

Comment définir, mesurer et améliorer  
la qualité, l'utilisabilité et l'utilité  
des systèmes de TAO de l'écrit et de l'oral ?

Une bataille contre le bruit, l'ambiguïté et le manque de contexte



**Hervé Blanchon**

Habilitation à Diriger des Recherches

20 décembre 2004

# Défis

# Défis

- À l'écrit

Permettre à un auteur monolingue de diffuser ses documents dans une ou plusieurs langues

- À l'oral

Permettre à un locuteur de converser dans sa langue maternelle avec un interlocuteur dont il ne connaît pas la langue

# moyens

- Des systèmes ...
  - utilisables
  - utiles
- ... de traduction automatique pour la **diffusion** ...
  - de l'écrit
  - de l'oral
- ... qui produisent des traductions de qualité

# problèmes

⚠ Il existe beaucoup d'obstacles à la qualité !

- Le bruit “physique” ⇒ capture du signal sonore
- Le bruit “syntaxique” ⇒ énoncés non conformes à la “norme”, réalisations “perturbées” de l'intention
  - Oral : “je veux heu une heur non deux chambres”
  - Écrit en tchat : “il y a klkun ou sui tute seulllllllll”
- L'ambiguïté ⇒ énoncés polysémiques
  - Oral & écrit

# But de mes travaux

## ⚠ Comment ...

- ... lutter contre ces obstacles ?
  - le **bruit** à l'oral
  - l'**ambiguïté** à l'écrit
- ... mesurer qualité, utilité et utilisabilité ?

# Plan de l'exposé

- Contextes
- *Première partie*
- *Deuxième partie*
- *Troisième partie*
- Conclusion
  - Contributions
  - Réflexions

# Deux contextes ...

- Traduction Automatique Fondée sur le Dialogue

- Cadre local : projet du GETA

- Projet : LIDIA 

- Traduction Automatique de Dialogues Oraux Finalisés

- Cadre international : Japon, États Unis, Allemagne, Corée du Sud, Italie, France

- Projets : C-STAR II, NESPOLE!, C-STAR III



... un domaine ...

- ... qui font finalement partie d'**un même domaine**

**La Communication Interpersonnelle Multilingue  
Médiatisée par la Machine**

## ... un domaine ...

- ... qui font finalement partie d'**un même domaine**

La Communication Interpersonnelle Multilingue  
Médiatisée par la Machine

- ... qui présentent **des caractéristiques opposées**

	Modalité	Message	Domaine
LIDIA	Écrit	Réfléchi	Tout
C-STAR & NESPOLE!	Oral	Spontané	Spécialisé

# ... un nouveau vocabulaire

Communication Interpersonnelle multilingue  
**Écrite Réfléchie Tout Terrain**  
médiatisée par la machine

- *Traduction Automatique Fondée sur le Dialogue*
- Nécessité : **lutter contre l'ambiguïté**

# ... un nouveau vocabulaire

Communication Interpersonnelle multilingue **Écrite**  
**Réfléchie Tout Terrain** médiatisée par la  
machine

- *Traduction Automatique Fondée sur le Dialogue*
- Nécessité : **lutter contre l'ambiguïté**

Communication Interpersonnelle multilingue  
**Orale Spontanée Spécialisée**  
médiatisée par la machine

- *Traduction Automatique de Dialogues Oraux Finalisés*
- Nécessité première : **lutter contre le bruit**

# Lutte contre l'ambiguïté

## Analyse de l'écrit

- Usage en analyse de l'écrit
  - Ne produire qu'une analyse, *interprétation*
    - Heuristiques (psycholinguistique, sémantique, corpus)
- ⚠ Risque
  - Ne pas produire la bonne analyse, *interprétation*
    - Une traduction qui ne transmet pas l'intention de l'auteur !

# Lutte contre l'ambiguïté

## Analyse de l'écrit

### Proposition

- Produire toutes les analyses, *interprétations*
  - Pas d'heuristique, ou très peu
- Poser des questions de **désambiguïstation** à l'auteur

*Si l'analyse est bonne alors la traduction est bonne*

- Évaluation de l'**utilisabilité**, de l'**utilité**
  - Vérifier que l'auteur comprend les questions et y répond correctement

# Lutte contre le bruit

## Analyse de l'oral

- Usage en analyse de l'oral
  - Relaxer les contraintes de l'analyseur
  - *en TA statistique ou par l'exemple la question ne se pose pas*

### Risque

- Ne pas produire de traduction
  - **<ref> je pense partir avec mes enfants et ma femme et donc j aimerais avoir des renseignements sur cette région**
  - **<hypo> j attache partir avec mes enfants il installe ce qui est en chat le savoir faire enseignement chez cette région**

# Lutte contre le bruit

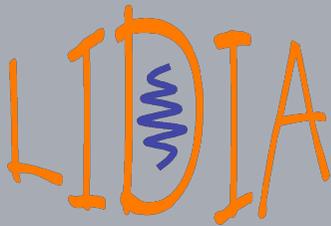
## Analyse de l'oral

### Proposition

- Chercher des **segments textuels pertinents**
  - Proposition fortement liée au résultat d'analyse
- Évaluation de la **qualité** des traductions
  - Vérifier que les traductions sont bonnes

# CI écrite réfléchie tout terrain

Première  
partie



LIDIA

1989-

- Introduction
- Lutte contre l'ambiguïté
  - Proposition
  - Évaluation
- Documents Auto-Explicatifs
  - Proposition
  - Exemple
- Verrous & pistes pour le futur

# Introduction

- **Projet LIDIA** → **la TAO de l'auteur monolingue**
  - Rédiger dans sa propre langue et faire traduire
    - sans utiliser de langage “contrôlé” (lexique, syntaxe)
    - sans devoir réviser les traductions
    - sans compréhension explicite
    - sans heuristique

# Introduction

- **Projet LIDIA** → **la TAO de l'auteur monolingue**

- Rédiger dans sa propre langue et faire traduire
  - sans utiliser de langage “contrôlé” (lexique, syntaxe)
  - sans devoir réviser les traductions
  - sans compréhension explicite
  - sans heuristique

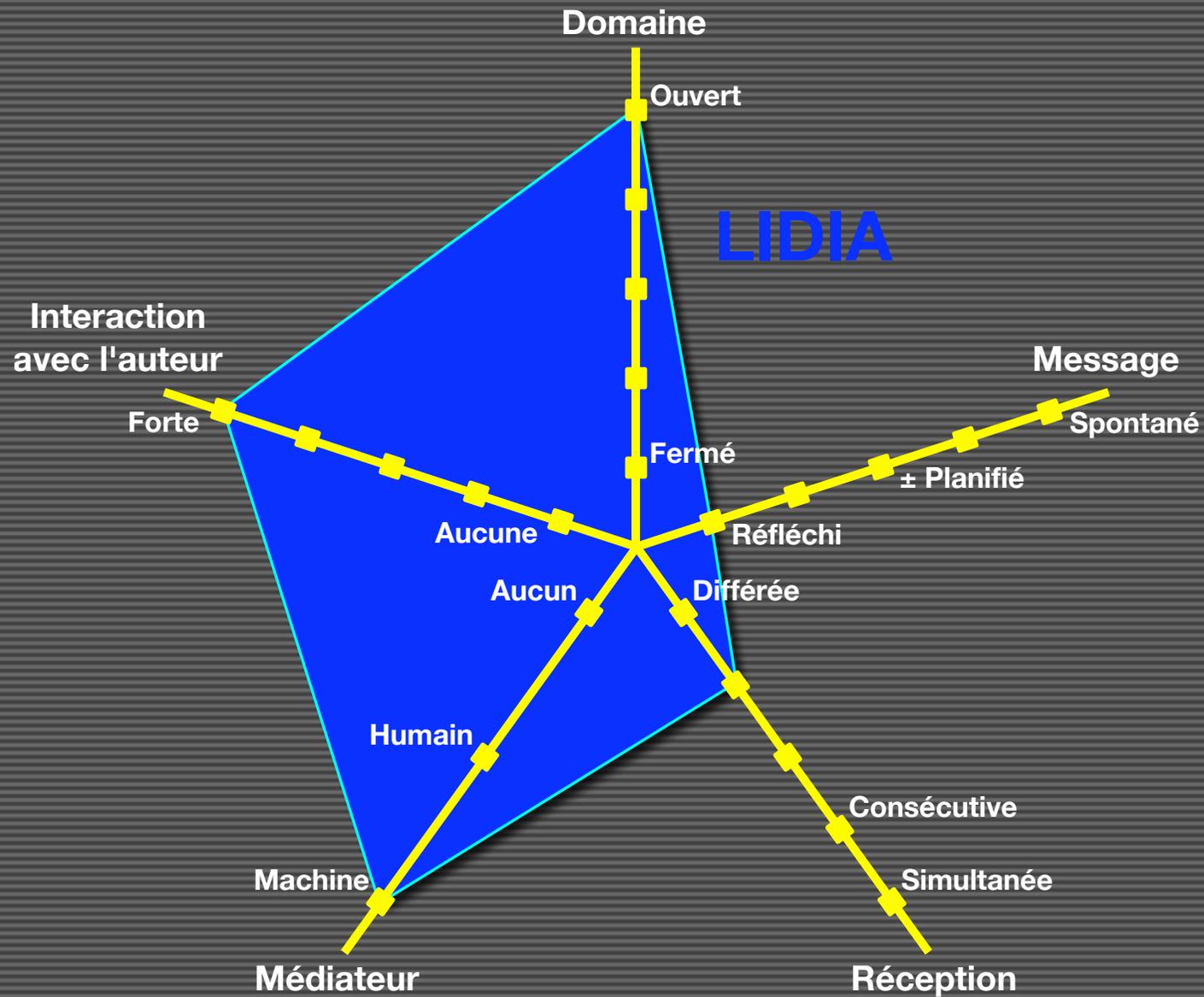
⚠ Toutes les ambiguïtés sont *a priori* possibles

- Comment lutter ?

- **La désambiguïtation interactive**

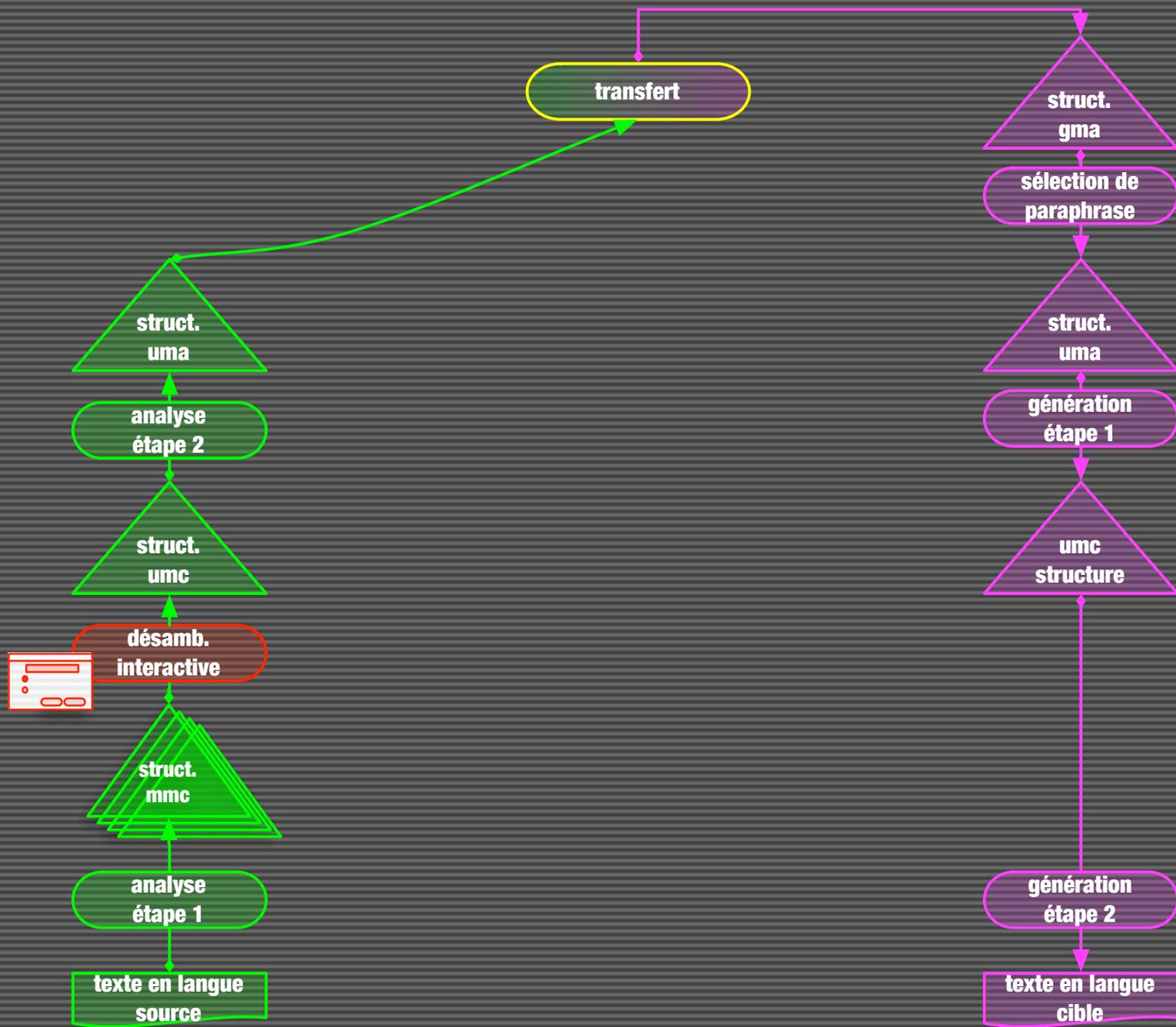
- L'auteur est le mieux placé pour savoir ce qu'il veut dire !

# communication interpersonnelle écrite réfléchie tout terrain

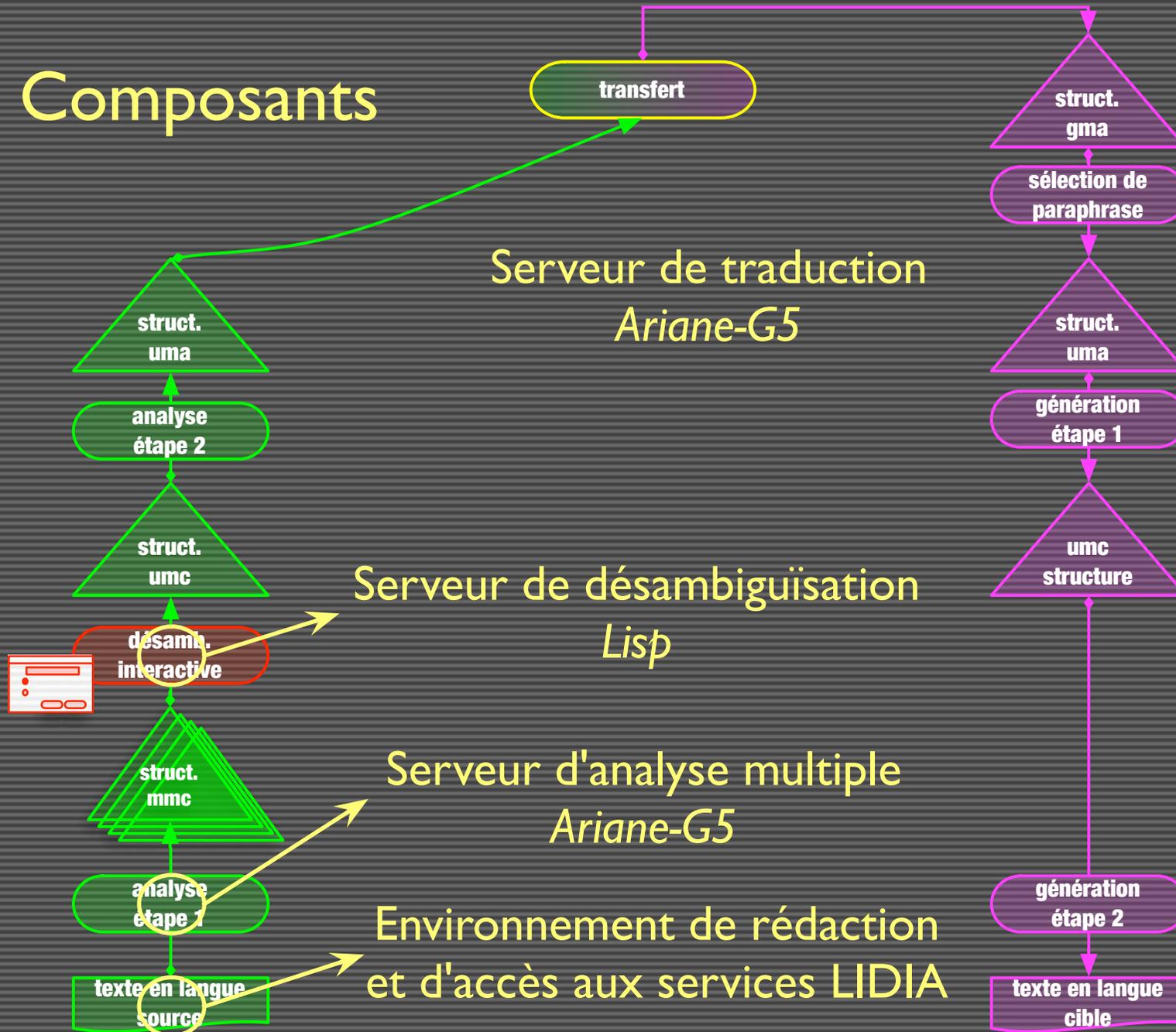


# Introduction

- Architecture linguistique
  - Transfert multiniveau à acceptations, propriétés et relations interlingues
- Structures manipulées
  - Des arbres



# Composants



# Introduction

## ● Bilan

### ● 4 maquettes

#### ● LIDIA-1.3270 & LIDIA-1.mail

[Blanchon 94]

[Lafourcade, Sérasset,  
Gaschler & Planas 94]

[Grasson 96]

#### ● LIDIA-2

[Belkhatir 01] [Ghorayeb 02] [Gressard 03]

#### ● LIDIA-3 (en cours)

[Romier 04] [Choumane 05]

### ● 5 linguistes

#### ● Analyse du français

Axtmeyer, Blanc, Guilbaud,  
Levenbach, Nédobejkine

#### ● Génération anglais, allemand, russe

### ● 2 informaticiens

Guillaume, Quezel-Ambrunaz

#### ● Accès aux services de traduction

# Introduction

## ● Bilan

● 4 ma

● LIDI

● LIDI

● LIDI

● 5 lingu

● Analyse

● Génération

● 2 informaticiens

● Accès aux services de traduction

*vers*

*- un environnement de  
rédaction portable.  
- des documents XML.  
- 1 protocole unique de  
communication entre  
composants.*

[Blanchon 94]

Lafourcade, Sérasset,  
Faschler & Planas 94]

[Grasson 96]

b 02] [Gressard 03]

04] [Choumane 05]

Blanc, Guilbaud,  
Levenbach, Nédobejkine

Guillaume, Quezel-Ambrunaz

# Désambiguïsation interactive (DI)

 Qu'est-ce qu'une ambiguïté ?

 Définition linguistique non opératoire

*[Fuchs 96]*

 **Définition formelle**

*[Blanchon & Boitet 04]*

# Désambiguïsation interactive

## Qu'est-ce qu'une ambiguïté ?

- Définition linguistique non opératoire

[Fuchs 96]

-  Définition formelle

[Blanchon & Boitet 04]

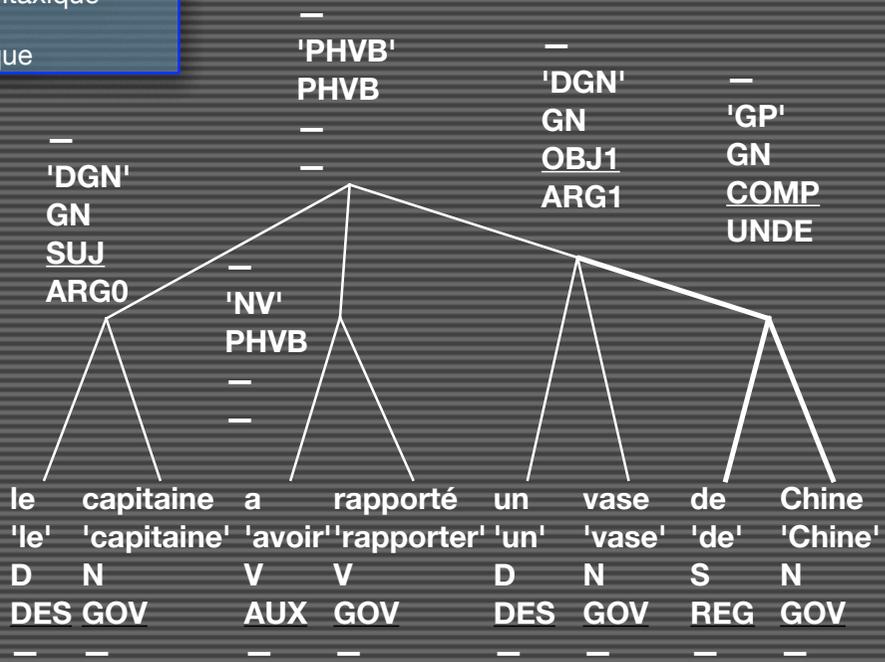
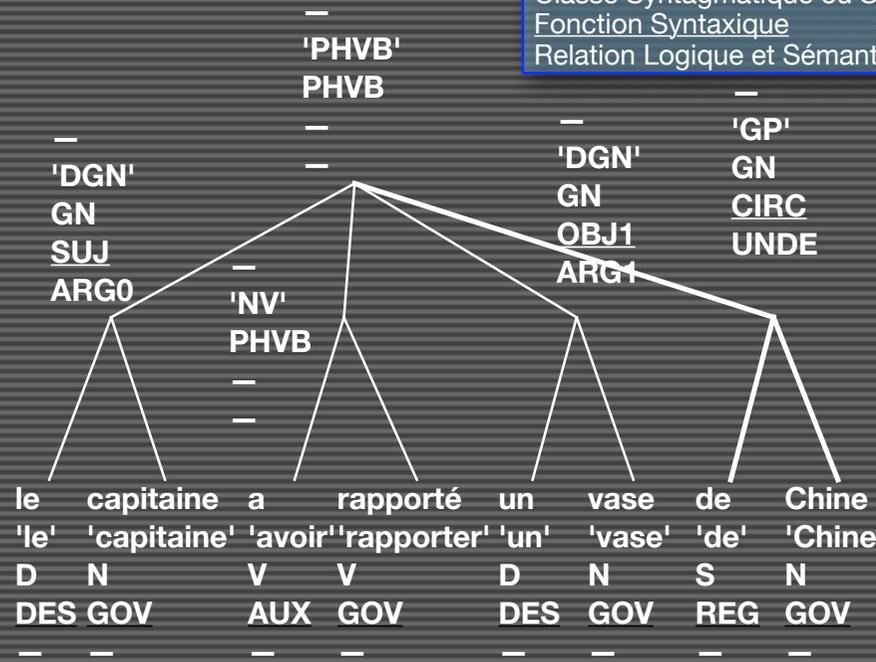
## Comment représenter une ambiguïté ?

[Blanchon 95]

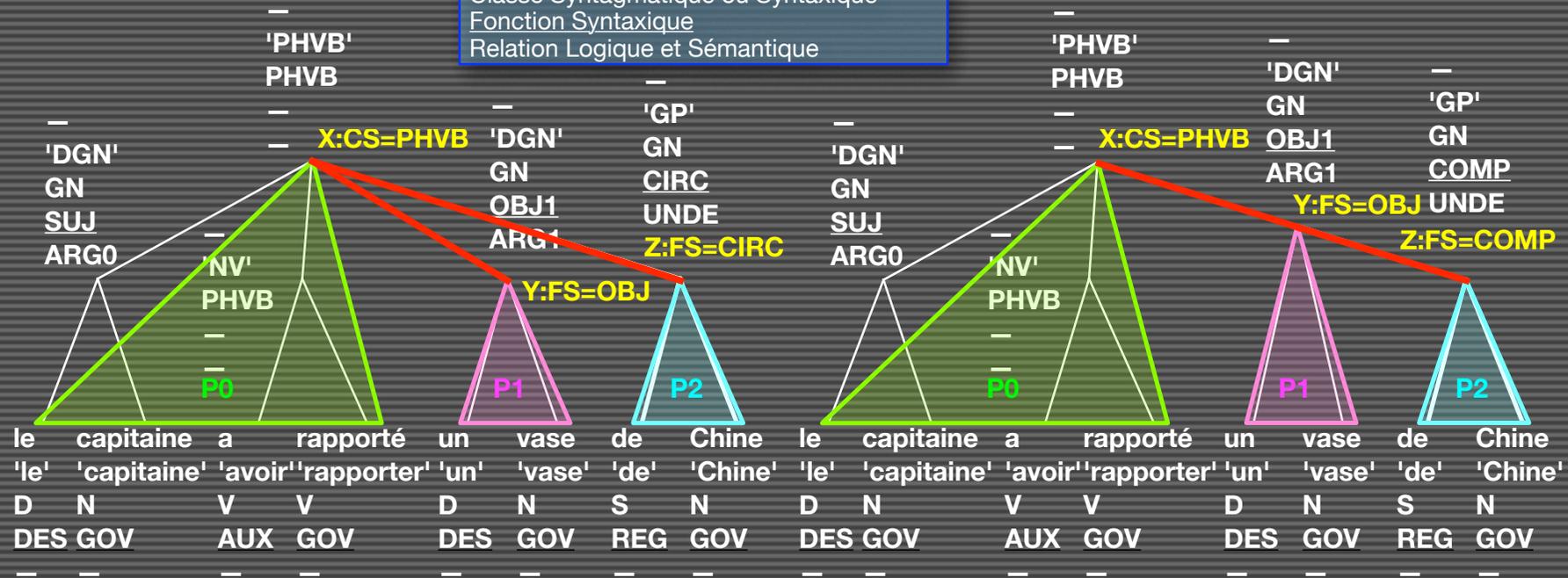
- Représentation non liée à un analyseur donné

-  **Ambiguïté = co-occurrence au sein d'un ensemble de solutions de plusieurs patrons "synchronisés" groupés en un faisceau**

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



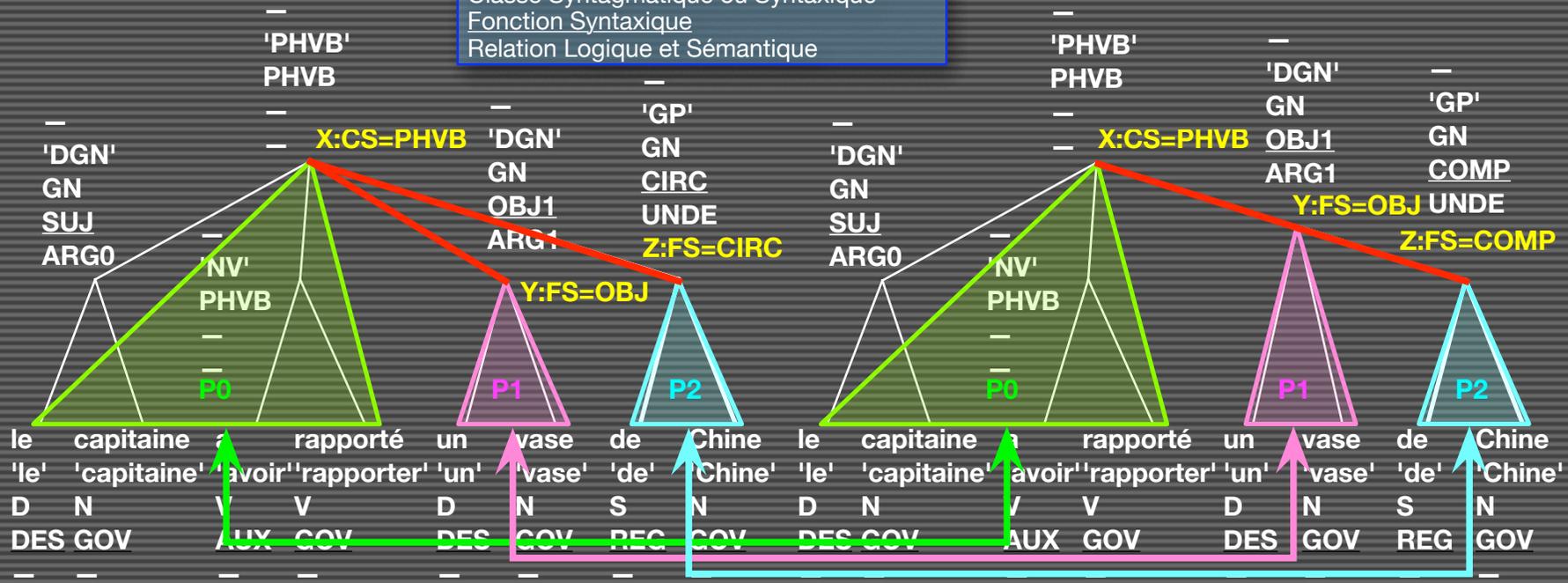
occurrence  
 Unité Lexicale  
 Classe Syntagmatique ou Syntaxique  
 Fonction Syntaxique  
 Relation Logique et Sémantique



### Un patron = un schéma d'arbre

- des variables de nœud (X, Y, Z)
- des variables de forêt (P0, P1, P2)

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



**Synchronisation ⇔ égalité du mot des feuilles des variables de forêt  
 partagées entre patrons d'un faisceau**

# Désambiguïsation interactive

⚠ Comment poser une question ?

⚠ Il ne faut pas que la question soit ambiguë !

# Désambiguïsation interactive

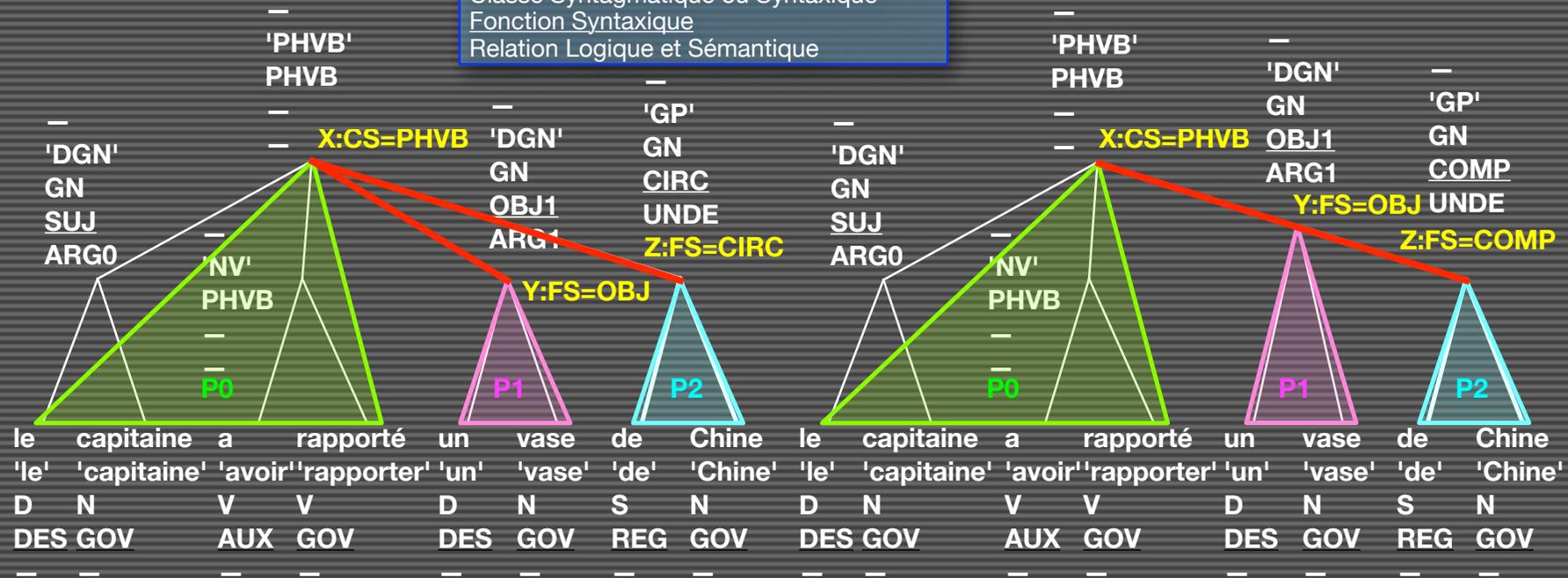
⚠ Comment poser une question ?

⚠ Il ne faut pas que la question soit ambiguë !

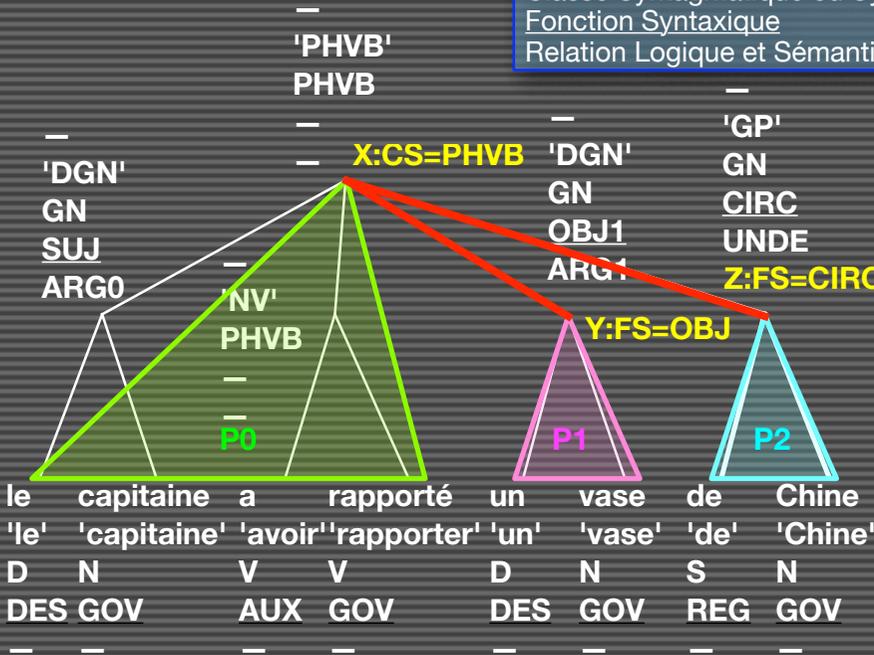
💡 **Faire un rephrasage contrôlé avec  $\neq$  opérateurs**

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

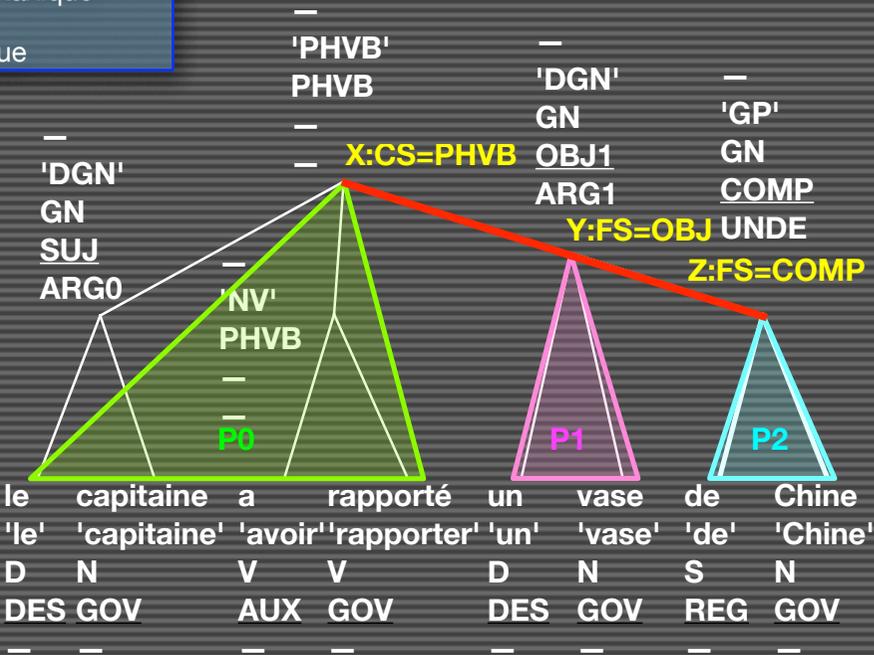
occurrence  
 Unité Lexicale  
 Classe Syntagmatique ou Syntaxique  
 Fonction Syntaxique  
 Relation Logique et Sémantique



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

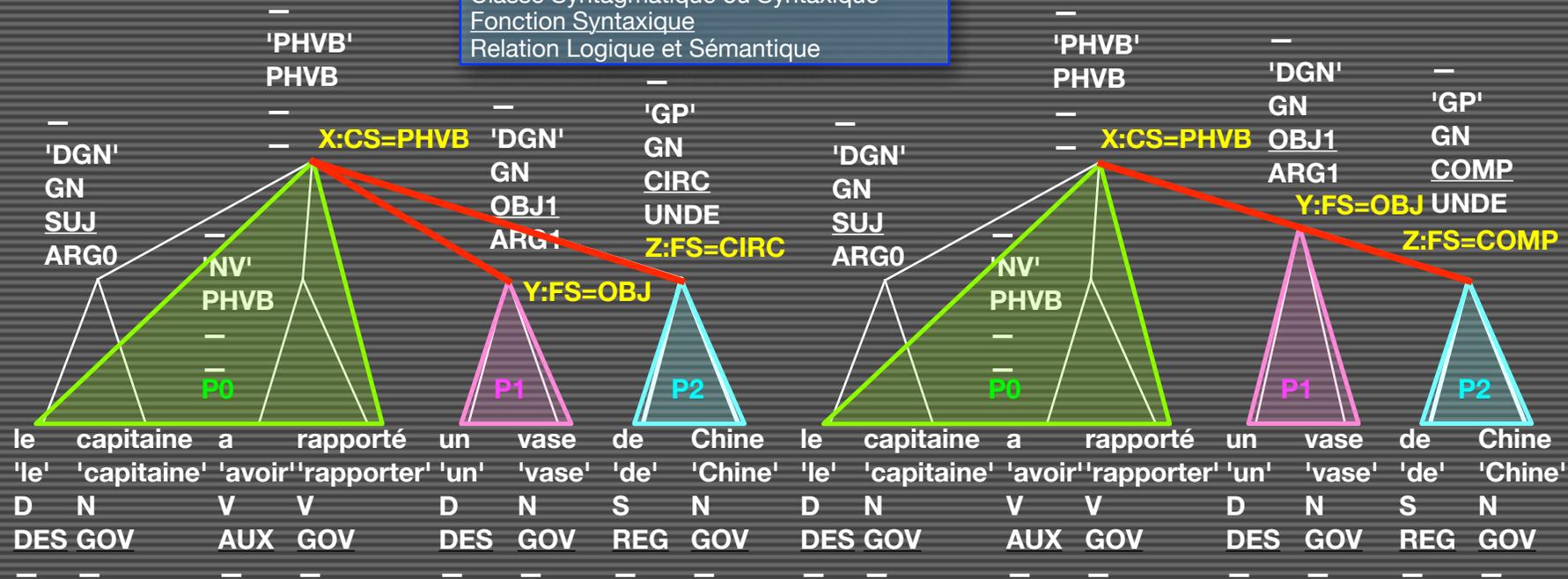


Text(P2), Text(P0) Text(P1)



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

**ambiguïté**

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

Choisissez la bonne.

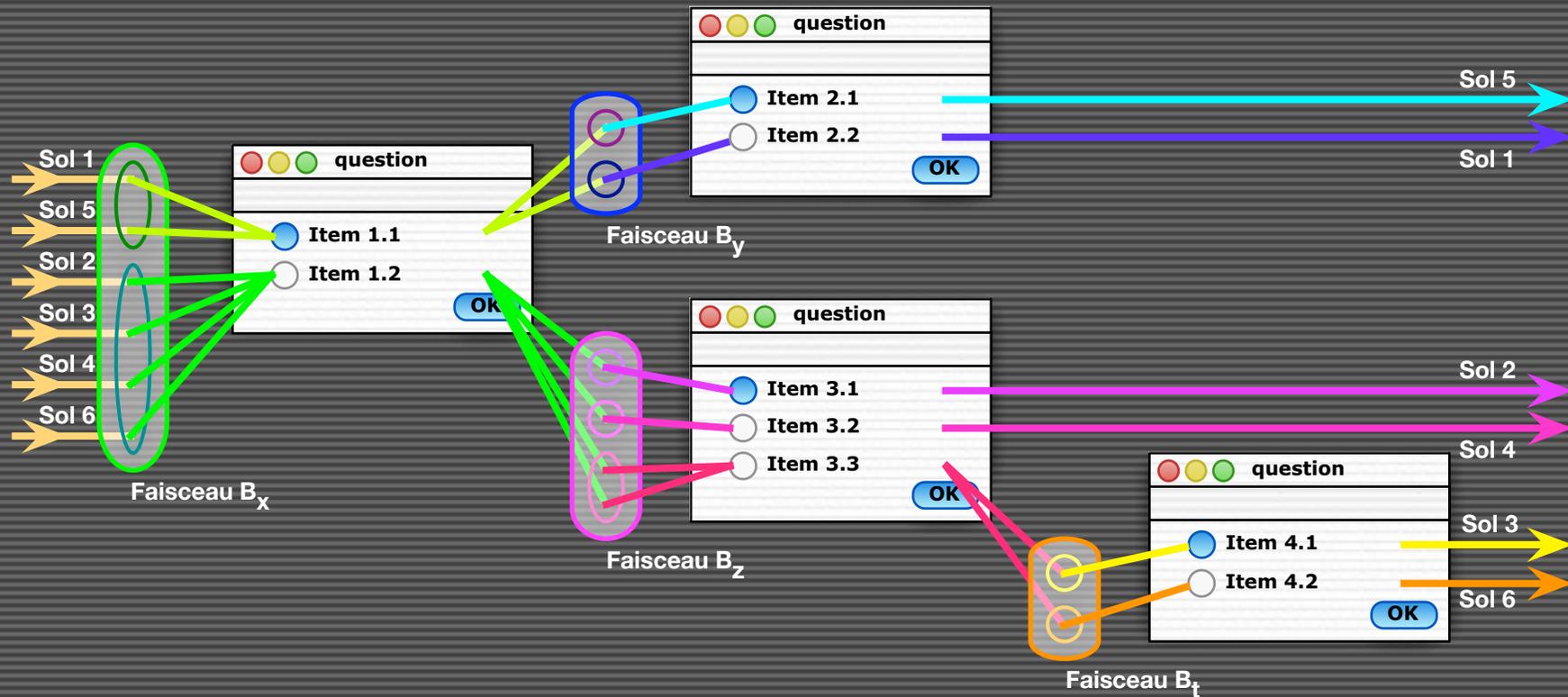
De Chine, le capitaine a rapporté un vase.

Le capitaine a rapporté (un vase de Chine).

OK

# Arbre de questions

Partitions successives de l'ensemble des solutions



# Désambiguïsation interactive

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

- d'un **linguiciel** particulier adapté à une application donnée
  - description des ambiguïtés
  - description de la construction des items de question
  - description d'une stratégie de désambiguïsation
    - ordre des questions

[Blanchon 95, pour l'anglais]

[Grasson 96, pour LIDIA]

● d'un **moteur**

- reconnaissance des ambiguïtés
- opérateurs de production des items de question

[Blanchon 95]

# Désambiguïsation interactive

💡 Un module de DI est l'

- d'un langage de description d'application donnée
  - description des données
  - description des opérations
  - description des relations
  - ordre de priorité
- d'un moteur de questionnement
  - reconnaissance des ambiguïtés
  - opérateurs de production des items de question

Préoccupation de  
l'informaticien.

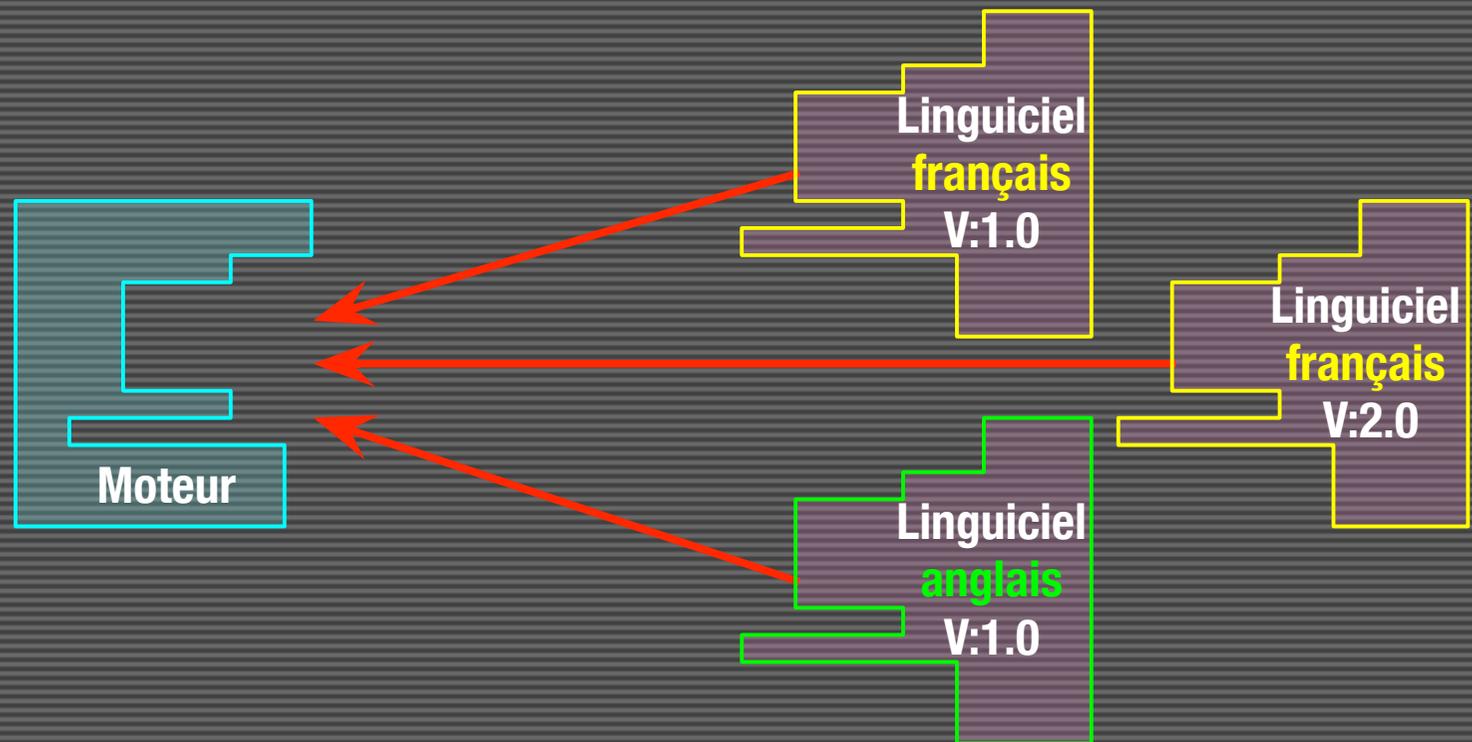
Séparer les données  
des programmes :  
- Réutilisabilité,  
- Généricité.

[Blanchon 95, pour l'anglais]  
[Blanchon 96, pour LIDIA]

[Blanchon 95]

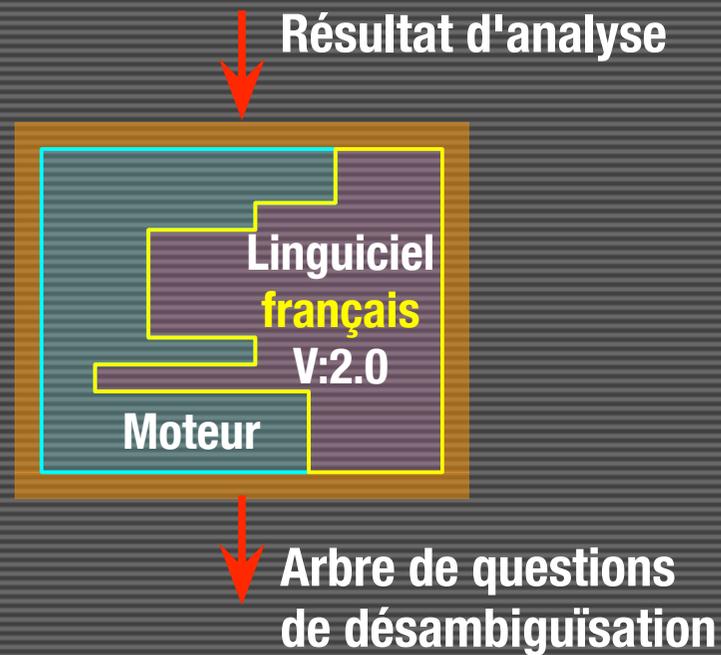
# Désambiguïsation interactive

Principe de développement



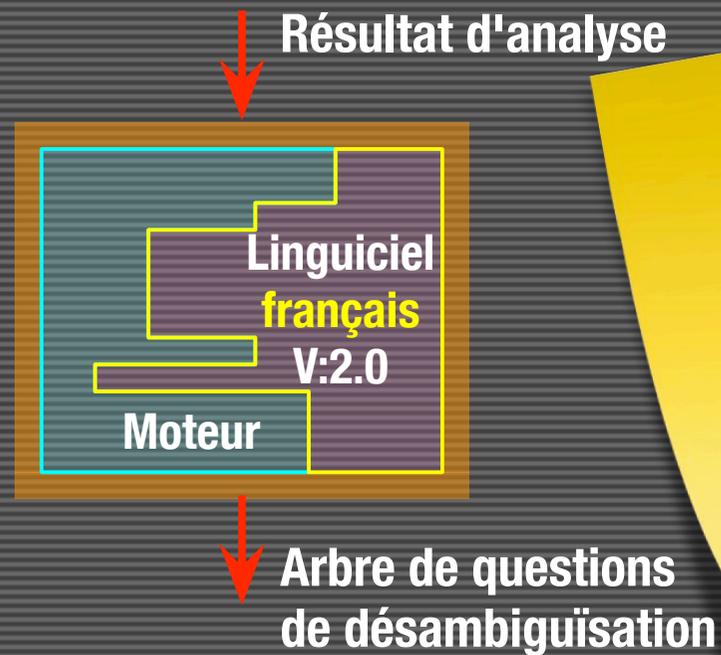
# Désambiguïsation interactive

Instanciation d'un module particulier



# Désambiguïsation interactive

Instanciation d'un module particulier



Résultats :

- un moteur
- 2 linguiciels
  - français LIDIA
  - anglais (réduit)

# Évaluation en utilisabilité et utilité

- *MIDDIM (ATR-CNRS, 3 ans)*
  - *Bases de données, annotation, ...*

*Boitet, Fafiotte,  
Tomokiyo, Winship*

# Évaluation en utilisabilité et utilité

- *MIDDIM (ATR-CNRS, 3 ans)*
  - *Bases de données, annotation, ...*

Boitet, Fafiotte,  
Tomokiyo, Winship

- **Seconde expérience**

[Blanchon & Fais 96]

- 60 sujets en 4 groupes
- 2 types de questions : humaines (H) & machines (M)
- 2 modalités : orales (O) & écrites (E)

# Évaluation en utilisabilité et utilité

- *MIDDIM (ATR-CNRS, 3 ans)*

- *Bases de données, annotation, ...*

Boitet, Fafiotte,  
Tomokiyo, Winship

- *Seconde expérience*

[Blanchon & Fais 96]

- 60 sujets en 4 groupes
- 2 types de questions : humaines (H) & machines (M)
- 2 modalités : orales (O) & écrites (E)
- **Résultats**
  - Pas de  $\neq$  significative entre les groupes Hx et Mx
  - Pas de  $\neq$  significative entre les groupes xO et xE



# Documents Auto-Explicatifs (DAE)

 Si un lecteur ne comprend pas, ou comprend mal, ce qu'il lit c'est un problème !

# Documents Auto-Explicatifs (DAE)

- ⚠ Si un lecteur ne comprend pas, ou comprend mal, ce qu'il lit c'est un problème !
- ⚠ La réponse aux questions de DI représentent le sens choisi par l'auteur !
- ⚠ Ne peut-on pas en faire profiter les lecteurs ?

# Documents Auto-Explicatifs

 Enrichir le document avec les réponses de l'auteur lors de la DI

- Présenter au **lecteur**, à **sa** demande, ...
  - ... le sens choisi par l'auteur parmi les interprétations potentielles
- Un visualiseur de DAE (analogie avec Acrobat Reader)

[Blanchon & Boitet --]

[Vigli 95]

[Ghorayeb 02]

[Choumane 05]

[Gressard 01-03]

[Schonek 03]

[Romier 04]

# Documents Auto-Explicatifs



Enrichir le document de réponses de l'auteur

- Prés

- ... le s  
inte

- Un vis

*Un concept novateur qui  
peut changer notre façon  
d'interagir avec le texte.*

*Une belle retombée  
imprévisible !*

le, ...

les

(boat Reader)

[Blanchon & Boitet --]

[Vigli 95]

[Ghorayeb 02]

[Choumane 05]

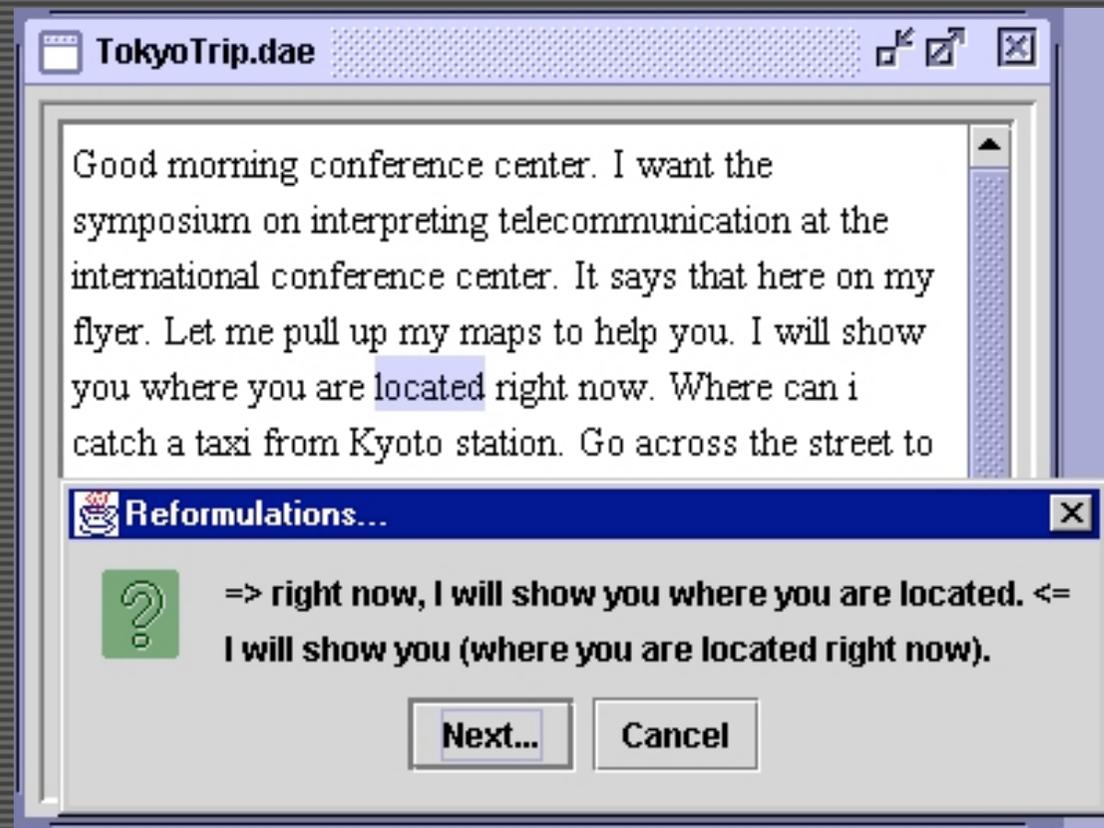
[Gressard 01-03]

[Schonek 03]

[Romier 04]

# Documents Auto-Explicatifs

- Visualisation d'une interprétation



# Verrous et pistes pour le futur

 Obtenir un module de DI couvrant

 DI construisant des descripteurs d'ambiguïté

*[Frey 96]  
[Irgadian 02]*

 **DI sans descripteur d'ambiguïté (DI par analogie)**

 Réduire le coût de la DI

 DI qui apprend, qui s'adapte au "style" de l'auteur

 Est-ce que l'auteur ne va pas aussi apprendre à rédiger ?

 DI incomplète & certification du sens

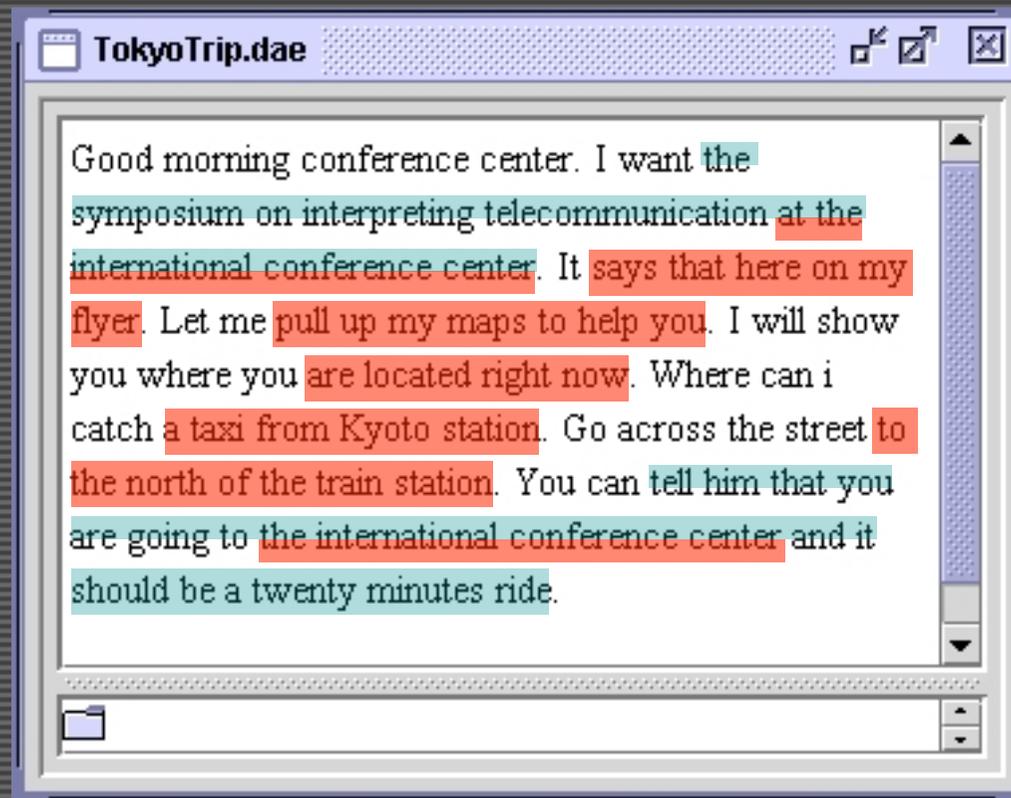
 Produire des DAE plus utiles

 **Localiser les ambiguïtés**



# Localiser les ambiguïtés

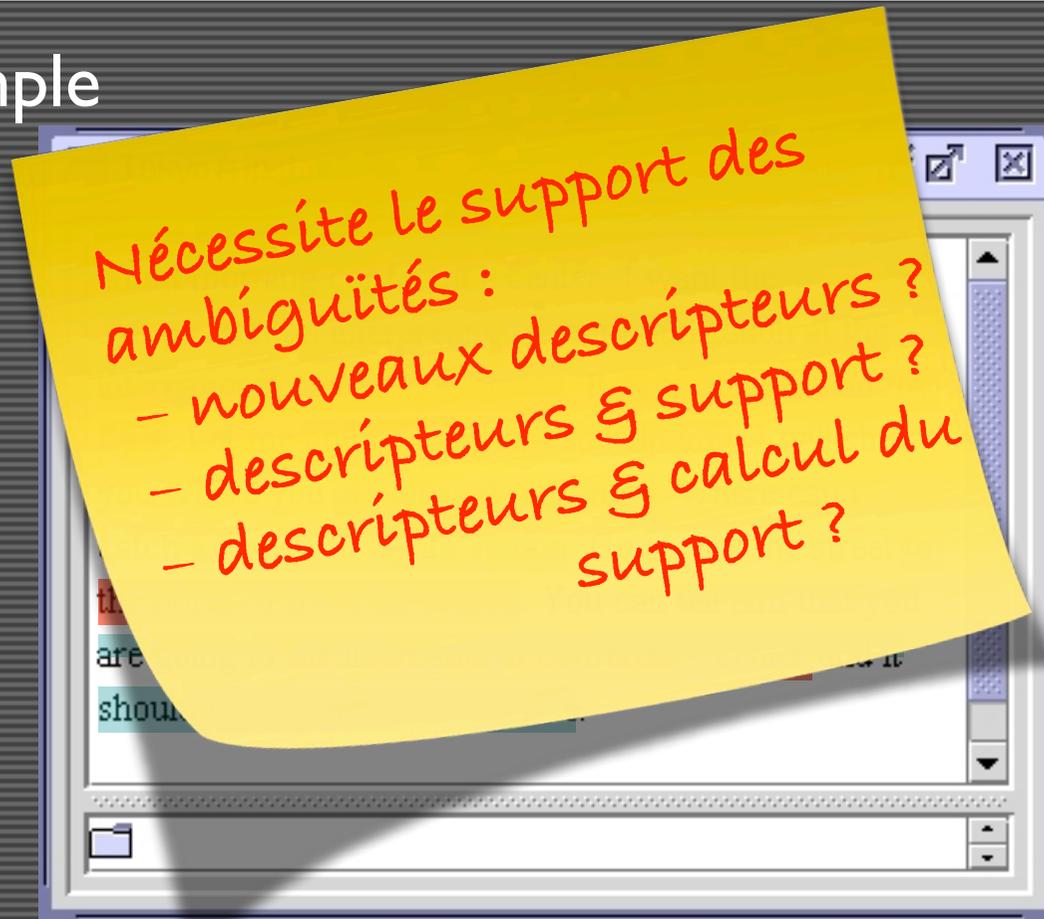
- Exemple





# Localiser les ambiguïtés

- Exemple

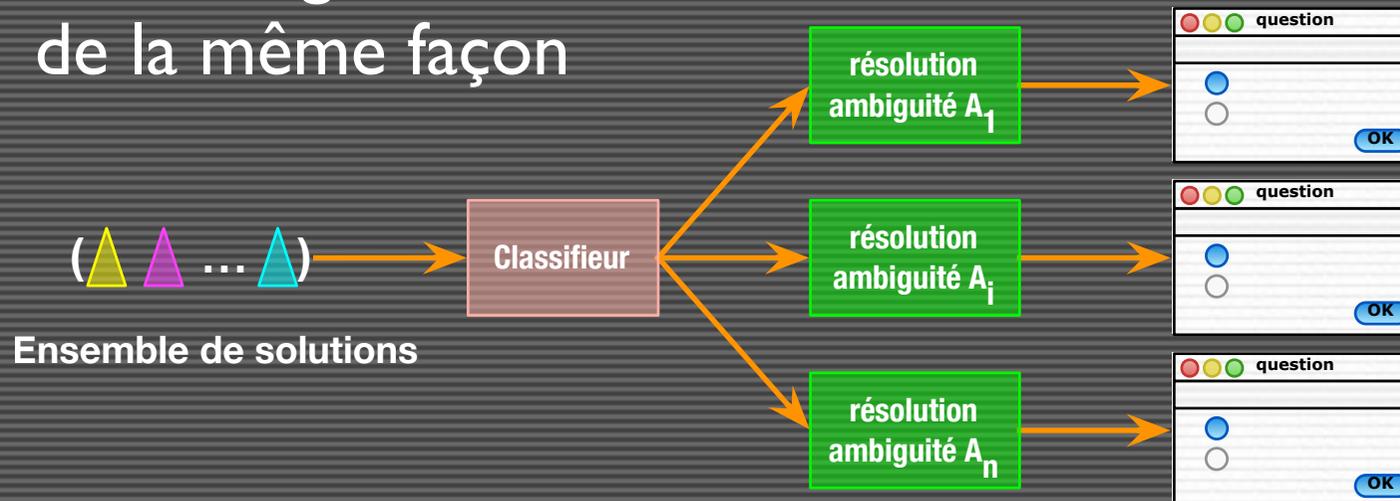




# DI par analogie

Coopération avec Y. Lepage, ATR

- 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



# Désambiguïsation analogique

## Idée

CONNAISSANCE

Il observe la photo et la classe.

Il observe la photo et il la classe.

Il observe la photo et il observe la classe.

français ambigu<sub>classe i</sub>

français désambiguïsé<sub>classe i</sub>

# Désambiguïsation analogique

## Idée

### CONNAISSANCE

Il observe la photo et la classe.

Il observe la photo et il la classe.

Il observe la photo et il observe la classe.

### QUESTION

Il atteint la grange et la ferme.

??

??

français ambigu classe i

français désambiguïsé classe i

# Désambiguïsation analogique

## Idée

### CONNAISSANCE

Il observe la photo et la classe.

Il observe la photo et il la classe.

Il observe la photo et il observe la classe.

### RÉPONSE

Il atteint la grange et la ferme.

Il atteint la grange et il la ferme.

Il atteint la grange et il atteint la ferme.

français ambigu classe i

français désambiguïsé classe i

# Désambiguïsation analogique

## Idée

CONNAISSANCE

Il observe la

erve la photo et il la classe.

to et il observe la classe.

Comment construire le  
français ambigu de  
chaque classe ?

RÉPONSE

Il atteint la grange et la ferme.

teint la grange et il la ferme.

Il atteint la grange et il atteint la ferme.

français ambigu classe i

français désambiguïsé classe i

# CI orale spontanée spécialisée

## Deuxième partie



1998-99



2000-02

- Introduction
- Lutte contre le bruit
  - Proposition
  - Réalisation
- Évaluation
  - Qualité des traductions
  - Enseignements
- Verrous & pistes pour le futur

# Introduction

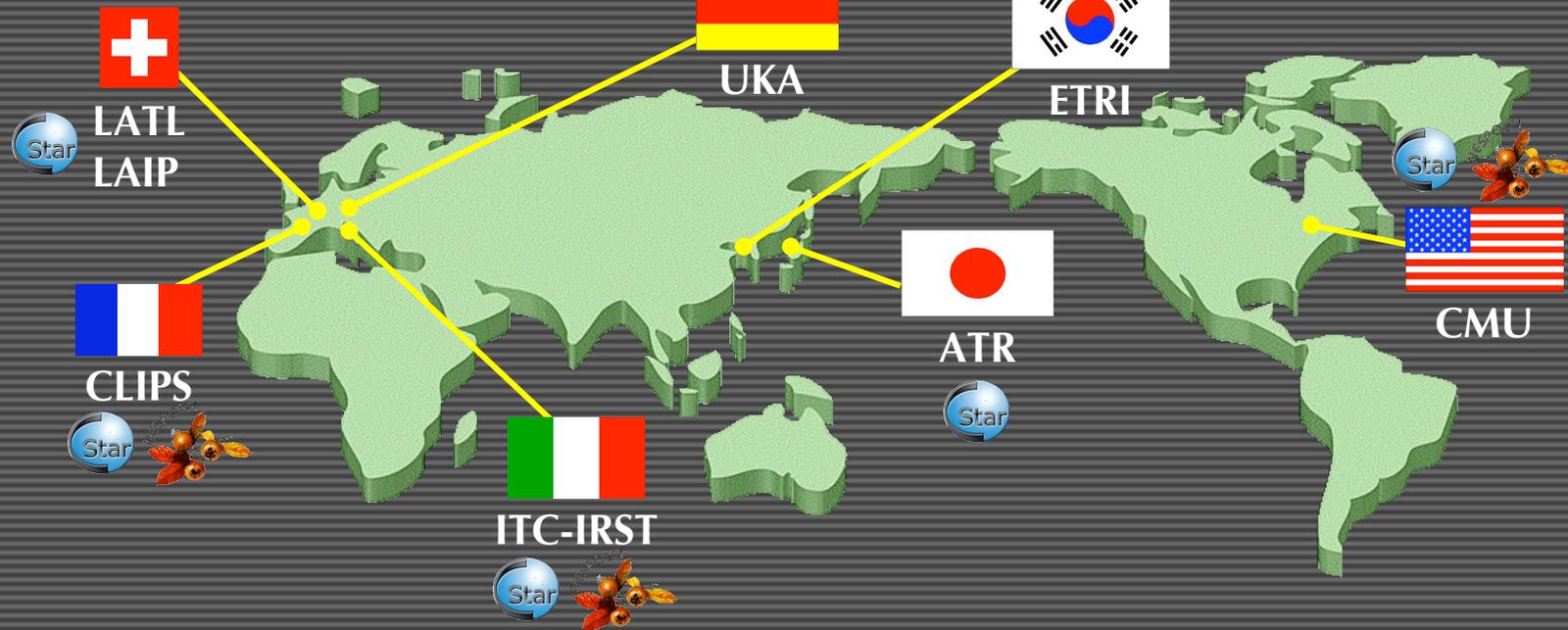


C-STAR II (1994-99)

CLIPS++ (1996-99)



NESPOLE! (2000-02)



# Introduction



**C-STAR II (1994-99)**  
*CLIPS++ (1996-99)*



**NESPOLE! (2000-02)**

## Équipes

*Aublet-Cuvelier (intégration)*

***Akbar (reco.)***

***Vaufreydaz (reco.)***

*Guilbaud (analyse)*

*Villard (analyse)*

*Blanchon (analyse ++)*

***Echegoyhen (génération)***

*Keller (synthèse)*

*Besacier (GEOD)*

*Vu Minh (ML sémantique)*

*Blanchon (a&g ++)*

*Ponsard (analyse, tagging)*

***Guilbaud (génération)***

*Fort (génération)*

*Helm (génération)*

***S. Mazenot (génération)***

*Lannelucq & Cau (génération)*

# Introduction

- C-STAR II ⇒ TA de dialogues oraux **réfléchis**
  - Domaine du tourisme *simulé* & **limité**
  - Démonstration devant les médias 22 juillet 1999
  
- NESPOLE! ⇒ TA de dialogues oraux **spontanés**
  - Domaine du tourisme *réel* & **réduit** puis **étendu**
    - Italien : **agent** de voyage professionnel
    - Anglais, Allemand, Français : **client**
  - 2 démonstrateurs : **2001** & **2002**

# Introduction

- C-STAR II ⇒ TA de dialogues oraux **réfléchis**

- Domaine du tourisme *simulé* & **limité**

hôtels, transports, informations touristiques

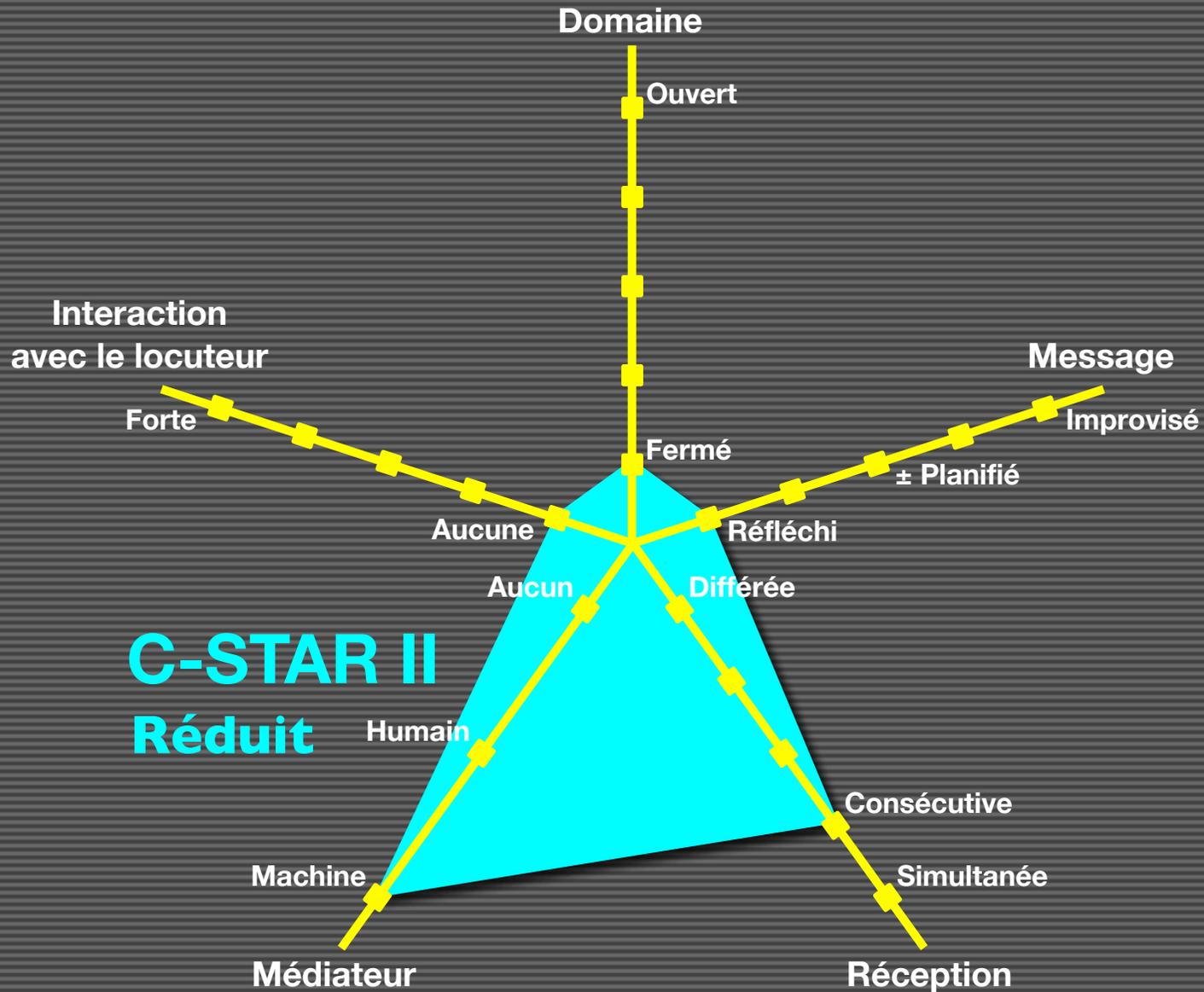
- NESPOLE! ⇒ TA de dialogues oraux **spontanés**

- Domaine du tourisme *réel* & **réduit** puis **étendu**

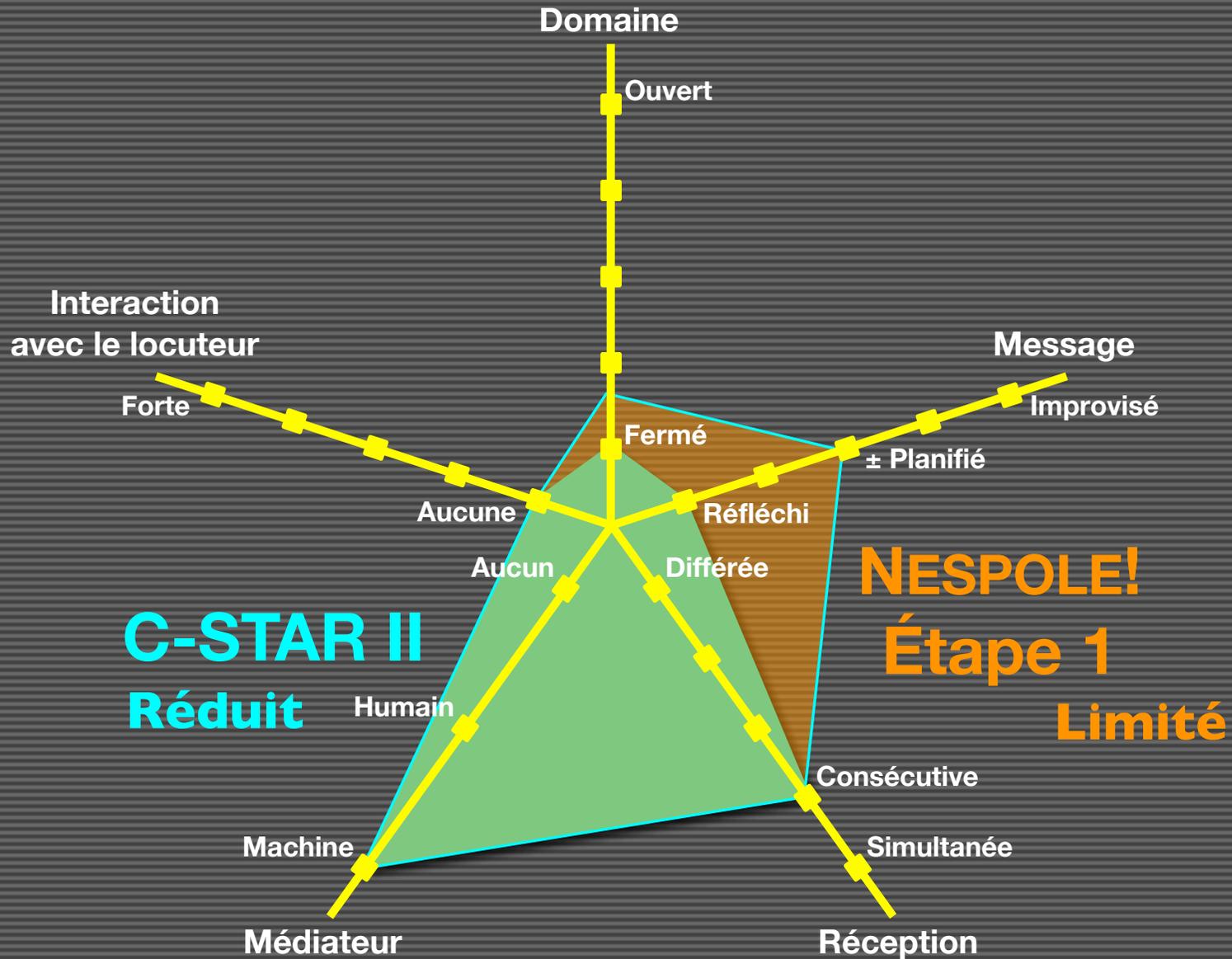
vacances été et hiver : logements, activités, itinéraires  
+ forfaits, châteaux et lacs, folklore

- 2 démonstrateurs : 2001 & 2002

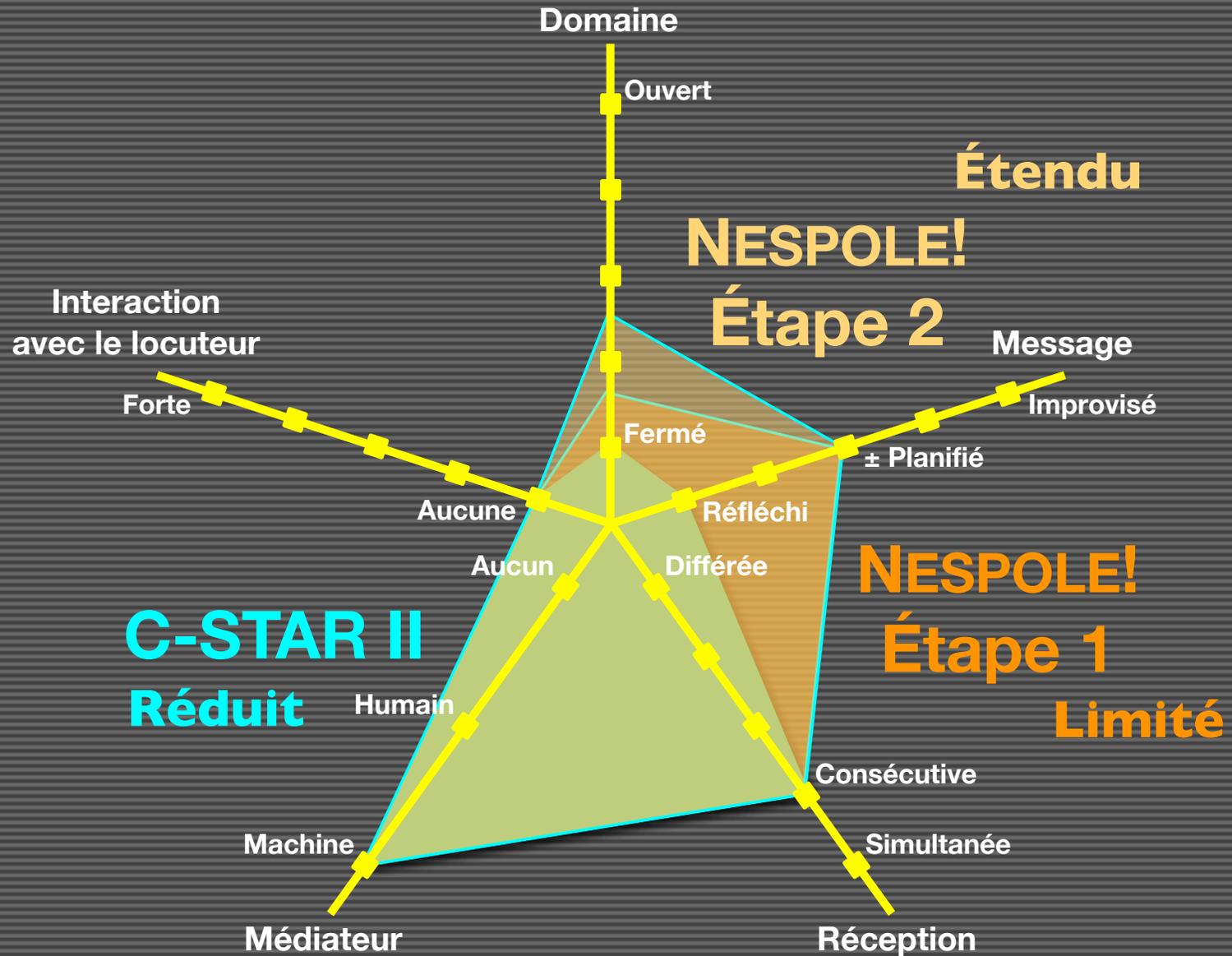
# communication interpersonnelle orale oponentielle spécialisée



# communication interpersonnelle orale opontanée spécialisée

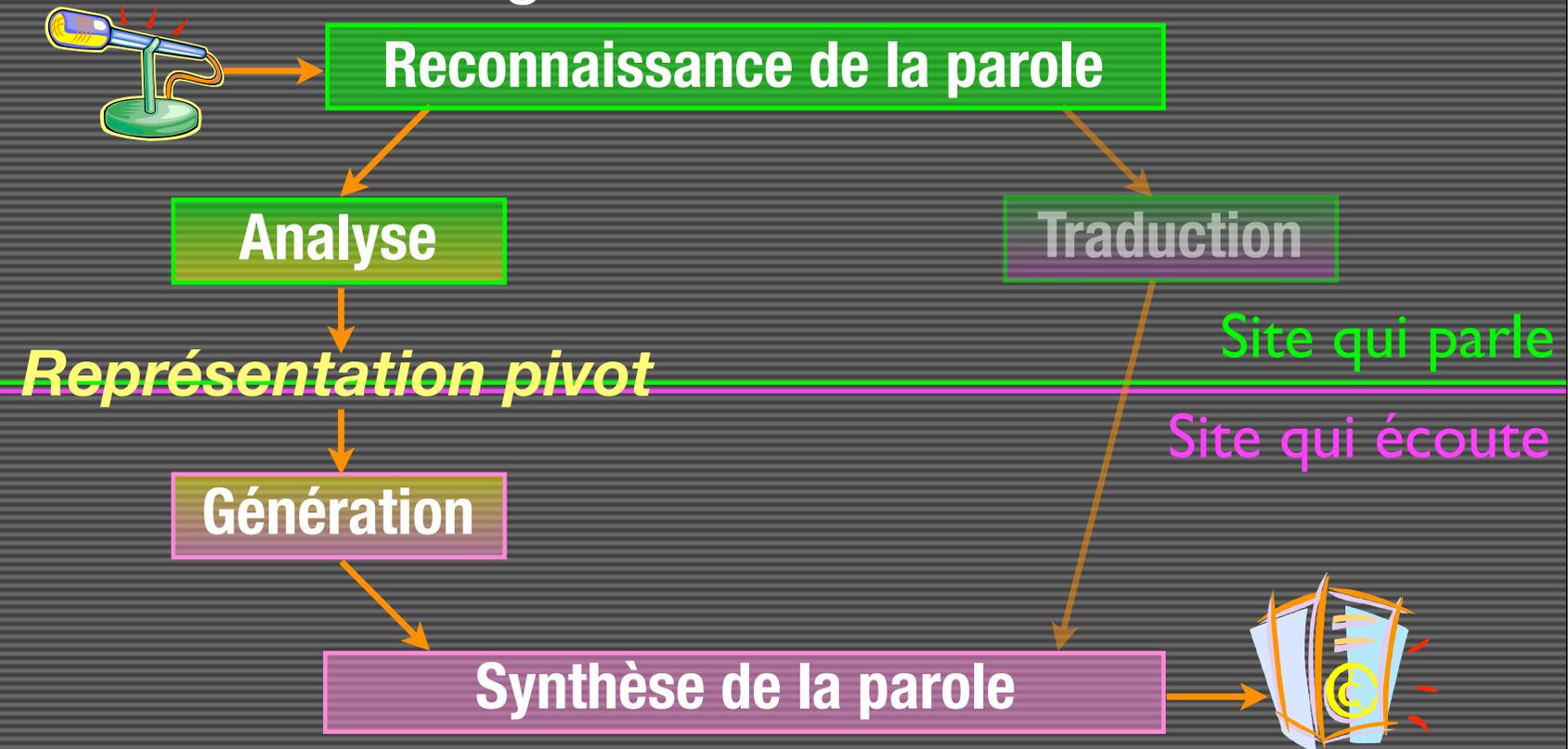


# communication interpersonnelle orale opontanée spécialisée



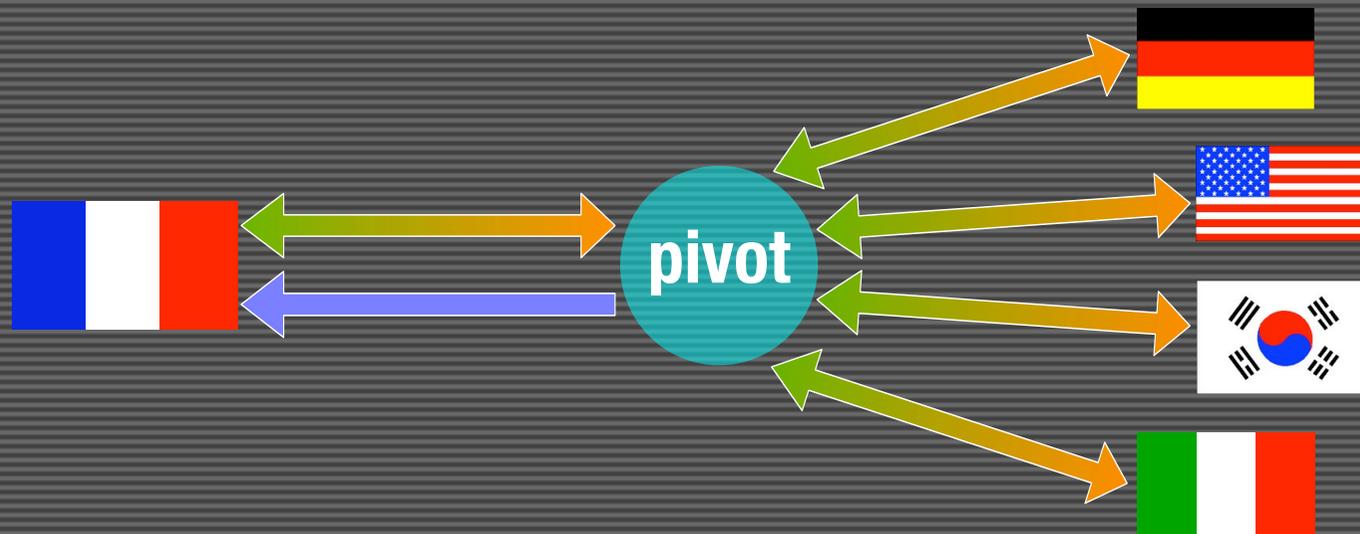
# Introduction

- Architecture linguistiques



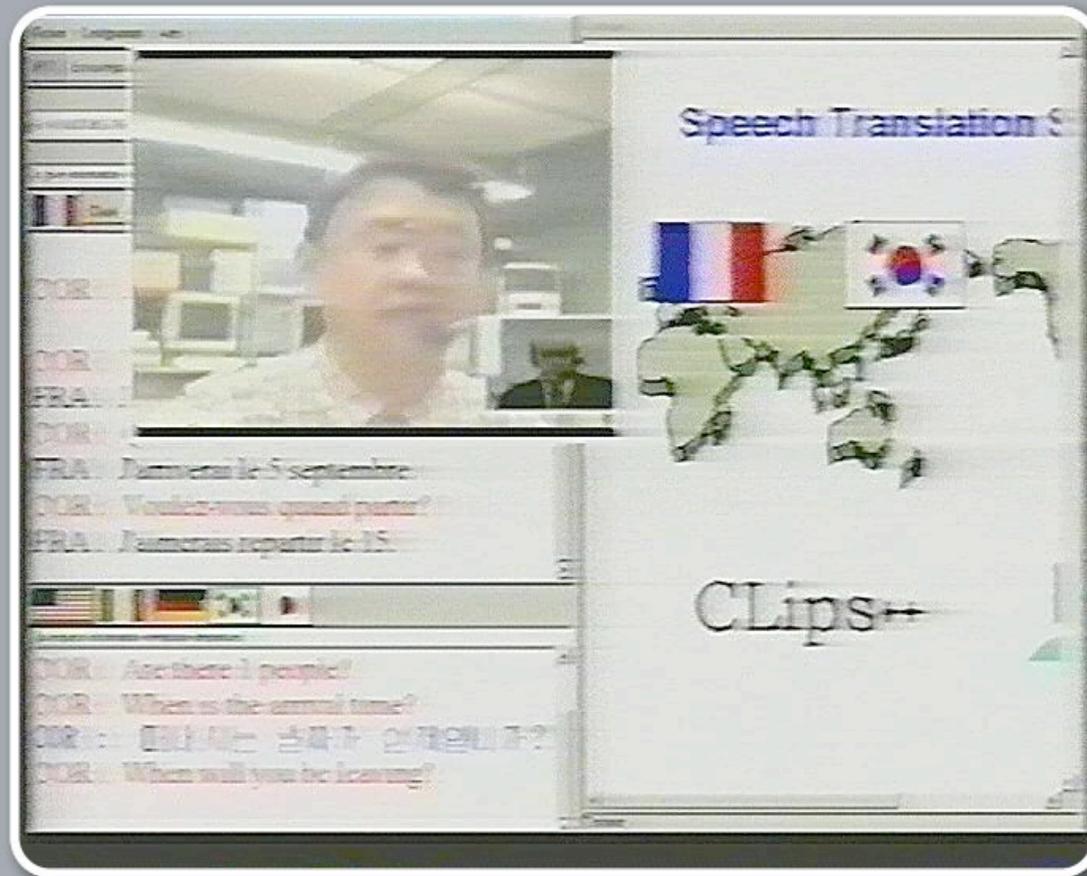
# Introduction

- Avantages de l'architecture par pivot
  - ⇒ n langues cibles pour le coût d'une seule
  - ⇒ n langues sources pour le coût d'une seule
  - ⇒ rétro-génération en langue source (contrôle)



# Introduction

- Démonstrations



# C-STAR 2

**Français-Coréen — septembre 1999**



# NESPOLE!

**Vidéo de présentation du projet**

# Lutte contre le bruit

 Qu'est-ce que le bruit ?

- Reconnaissance imparfaite de la parole
- Syntaxe non standard du français

# Lutte contre le bruit

⚠️ Qu'est-ce que le bruit ?

- Reconnaissance imparfaite de la parole
- Syntaxe non standard du français

⚠️ Pas de définition, pas de représentation positive

💡 Ne voir que ce qui est pertinent dans l'entrée à analyser (être aveugle au bruit)

- Savoir ce que l'on doit chercher
  - Quelles informations dans le résultat d'analyse ?
- Faire la recherche

# Lutte contre le bruit

- Résultat d'analyse : représentation IF
  - Un acte de Dialogue
    - Un **acte de parole**
    - Des **concepts**
  - Des **arguments**
    - Actualisation des variables du discours
- Exemple de représentation IF
  - **<client>** c est une chambre simple
  - {**c:give-information+room**(**room-spec=(identifiability=no, single-room)**)}

# Lutte contre de bruit



## Méthode fondée sur des patrons

[Blanchon 02-04]

### Automates de reconnaissance de séquences pertinentes

- Pour trouver quoi ?

- ... des valeurs d'arguments

*Les arguments permettent de trouver les concepts + l'acte de parole, soit l'**acte de dialogue***

- Exemple de patron

- " (\[0\ -9\ ]+) (**\$fifservdico::frenchcurrency**)(?:xls)? " \$inString IMatch **ICurQty** **ICurrency**

- "**price**=(quantity=" **\$ICurQty** ", currency=" [**fifservdico::Currency2If** **\$ICurrency**] ")"

# Analyse

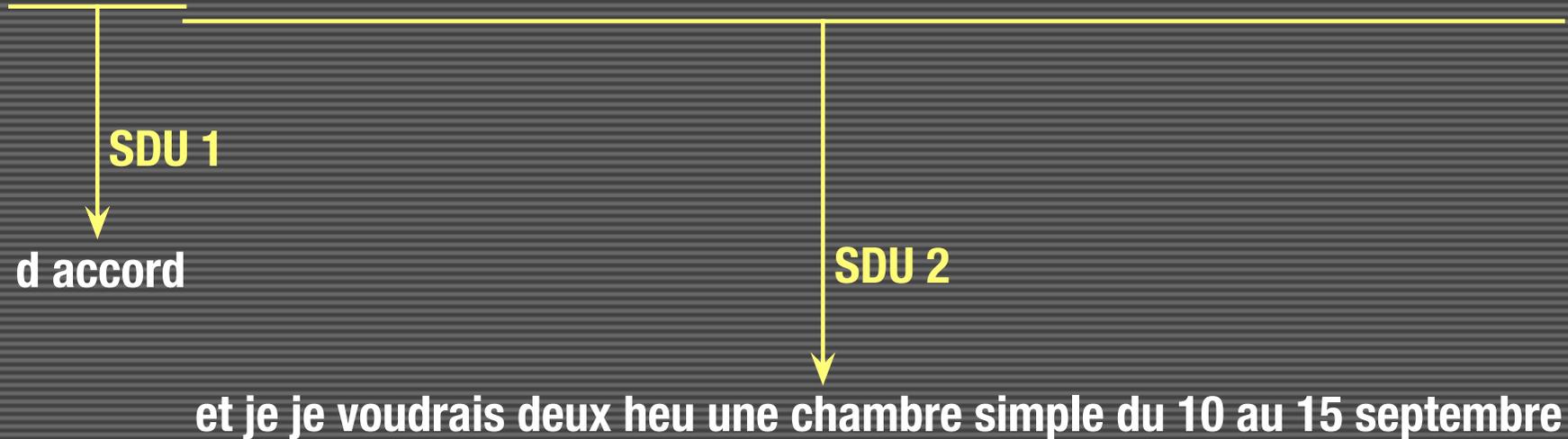
## Étape I : Découpage en SDU

**c : d accord et je je voudrais deux heu une chambre simple du 10 au 15 septembre**

# Analyse

## Étape I : Découpage en SDU

c : d accord **et je** je voudrais deux heu une chambre simple du 10 au 15 septembre



# Analyse

## Étape 2 : Détection du *focus concept*

c : et je je voudrais deux heu une chambre simple du 10 au 15 septembre

# Analyse

## Étape 2 : Détection du *focus concept*

c : et je je voudrais deux heu une **chambre simple** du 10 au 15 septembre



**room**

# Analyse

## Étape 3 : Construction de l'acte de dialogue

**c :** et je je voudrais deux heu une chambre simple du 10 au 15 septembre

# Analyse

## Étape 3 : Construction de l'acte de dialogue

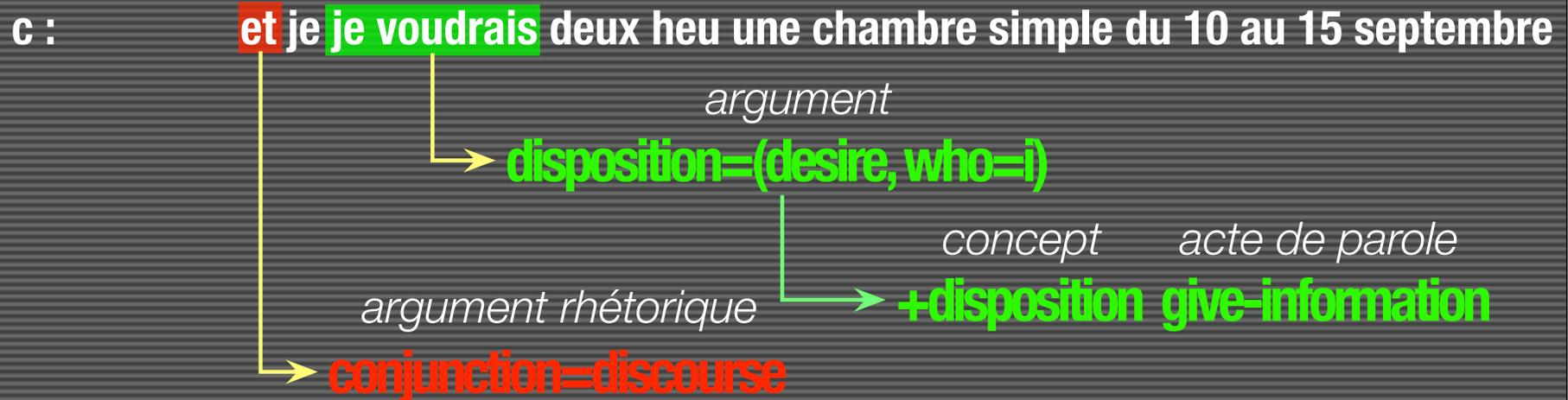
c : **et** je je voudrais deux heu une chambre simple du 10 au 15 septembre

*argument rhétorique*

→ **conjunction=discourse**

# Analyse

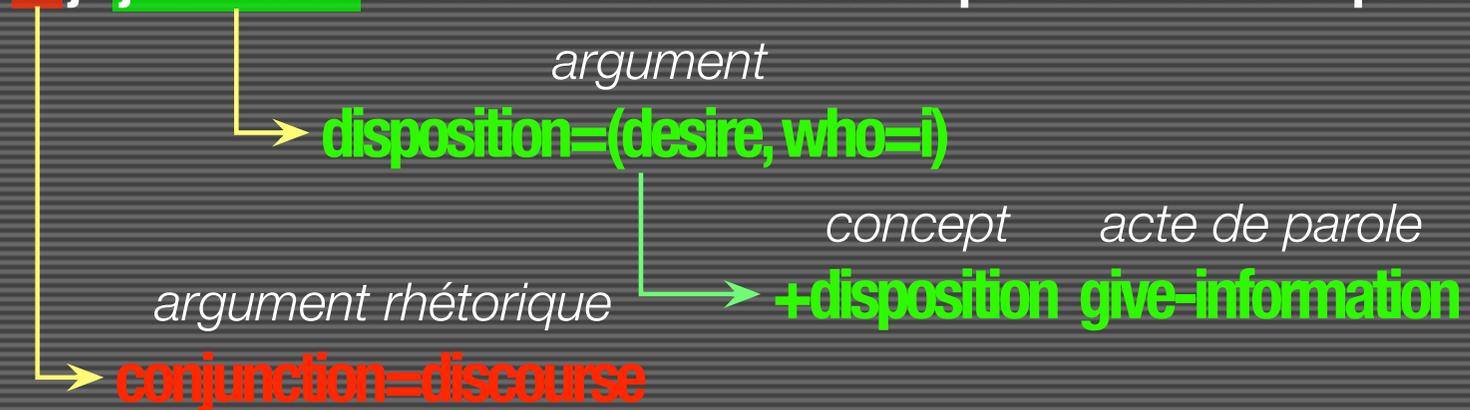
## Étape 3 : Construction de l'acte de dialogue



# Analyse

## Étape 3 : Construction de l'acte de dialogue

c : **et je je voudrais** deux heu une chambre simple du 10 au 15 septembre



{ **give-information+disposition**  
( **conjunction=discourse, disposition=(desire, who=i)**

}}

# Analyse

## Étape 4 : Construction des arguments

c : et je je voudrais deux heu une chambre simple du 10 au 15 septembre

# Analyse

## Étape 4 : Construction des arguments

c : et je je voudrais deux heu **une chambre simple** du 10 au 15 septembre

*argument*

→ **room-spec=(identifiability=no, single\_room)** *focus concept*

→ **+room**

# Analyse

## Étape 4 : Construction des arguments

c :

et je je voudrais deux heu **une chambre simple** du 10 au 15 septembre

*argument*

→ **room-spec=(identifiability=no, single\_room)**

*focus concept*

→ **+room**

*argument*

→ **time=(start-time=(md=10), end-time=(md=15, month=9))**

# Analyse

## Étape 4 : Construction des arguments

c :

et je je voudrais deux heu **une chambre simple** du 10 au 15 septembre

*argument*

→ **room-spec=(identifiability=no, single\_room)**

*focus concept*

→ **+room**

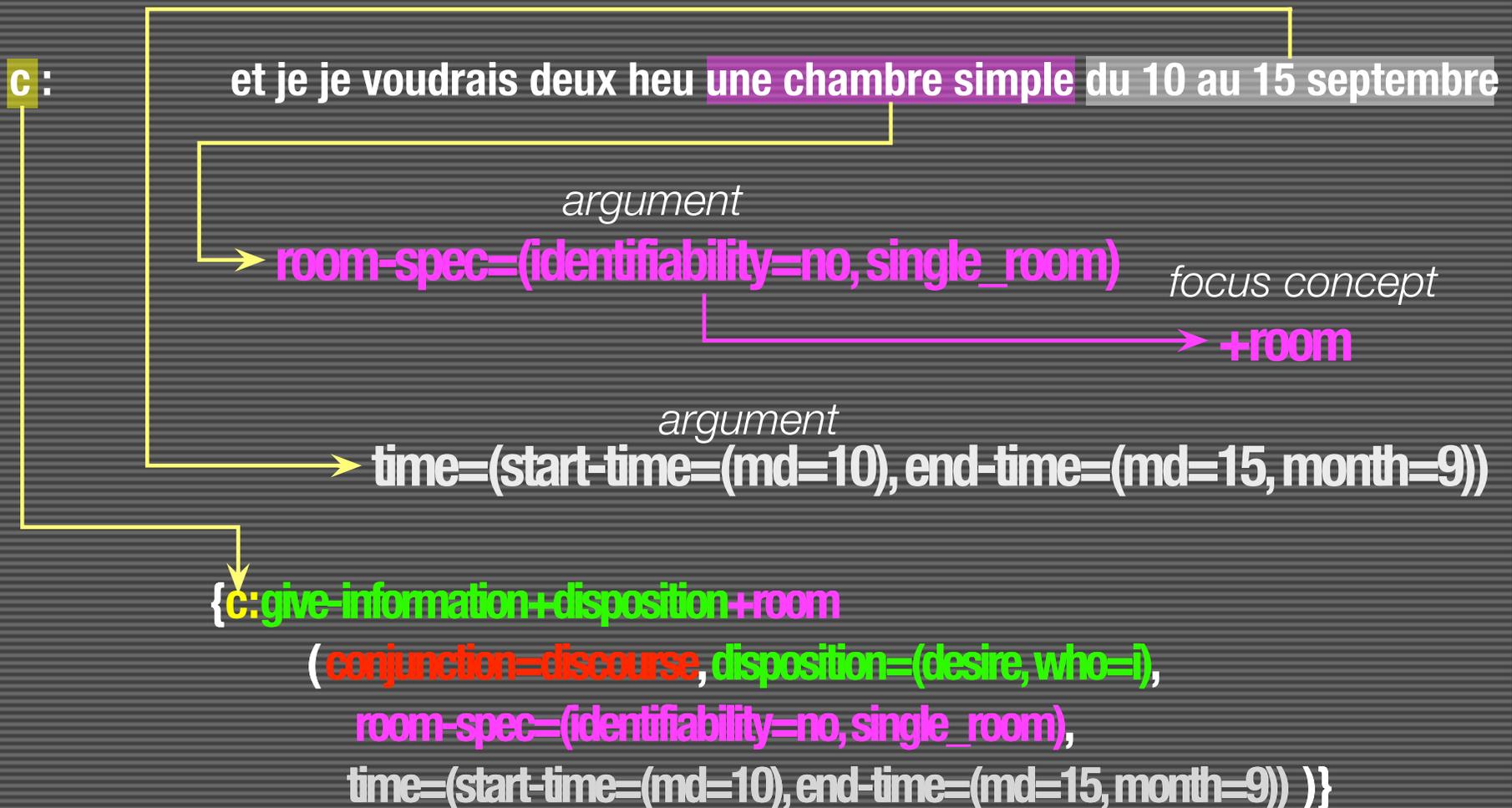
*argument*

→ **time=(start-time=(md=10), end-time=(md=15, month=9))**

```
{ give-information+disposition+room  
  (conjunction=discourse, disposition=(desire, who=i),  
   room-spec=(identifiability=no, single_room),  
   time=(start-time=(md=10), end-time=(md=15, month=9)) ) }
```

# Analyse

## Étape 4 : Construction des arguments



# Évaluation

## Que veut-on évaluer ?

- Le module de reconnaissance de la parole
  - En pourcentage de mots correctement reconnus (WAR)
  - En terme de paraphrase de la transcription du signal

# Évaluation

## ⚠ Que veut-on évaluer ?

- Le module de reconnaissance de la parole
  - En pourcentage de mots correctement reconnus (WAR)
  - En terme de paraphrase de la transcription du signal
- La chaîne de traduction monolingue, et
- La chaîne de traduction bilingue
  - Sur les références (transcriptions manuelles du signal)
    - Performances atteignables avec un WAR à 100%
  - Sur les hypothèses de reconnaissance
    - Performances réelles du système

# Évaluation

## ⚠️ Que veut-on évaluer ?

- Le mode de fonctionnement du système
  - En position de veille
  - En termes de performance (WAR)
  - En termes de performance du signal
- La chaîne de commandement
- La chaîne de transmission
  - Sur les performances réelles (performances réelles du signal)
  - Performances réelles du système avec un WAR à 100%
- Sur les hypothèses de reconnaissance
  - Performances réelles du système

Évaluation SUBJECTIVE :

Des juges humains  
donnent des scores à  
chaque SDU.

# Évaluation — Second démonstrateur

## ● Résultats (Traductions acceptables)

WAR à 100%

Reco	56%	51%	58%	75%
Hypo as Paraphrase	67%	62%	60%	76%
Monolingue	An→An <b>c</b>	All→All <b>c</b>	Fr→Fr <b>c</b>	It→It <b>a</b>
sur Refs / Hypos	68 / 50	61 / 51	77 / 58	51 / 42
Bilingue	An→It <b>c</b>	All→It <b>c</b>	Fr→It <b>c</b>	
sur Refs / Hypos	70 / 50	x / x	77 / 58	
	It→An <b>a</b>	It→All <b>a</b>	It→Fr <b>a</b>	
	33 / 30	45 / 38	37 / 33	

# Évaluation — Second démonstrateur

## ● Résultats (Traductions acceptables)

WAR à 100%

Reco	56%	51%	58%	75%
Hypo as Paraphrase	67%	62%	60%	76%
Monolingue	An→An <b>c</b>	All→All <b>c</b>	Fr→Fr <b>c</b>	It→It <b>a</b>
sur Refs / Hypos	68 / 50	61 / 51	77 / 58	51 / 42
Bilingue	An→It <b>c</b>	All→It <b>c</b>	Fr→It <b>c</b>	
sur Refs / Hypos	70 / 50	x / x	77 / 58	
	It→An <b>a</b>	It→All <b>a</b>	It→Fr <b>a</b>	
	33 / 30	45 / 38	37 / 33	

# Évaluation — Second démonstrateur

## ● Résultats (Traductions acceptables)

WAR à 100%

Reco	56%	51%	58%	75%
Hypo as Paraphrase	67%	62%	60%	76%
Monolingue	An→An <b>c</b>	All→All <b>c</b>	Fr→Fr <b>c</b>	It→It <b>a</b>
sur Refs / Hypos	68 / 50	61 / 51	77 / 58	51 / 42
Bilingue	An→It <b>c</b>	All→It <b>c</b>	Fr→It <b>c</b>	
sur Refs / Hypos	70 / 50	x / x	77 / 58	
	It→An <b>a</b>	It→All <b>a</b>	It→Fr <b>a</b>	
	33 / 30	45 / 38	37 / 33	

# Évaluation — Second démonstrateur

## ● Résultats (Traductions acceptables)

WAR à 100%

Reco	56%	51%	58%	75%
Hypo as Paraphrase	67%	62%	60%	76%
Monolingue	An→An <b>c</b>	All→All <b>c</b>	Fr→Fr <b>c</b>	It→It <b>a</b>
sur Refs / Hypos	68 / 50	61 / 51	77 / 58	51 / 42
Bilingue	An→It <b>c</b>	All→It <b>c</b>	Fr→It <b>c</b>	
sur Refs / Hypos	70 / 50	x / x	77 / 58	
	It→An <b>a</b>	It→All <b>a</b>	It→Fr <b>a</b>	
	33 / 30	45 / 38	37 / 33	

# Évaluation — Second démonstrateur

## ● Résultats (Traductions acceptables)

WAR à 100%

Reco	56%	51%	58%	75%
Hypo as Paraphrase	67%	62%	60%	76%
Monolingue	An→An <b>c</b>	All→All <b>c</b>	Fr→Fr <b>c</b>	It→It <b>a</b>
sur Refs / Hypos	68 / 50	61 / 51	77 / 58	51 / 42
Bilingue	An→It <b>c</b>	All→It <b>c</b>	Fr→It <b>c</b>	
sur Refs / Hypos	70 / 50	x / x	77 / 58	
	It→An <b>a</b>	It→All <b>a</b>	It→Fr <b>a</b>	
	33 / 30	45 / 38	37 / 33	

# Évaluation — Second démonstrateur

## ● Résultats (Traductions)

WAR à 100%

Reco			58%	75%
Hypo as Paraph			0%	76%
Monolingue			Fr <b>c</b>	It→It <b>a</b>
sur Refs / Hyp			58	51 / 42
Bilingue			<b>c</b>	
sur Refs / Hypos			77 / 58	
	It→An <b>a</b>	It→All <b>a</b>	It→Fr <b>a</b>	
	33 / 30	45 / 38	37 / 33	

Les Italiens ont mal travaillé !

# Changement d'échelle

⚠ Pourquoi ces mauvais scores à partir de l'italien ?

- Classification des données en 3 catégories

1. Domaine du premier démonstrateur

2. Domaine du second démonstrateur

3. Hors du domaine des 2 démonstrateurs

- Pour le français, l'allemand et l'anglais (clients)

- moins de 5% dans les catégories 2. & 3. !!!!

# Changement d'échelle

## ⚠ Pourquoi ces mauvais scores à partir de l'italien ?

- Classification des données en 3 catégories
  1. Domaine du premier démonstrateur
  2. Domaine du second démonstrateur
  3. Hors du domaine des 2 démonstrateurs
- Pour le français, l'allemand et l'anglais (clients)
  - moins de 5% dans les catégories 2. & 3. !!!!
- Pour l'italien (agents)
  - 13% dans la catégorie 2.
  - 25% dans la catégorie 3.

[Blanchon 04]

# Changement d'échelle

⚠ Pourquoi ces mauvais résultats sur le tir de l'italien ?

- Classification des tours de parole en 3 catégories
  - 1. Domaines
  - 2. Domaines
  - 3. Hors-domaine
- Pour le tir de l'italien (agents)
  - moins de 25% de tours de parole analysables
- Pour l'italien (agents)
  - 13% dans la catégorie 2.
  - 25% dans la catégorie 3.

Le domaine devait croître  
... il a trop crû !!!!

25% des tours de parole  
de l'agent ne sont pas  
analysables !!!!

# Changement d'échelle

## ● Résultats bruts

Reco	56%	51%	58%	75%
Hypo as Paraphrase	<b>67%</b>	<b>62%</b>	<b>60%</b>	<b>76%</b>
Monolingue	An→An <b>c</b>	All→All <b>c</b>	Fr→Fr <b>c</b>	It→It <b>a</b>
sur Refs / Hypos	68 / 50	61 / 51	77 / 58	51 / 42
Bilingue	An→It <b>c</b>	All→It <b>c</b>	Fr→It <b>c</b>	
sur Refs / Hypos	70 / 50	x / x	77 / 58	
	It→An <b>a</b>	It→All <b>a</b>	It→Fr <b>a</b>	
	33 / 30	45 / 38	37 / 33	

# Changement d'échelle

## ● Résultats corrigés (+9)

Reco	56%	51%	58%	75%
Hypo as Paraphrase	<b>67%</b>	<b>62%</b>	<b>60%</b>	<b>76%</b>
Monolingue	An→An <b>c</b>	All→All <b>c</b>	Fr→Fr <b>c</b>	It→It <b>a</b>
sur Refs / Hypos	68 / 50	61 / 51	77 / 58	<b>59 / 51</b>
Bilingue	An→It <b>c</b>	All→It <b>c</b>	Fr→It <b>c</b>	
sur Refs / Hypos	70 / 50	x / x	77 / 58	
	It→An <b>a</b>	It→All <b>a</b>	It→Fr <b>a</b>	
	33 / 30	45 / 38	<b>46 /</b>	

# Changement d'échelle

## ● Résultats corrigés

Reco			58%	75%
Hypo as Par			60%	76%
Monolingue			→Fr <b>c</b>	It→It <b>a</b>
sur Refs /			/ 58	59 / 51
Bilingue			- c	
sur Refs / Hyp			77 / 58	
	It→An <b>a</b>	It→All <b>a</b>	It→Fr <b>a</b>	
	33 / 30	45 / 38	46 /	

Quand on ouvre le domaine, on ouvre la boîte de Pandore !!



Rosso Fiorentino  
1495-1540

# Verrous et pistes pour le futur

## Intégration des composants

 En entrée *[Vu Minh 04, modèle de langage sémantique]*

 En sortie

 Gestion du contexte et du dialogue

## Augmentation de la couverture

## Architecture linguistique

 Rester multilingue : approche pivot

- Attention au choix du pivot

- L'approche statistique reste possible [MASTOR - IBM]

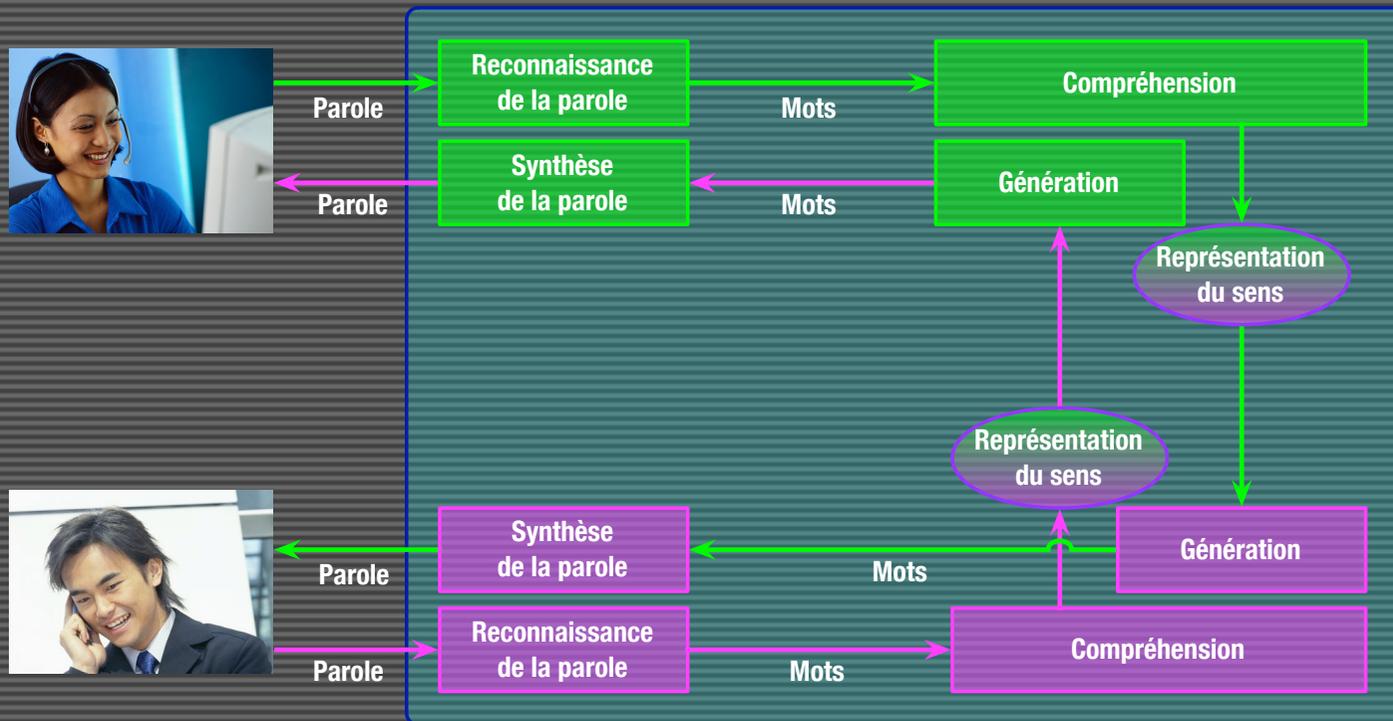


# Gestion du contexte

[Boitet, Blanchon & Guilbaud 00]

- Trois contextes : **Global**, **Dialogique**, **Linguistique**
  - Type de dialogue
  - Caractéristique des participants (田中さん。 ⇨ Mr/Ms Tanaka?)
  - Rôle des participants, ...
  - Représentation du passé
  - État du dialogue
  - Prédications sur le futur
  - Centres possibles pour les ellipses
  - Choix lexicaux antérieurs (consistance)
  - Domaine pour la désamb. lexicale (express ⇨ coffee, train)

# Communication Interpersonnelle Orale Spontanée Spécialisée **actuelle**



# Communication Interpersonnelle Orale Spontanée Spécialisée **actuelle**

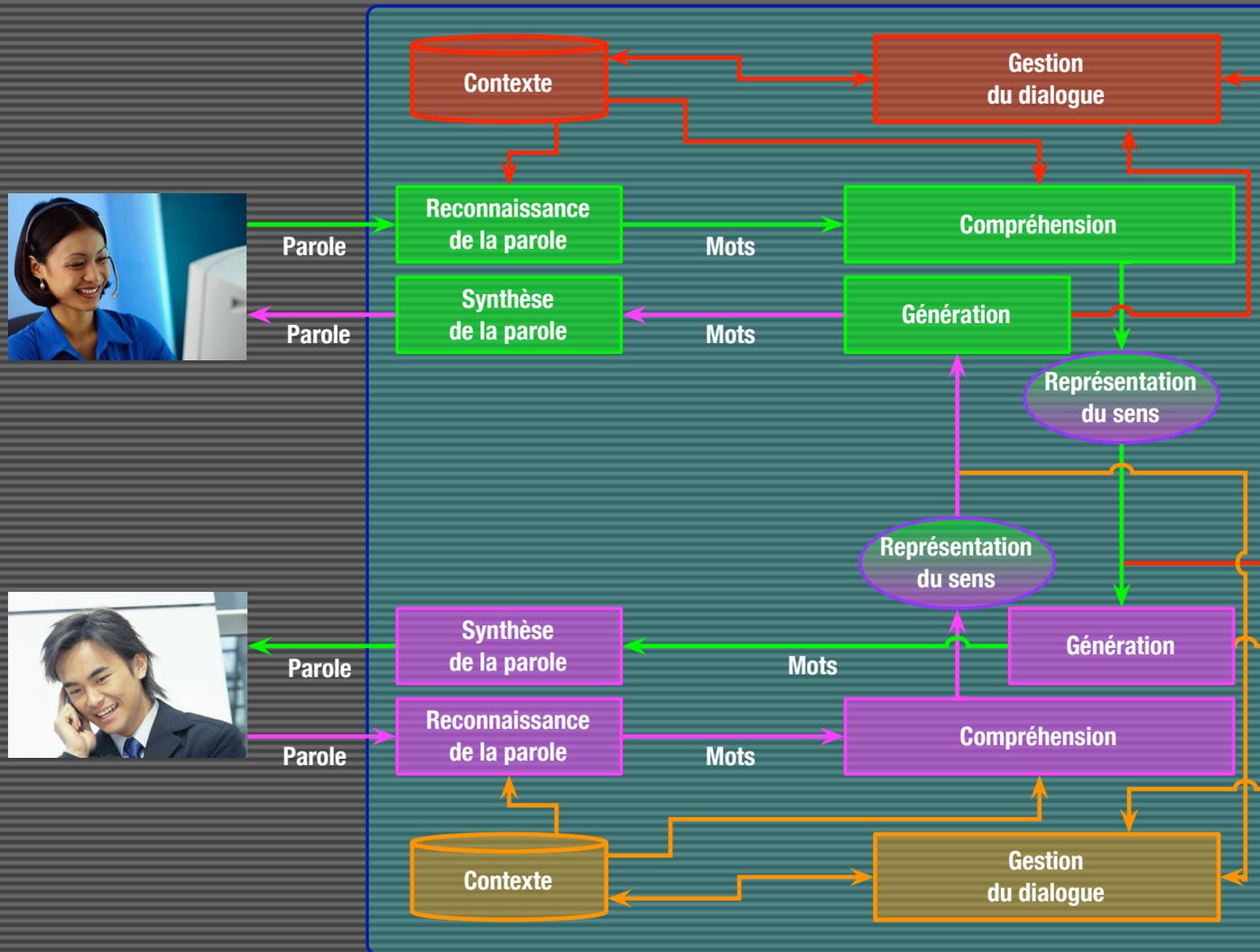


Pour l'instant, on traduit  
des énoncés isolés !  
OR ...  
On veut traduire du  
dialogue, donc ...  
des énoncés en contexte !



# Communication Interpersonnelle

## Orale Spontanée Spécialisée **proposée** en opposition

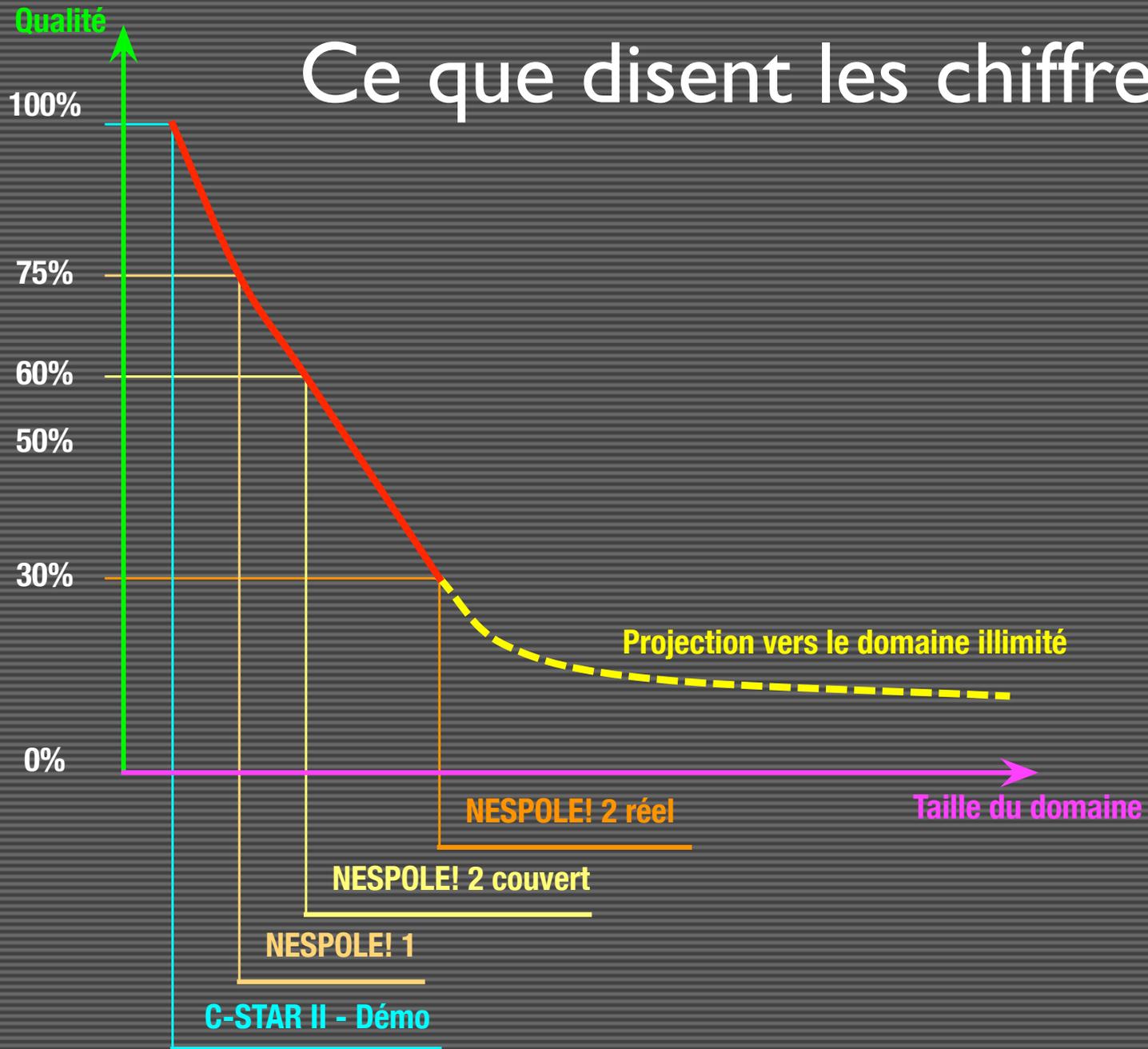




# Augmentation de la couverture

 **LA** nouvelle frontière : le domaine illimité ?

# Ce que disent les chiffres





# Augmentation de la couverture

⚠️ **LA** nouvelle frontière : le domaine illimité ?

- Bel objectif scientifique

- ⚠️ Être raisonnable si on veut viser des utilisateurs !



# Augmentation de la couverture

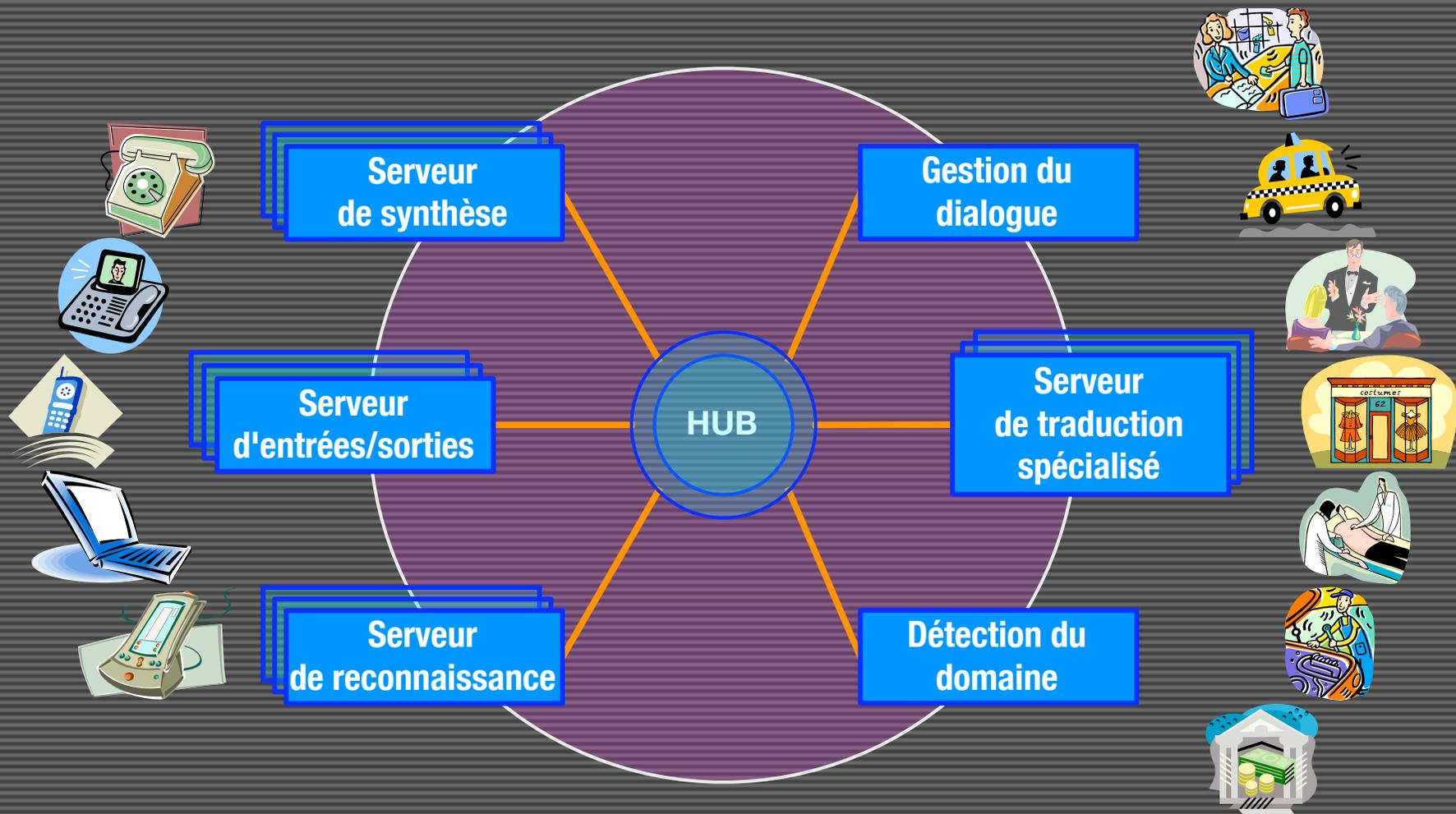
 **LA** nouvelle frontière : le domaine illimité ?

- Bel objectif scientifique

-  Être raisonnable si on veut viser des utilisateurs !

 **Intégration de systèmes spécialisés à travers une interface unique**

# Intégration de systèmes spécialisés





# Augmentation de la couverture

 **LA** nouvelle frontière : le domaine illimité ?

- Bel objectif scientifique

-  Être raisonnable si on veut viser des utilisateurs !

 **Intégration de systèmes spécialisés à travers une interface unique**

- Tester rapidement les systèmes !

- Utiliser les résultats de la recherche !

- **Progresser sur la couverture en parallèle !**

# Retour sur l'évaluation

Troisième  
partie



2000-06

Présidence 2004-

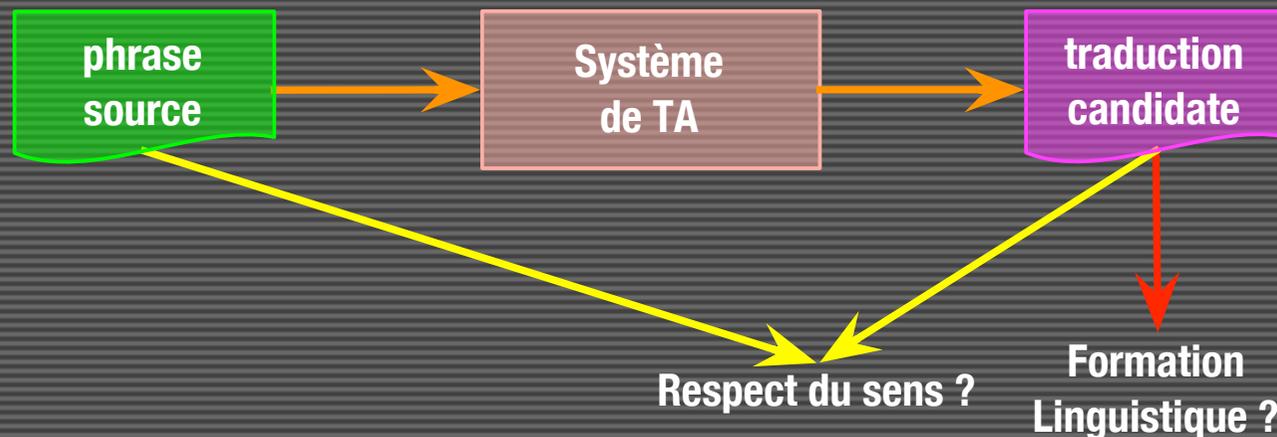
- Évaluation subjective
- Problèmes & Idées
- Évaluation objective
- Expérience pratique
  - Campagne IWSLT 2004
  - Résultats & commentaires
- Verrous et pistes pour le futur

# Évaluation subjective

- Ancienne pratique : juger ...
  - ... la bonne formation linguistique de la traduction
  - ... le respect du sens
    - **par rapport à la phrase source**

# Évaluation subjective

- Ancienne pratique : juger ...
  - ... la bonne formation linguistique de la traduction
  - ... le respect du sens
- **par rapport à la phrase source**

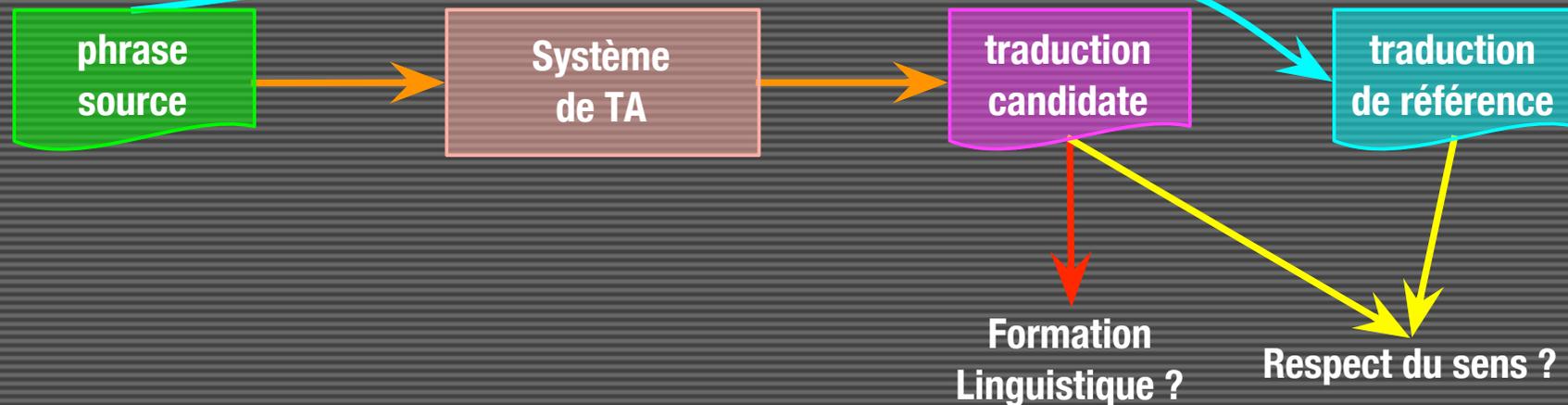


# Évaluation subjective

- Nouvelle pratique (NIST) : juger ...
  - ... la bonne formation linguistique de la traduction
  - ... le respect du sens
    - **par rapport à une traduction de référence**

# Évaluation subjective

- Nouvelle pratique (NIST) : juger ...
  - ... la bonne formation linguistique de la traduction
  - ... le respect du sens
- **par rapport à une traduction de référence**  
Traduction humaine



# Évaluation subjective

- Nouvelle pratique (NIST) : juger ...
- ... la linguistique de la traduction

*La langue source a disparu !!!!*

**traduction de référence**  
maine

traduction candidate

traduction de référence

Formation Linguistique ?

Respect du sens ?

# Problèmes cités par la communauté

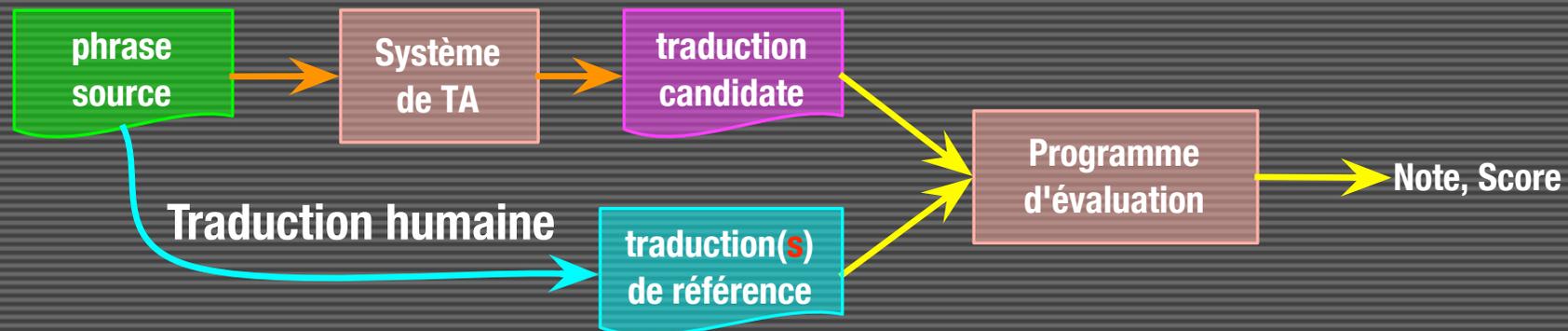
- L'évaluation subjective demande des ressources !
  - des juges bilingues
  - des juges monolingues + des traductions de références
- L'évaluation subjective prend beaucoup de temps !
- Les juges ne sont pas toujours ...
  - d'accord entre eux !
  - d'accord avec eux-mêmes dans le temps !

# Idée proposée d'abord par IBM

- Automatiser l'évaluation ...
  - **Un programme déterministe qui calcule un score**
    - Plus de problème de désaccord entre juges
    - Plus de problème de variation dans le temps
- ... en utilisant les données disponibles, construites pour l'évaluation subjective

# Idée proposée d'abord par IBM

- Automatiser l'évaluation ...
  - **Un programme déterministe qui calcule un score**
    - Plus de problème de désaccord entre juges
    - Plus de problème de variation dans le temps
- ... en utilisant les données disponibles, construites pour l'évaluation subjective



# Idée proposée d'abord par IBM

- Automatiser l'évaluation

- Un **programme qui calcule un score**

- F

- P

- ... en **temps** **limités**, construites

pour



# Évaluation objective

- Mesures fondées sur ...
  - la co-occurrence de n-grammes
    - **BLEU** [Papineni & al. 02]
    - **NIST** [Doddington 02]
  - la co-occurrence de sous-séquences communes connexes ou non
    - **ROUGE** (Famille de mesures) [Lin 04]
- Mesure d'évaluation des méthodes d'évaluation
  - **ORANGE** [Lin 04]

# Évaluation objective

- Mesures fondées sur

- la co-oc

- **BLEU**

- **NIST**

- la co-oc

- connexe

- **ROUGE**

- Mesure d'évaluation des méthodes d'évaluation

- **ORANGE** [Lin 04]

*On aura bientôt l'arc en ciel !!*



communes

# Expérience pratique

- IWSLT 2004

- Évaluation compétitive sur les mêmes données

- 2 Couples de langues

- chinois  $\rightleftarrows$  anglais

- japonais  $\rightleftarrows$  anglais

- Évaluation subjective

- NIST (formation linguistique, respect du sens)

- Évaluation objective

- 5 mesures (BLEU, NIST, GTM, WER, PER)

*Bey*

*Blanchon*

*Boitet*

*Brunet-Manquat*

*Cao*

*Chalabreysse*

*Hamon*

*Hung*

*Tomokiyo*

# Expérience pratique

- Participation du CLIPS
  - Utiliser les **systemes Systran**
    - souvent utilisés comme systèmes de référence
  - Voir des traductions candidates et les résultats d'évaluation associés

# Expérience pratique

- Participation du CLIPS
  - Utiliser les **systemes Systran**
    - souvent utilisés comme systèmes de référence
  - Voir des traductions candidates et les résultats d'évaluation associés
- Résultats
  - Systran chinois  $\Rightarrow$  anglais : **7<sup>ième</sup> sur 8**
  - Systran japonais  $\Rightarrow$  anglais : **4<sup>ième</sup> sur 4**

# Expérience pratique

## Nouvelles données

- Traductions Systran japonais → anglais révisées manuellement (de nouveaux candidats)

# Expérience pratique

## Nouvelles données

- Traductions Systran japonais → anglais révisées manuellement (de nouveaux candidats)

## ● Nouveaux résultats

 Les traductions révisées se classent 3<sup>ème</sup> sur 4 !!!

# Expérience pratique

## Nouvelles données

- Traductions candidates Systran japonais → anglais révisées manuellement

## ● Nouveaux résultats

⚠ Les traductions révisées se classent 3<sup>ième</sup> sur 4 !!!

⚠ Les traductions humaines ne sont pas bonnes ?

- **Elles sont très bonnes !!!**
- **..., mais elles n'imitent pas les références !!!**

# Expérience pratique

⚠ On n'évalue pas la qualité des traductions

- ..., **mais la "similarité" entre**
  - une **traduction candidate**, et
  - une ou plusieurs **traductions de références**

# Expérience pratique

⚠ On n'évalue pas la qualité des traductions

● ..., mais la "similarité" entre

● une **traduction candidate**, et

● une ou plusieurs **traductions de références**

⚠ Donc ...

● Si le score est bon, les traductions candidates sont bonnes pour peu que les références le soient

● **Si le score est mauvais, on ne peut rien dire !!!**

● Sauf si le système est conçu pour imiter les références

● TA par l'exemple, TA statistique

# Expérience pratique

⚠ On n'évalue pas la qualité des productions

● ..., m

● une t

● une o

ences

⚠ Donc ...

● Si le score est mauvais, les références sont  
bonnes pour que les références soient

● **Si le score est mauvais, on ne peut rien dire !!!**

● Sauf si le système est conçu pour imiter les références

● TA par l'exemple, TA statistique

*Attention à la  
communication sur les  
résultats !!*

# Verrous et pistes pour le futur

## *Pas de mesure de la qualité linguistique*

- *Corrélation avec l'évaluation subjective en question*
- *Mauvais scores des traductions humaines de qualité*
- *Pas de diagnostic sur les faiblesses des systèmes*

## Proposition

- Mesurer l'effort minimal requis pour produire une bonne traduction à partir de la traduction candidate
- produire de nouvelles références comme effet de bord

# Verrous et pistes pour le futur

## *Pas de mesure de l'utilisabilité*

- *Qualité linguistique non corrélée avec l'utilisabilité*
  - *Systran-Euratom (1972) : Qualité=1/5, utilité=4,5/5*
  - *Transcription de monologues ou de dialogues interprétés*

## Proposition

- Mesurer l'utilisabilité avec des utilisateurs durant le développement
- Mettre des prototypes sur le web (collecte de données)
- Montrer les résultats
- Essayer de corréler utilisabilité et mesures "usuelles"

# Verrous et pistes pour le futur

## *D'autres critères doivent être pris en compte*

### ● Ergonomie (utilité, utilisabilité)

- Vitesse

- Flexibilité

- Le système accepte-il des dictionnaires utilisateur ?

- Le système peut-il "apprendre" ?

JEIDA, EAGLES, FEMTI  
[Hovy & al. 02, MT vol. 17]

### ● Facteurs de conception interne

- Facilité de maintenance et d'amélioration du linguiciel

- Dictionnaires, traitement de phénomènes linguistiques

- Utilisation du contexte

- Richesse de la structure interface entre composants

# Conclusion – apports personnels

- **en CI écrite réfléchie tout terrain**

- Animation d'une équipe

- 💡 Une architecture pour la DI

- 💡 Des questions compréhensibles

- 💡 Le concept de DAE

- 🔒 Des pistes de recherche pour aller plus loin

- Thèse en cours *[Supeeti Kulchan ; thèse franco-thai]*

- 💡 **La DI pour l'acquisition de qualité**

- Un Thaï non francophone veut obtenir une traduction de qualité en thaï d'un document en français

# Conclusion – apports personnels

- **en CI orale spontanée spécialisée**

- Animation d'une équipe

- Des résultats au niveau de l'état de l'art

- 💡 Analyse par patrons & génération concaténative

- 🔒 Des pistes de recherche originales qui se démarquent de la mode

- Thèse en cours *[Achille Falaise ; CIFRE Prosodie]*

- autour de la TA de tchat

- **Objectif : rester dans la course**

- Mobiliser les énergies (pas assez nombreux, des linguistes)

# Conclusion – réflexions générales

## en traduction automatique

- Premier Constat

- La DI n'est plus un sujet de recherche qui réunit une communauté importante

- Pourtant ...

- La DI est la clé de la démocratisation de la TA de qualité

# Conclusion – réflexions générales

## en traduction automatique

- La DI est la clé de la démocratisation de la TA
- **Orientation pour le futur**
  - Jouer la complémentarité entre Heuristiques & DI
    - Apprentissage, personnalisation
    - Certification du sens
- Un encouragement
  - Les industriels viennent à la DI
    - Pionniers : Systran (écrit), Spoken Inc. (oral)

# Conclusion – réflexions générales

## en traduction automatique

- **Second constat**
  - Engouement général pour les méthodes statistiques
- Un changement en cours
  - De l'exploitation de formes de surface ...
    - *On n'a plus besoin de linguistes*
  - ... vers l'exploitation de structures
    - *On refait appel aux linguistes*

# Conclusion – réflexions générales

## en traduction automatique

- Engouement général pour les méthodes statistiques
- **Orientation pour le futur**
  - Jouer la complémentarité des approches ...
    - linguistique, statistique & par analogie (méthodes mixtes)
  - ... elles ont montré leur utilité
    - Exemple : [MASTOR - IBM]

*Il faut éviter d'être sectaire !*

# Conclusion – réflexions générales

## en traduction de parole

- **Constat**
  - Les approches actuelles limitent les performances
- **Pourtant ...**
  - La traduction de parole est l'héritière de la communication entre personnes et systèmes d'information

# Conclusion – réflexions générales

## en traduction de parole

- La traduction de parole est l'héritière de la communication entre personnes et systèmes d'information
- **Orientation pour le futur**
  - Mettre en œuvre la gestion dialogue (le savoir-faire existe)
  - Dépasser l'intégration sous forme de boîte noire
    - Interaction mutuelle entre reconnaissance et analyse
    - Interaction entre génération et synthèse
  - Ajouter de la DI [Spoken Inc.]

# Conclusion – réflexions générales

## en évaluation

- **Constat**

- L'évaluation objective est une bonne idée, mais sa mise en œuvre n'est pas satisfaisante
  - Les méthodes de mesures actuelles mesurent seulement la similarité entre 2 textes
  - Beaucoup de méthodes de mesures sont proposées

- **Pourtant ...**

- l'important est de savoir si la tâche est correctement accomplie ou non

# Conclusion – réflexions générales

## en évaluation

- l'important est de savoir si la tâche est correctement accomplie ou non
- **Orientation pour le futur**
  - Faire des évaluations liées à la tâche ...
    - Mesurer le coût de la post-édition
      - Produire de nouvelles traductions de référence proches des systèmes à évaluer par effet de bord
    - Diagnostiquer les faiblesses des systèmes
  - ... et mutualiser l'évaluation subjective [PolyphraZ]

*Merci !*