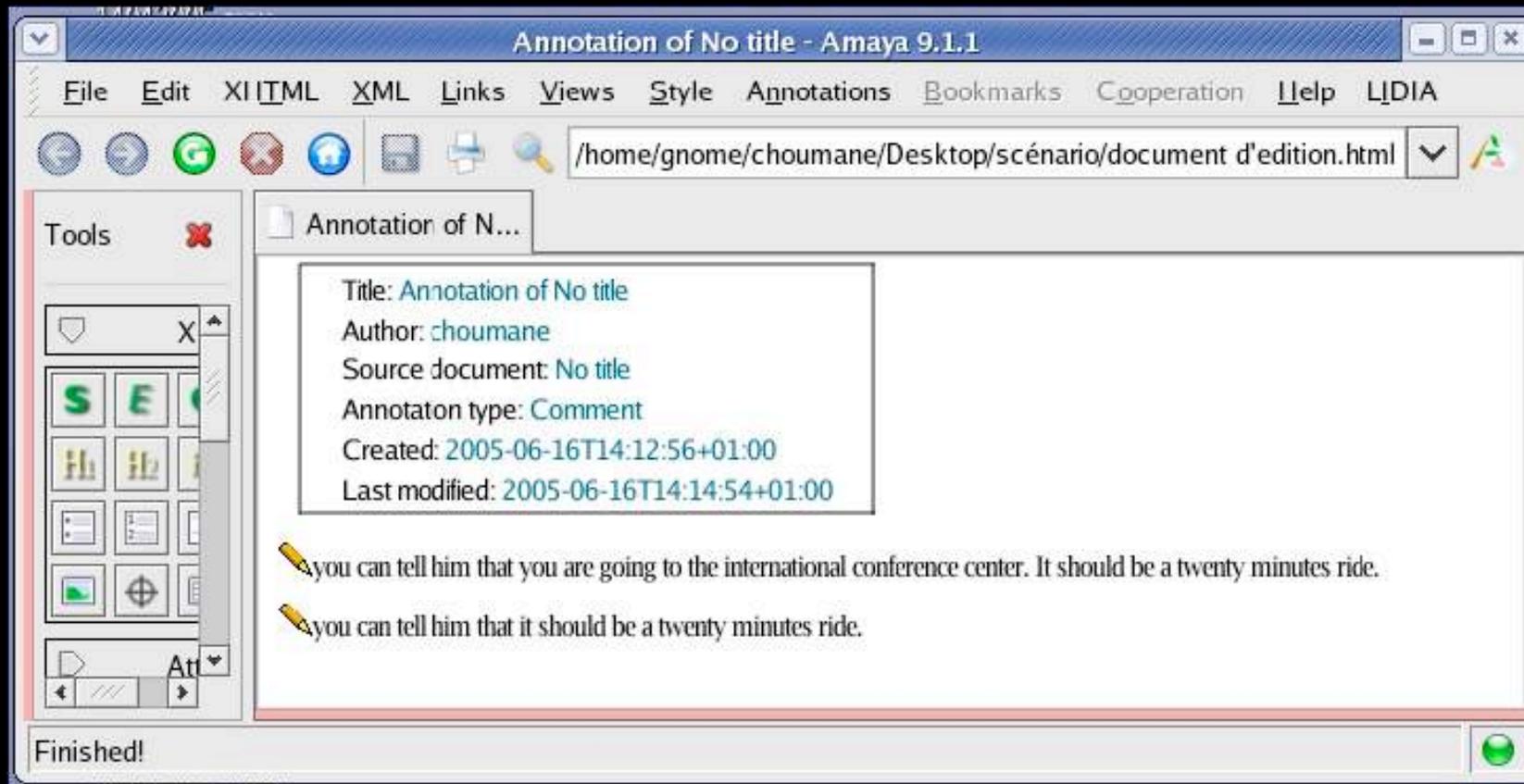


LIDIA-3 : Démonstration (2/6)

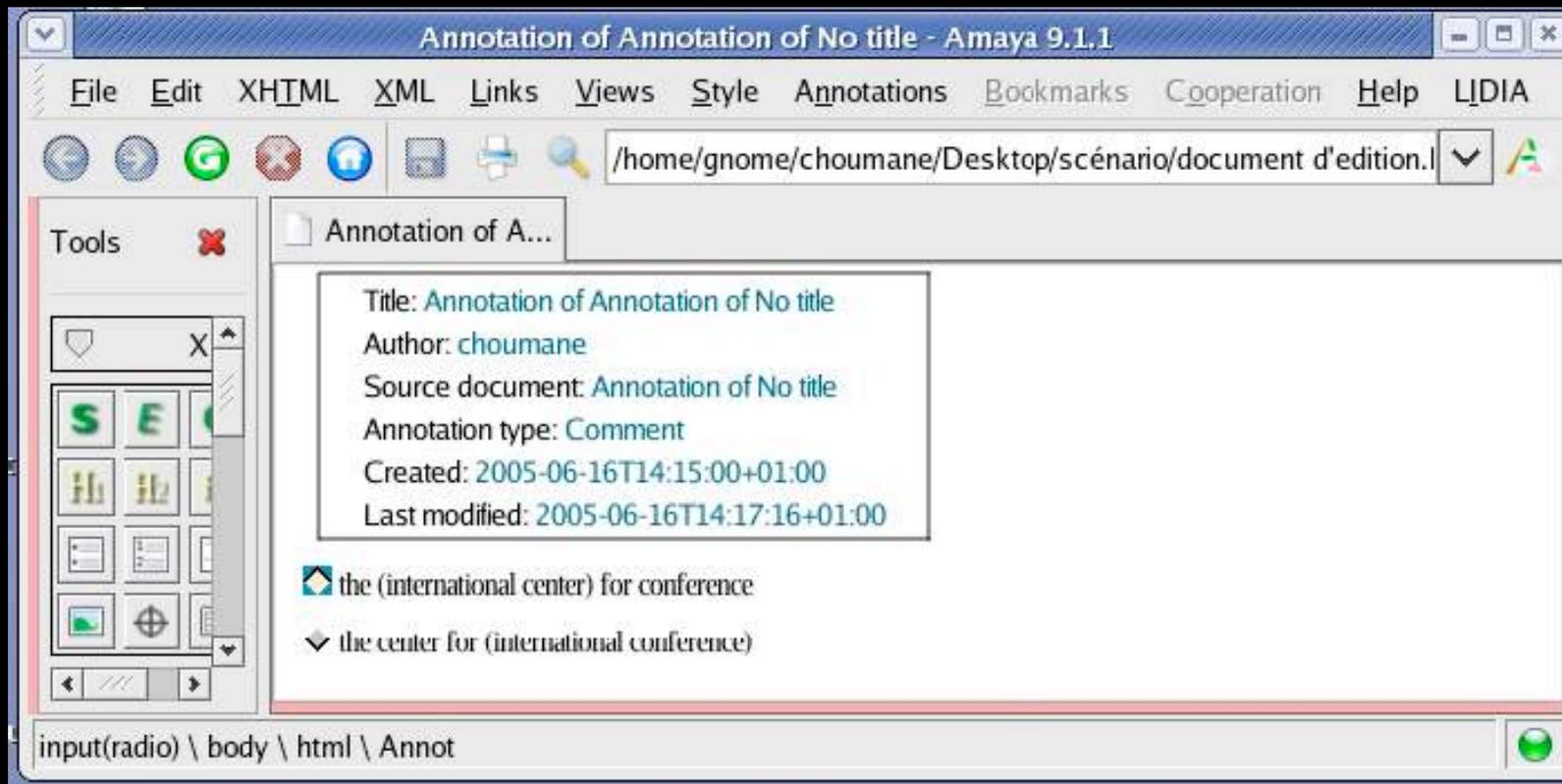
- Document édité annoté



LIDIA-3 : Démonstration (3/6)

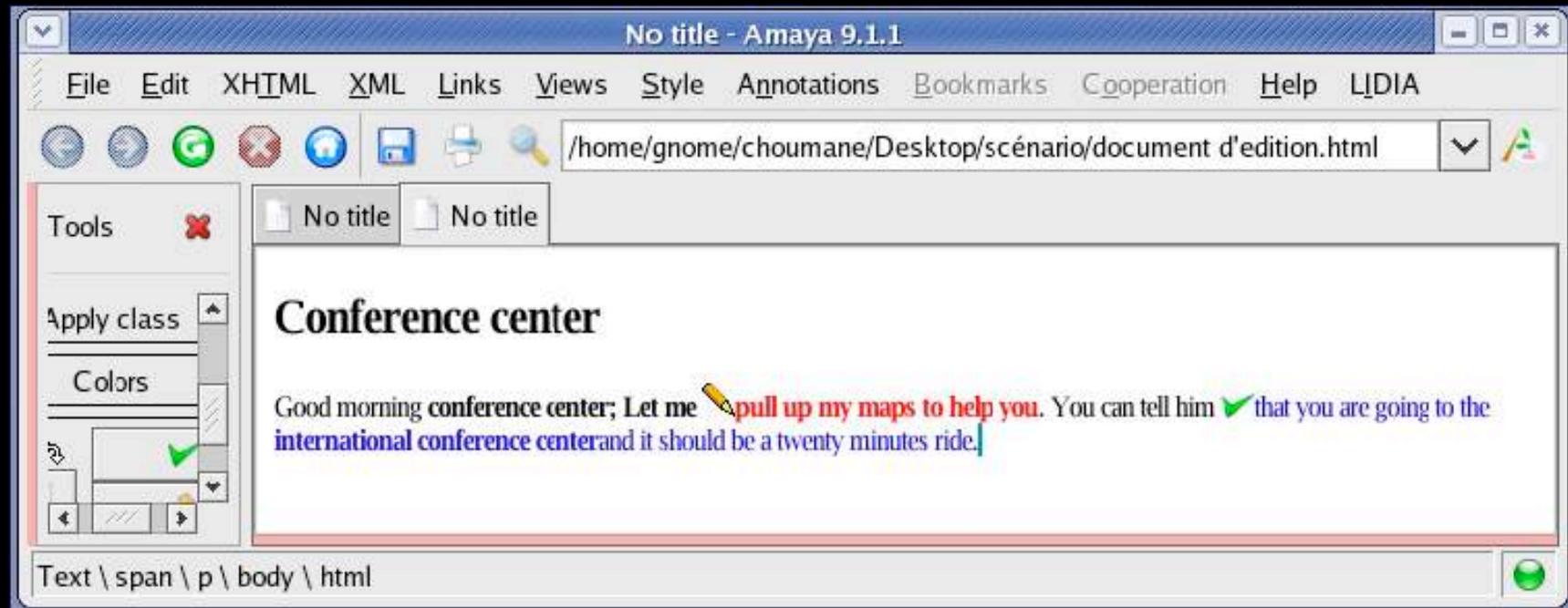


LIDIA-3 : Démonstration (4/6)



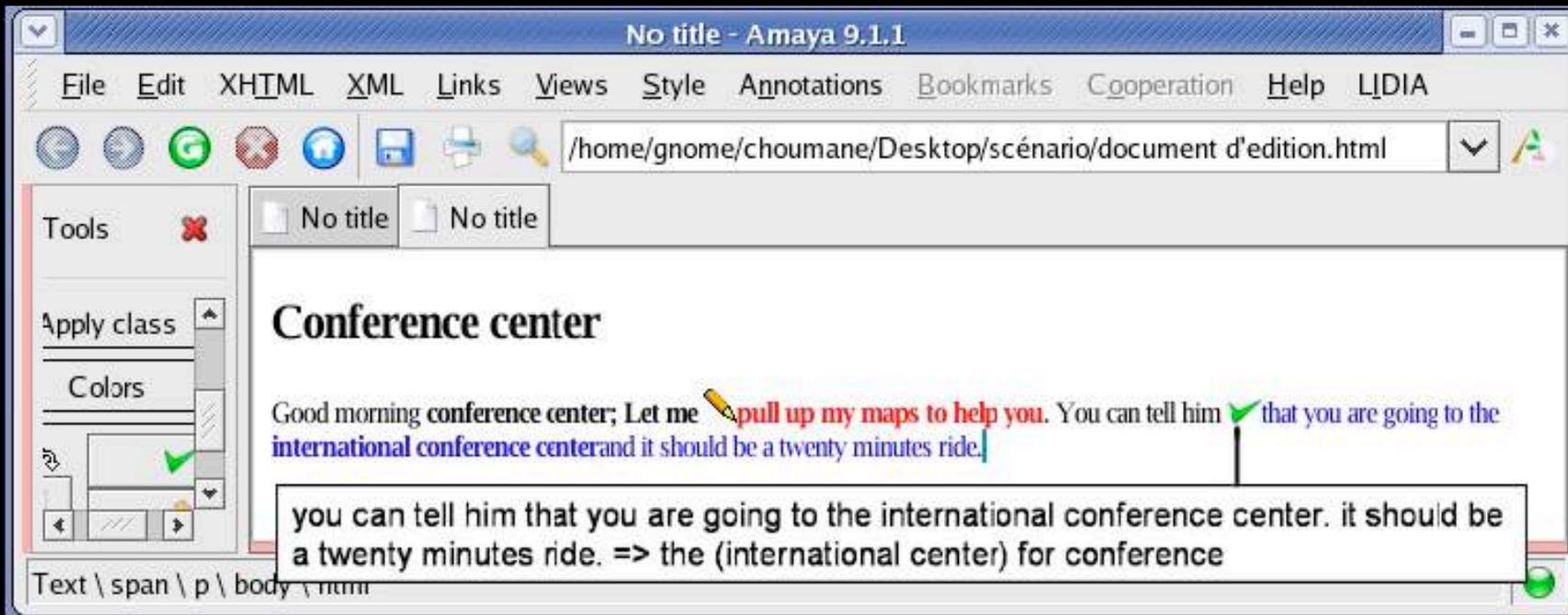
LIDIA-3 : Démonstration (5/6)

- Document édité mis à jour



LIDIA-3 : Démonstration (6/6)

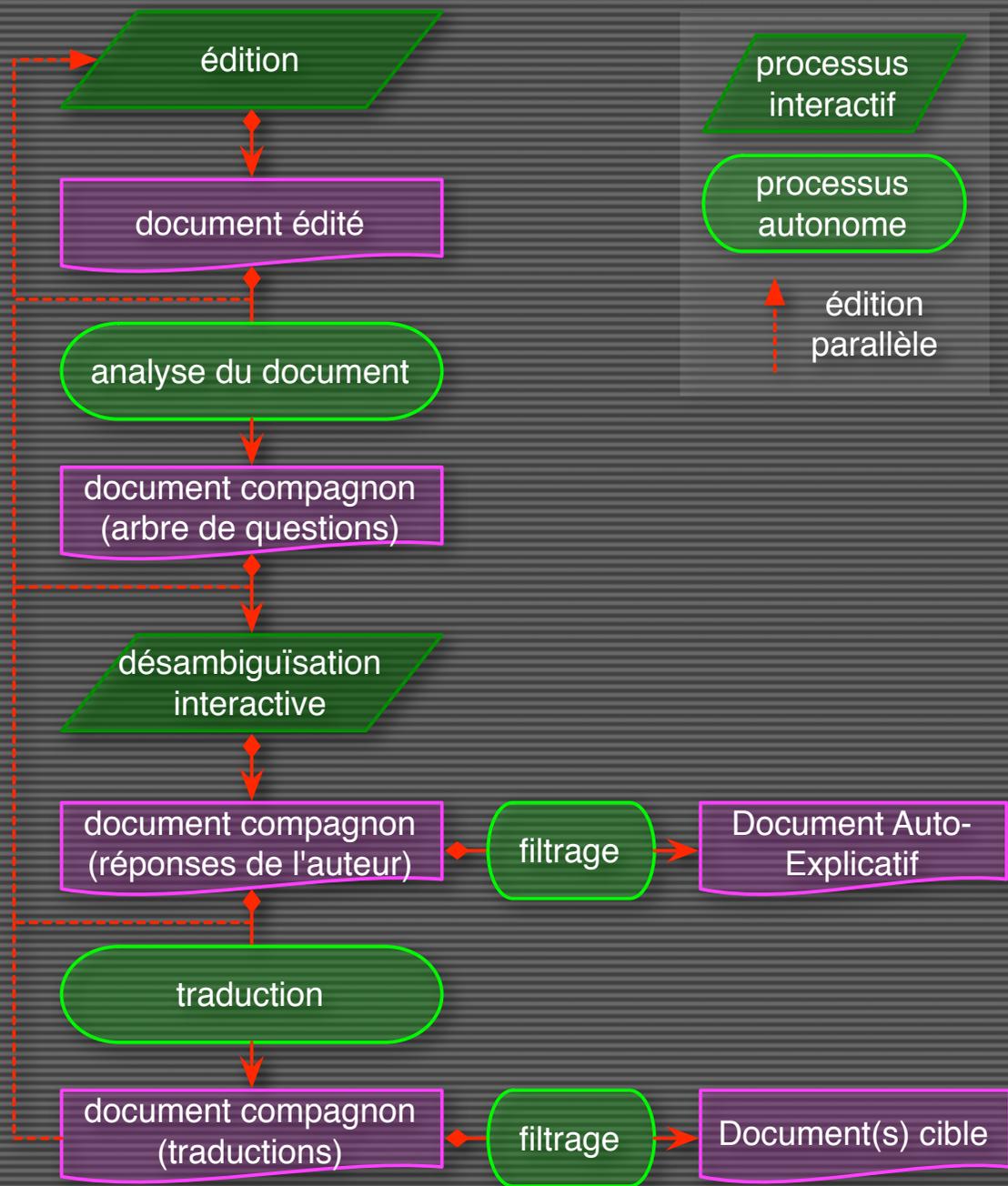
● Présentation des explications



The screenshot shows the Amaya 9.1.1 web editor window. The title bar reads "No title - Amaya 9.1.1". The menu bar includes File, Edit, XHTML, XML, Links, Views, Style, Annotations, Bookmarks, Cooperation, Help, and LIDIA. The address bar shows the file path: /home/gnome/choumane/Desktop/scénario/document d'edition.html. The main editing area displays the text "Conference center" followed by a paragraph: "Good morning conference center; Let me pull up my maps to help you. You can tell him that you are going to the international conference center and it should be a twenty minutes ride." The text "pull up my maps to help you" is highlighted in red, and "international conference center" is highlighted in blue. A green checkmark is placed above the word "that". A tooltip box is open over the word "that", containing the text: "you can tell him that you are going to the international conference center. it should be a twenty minutes ride. => the (international center) for conference". The status bar at the bottom shows the path: Text \ span \ p \ body \ rrrrr.

LIDIA 3 : implémentation *(1/7)*

- Scénario implémenté



LIDIA-3 : implémentation *(2/7)*

- Document compagnon **avant analyse**

LIDIA-3 : implémentation (3/7)

- Document compagnon **après analyse**

LIDIA-3 : implémentation (4/7)

LIDIA-3 : implémentation (5/7)

LIDIA-3 : implémentation *(6/7)*

- Synchronisation entre documents édité et document compagnon

LIDIA-3 : implémentation (7/7)

- Préparation des questions et des explications
 - Une question
 - représentée dans un fichier XHTML affiché
 - Une explication
 - préparée à la volée
 - paragraphes sélectionnés par l'auteur
 - affichée dans un cadre

Verrous et pistes pour le futur

Obtenir un module de DI couvrant

[Frey 96] &

 DI construisant des descripteurs d'ambiguïté *[Irgadian 02]*

 DI sans descripteur d'ambiguïté (DI par analogie)

Réduire le coût de la DI

 DI qui apprend le “style” de l’auteur

 Est-ce que l’auteur ne peut pas aussi apprendre à rédiger ?

 DI incomplète & certification du sens

Produire des DAE plus utiles

 Présenter le support de l’ambiguïté



Automatiser la construction

- Construire automatiquement les descripteurs d'ambiguïté
 - N'utiliser que le support pour plus de généralité
 - *[Frey 96]*
- Proposer une première ébauche des méthodes de construction des items de question
 - *[Irgadian 02]*



Apprendre le style de l'auteur

- Au début

- Ordre psycholinguistique des items de question
- Équirépartition du poids des items

-  En cours d'usage (apprentissage)

- Items ordonnés par fréquence de choix
 - “Style” de l'auteur
 - Heuristique personnalisée
- Si le poids d'un item dépasse un certain seuil pas de question



Autoriser une DI incomplète

- Il faudrait ...
 - Produire une traduction unique
 - ... utilisation de préférences et heuristiques
 - Produire une traduction multiple factorisée
- 💡 On pourrait alors certifier de sens
 - ... en fonction des ambiguïtés non résolues

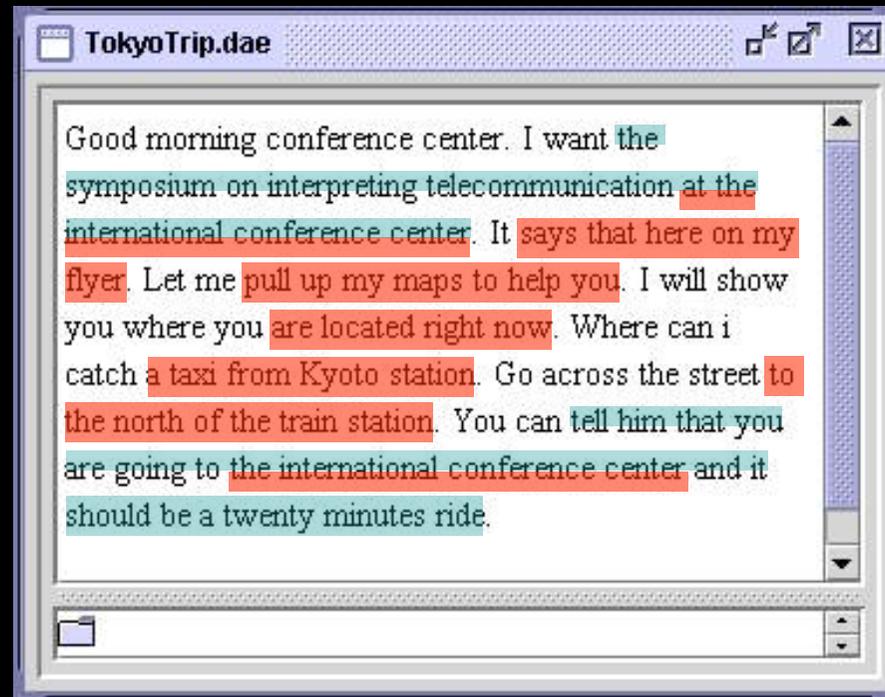


Support d'ambiguïté & DAE



Localiser précisément les ambiguïtés

- Fournir des informations pertinentes au lecteur





Support d'ambiguïté & DAE



Localiser précisément les ambiguïtés

- Fournir des informations au lecteur

Nécessite le support des ambiguïtés :

- nouveaux descripteurs ?*
- descripteurs & support ?*
- calcul du support ?*

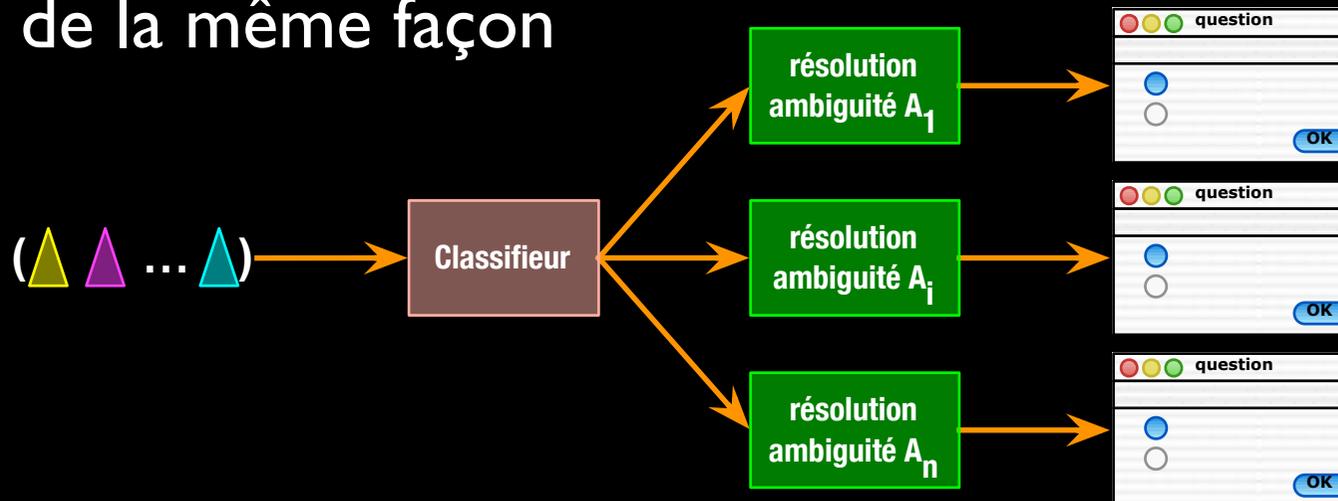


DI par analogie

💡 Se passer d'une description explicite

- ... des ambiguïtés, et
- ... de la manière de poser les questions

💡 Les ambiguïtés d'une même classe se résolvent de la même façon





Merci !

Commentaires & Questions