

Modélisation cognitive d'interactions humaines dans un cadre de planification multi-agents

Soutenance de thèse

Alexandre Pauchet

Laboratoire Informatique de Paris Nord
Institut Galilée - Université Paris 13

Vendredi 22 Septembre 2006

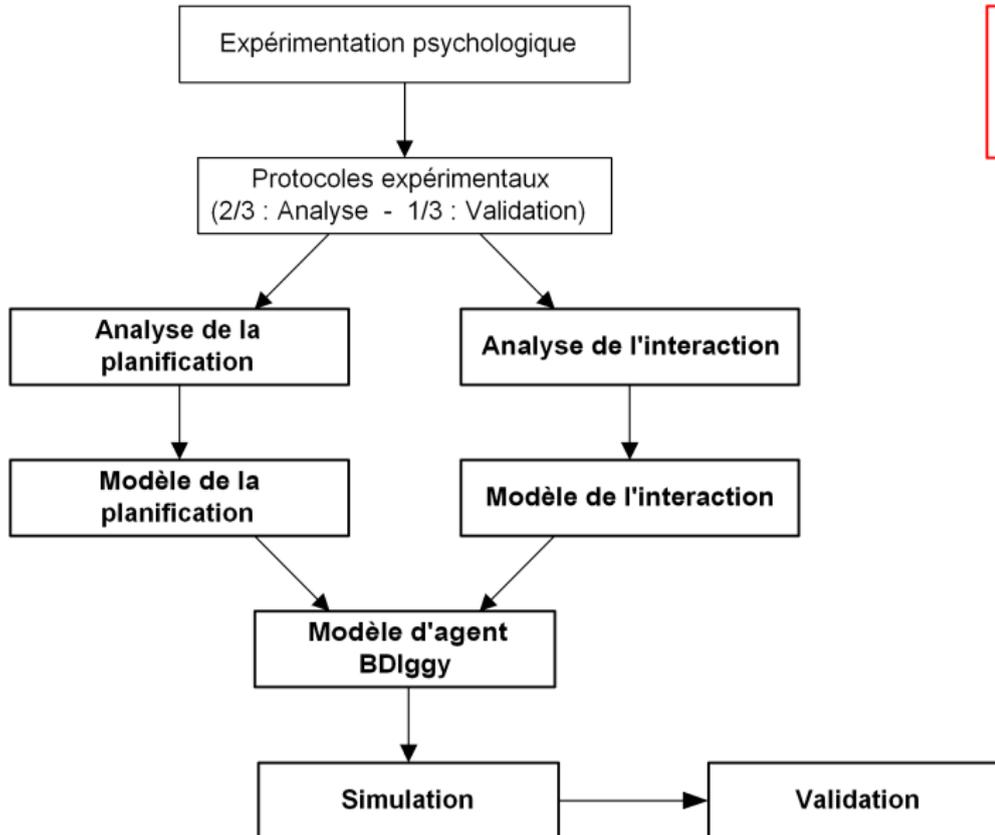
Encadrantes

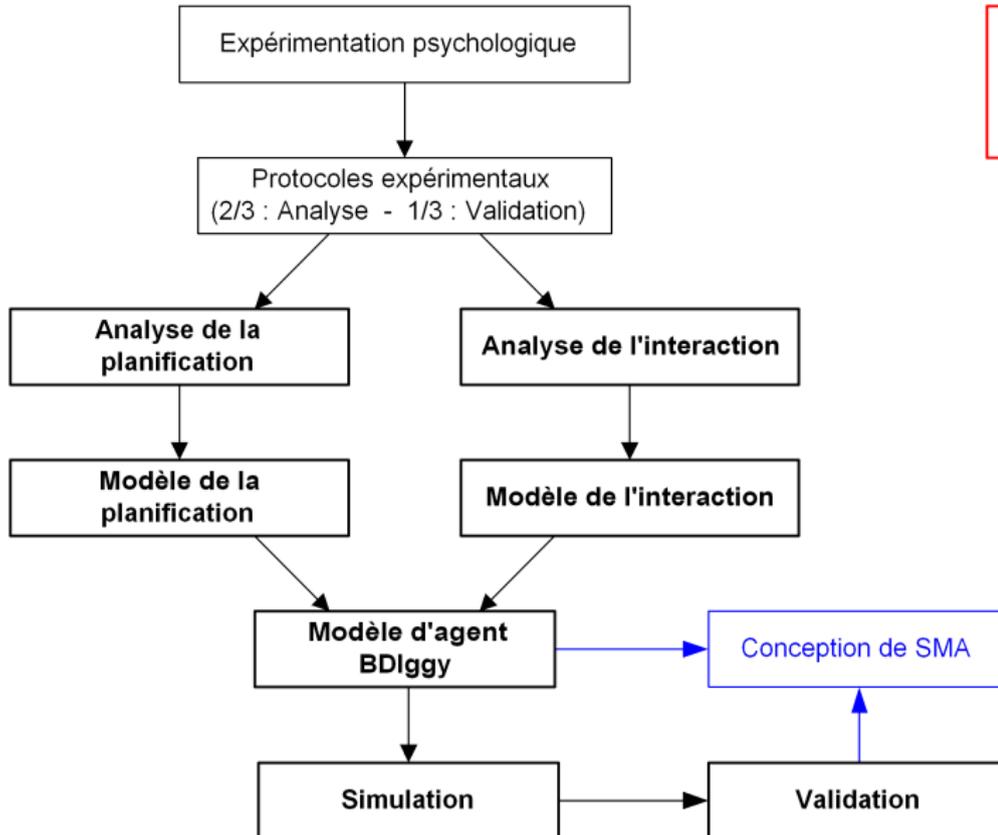
Amal El Fallah Seghrouchni (LIP6)
Nathalie Chaignaud (PSI)

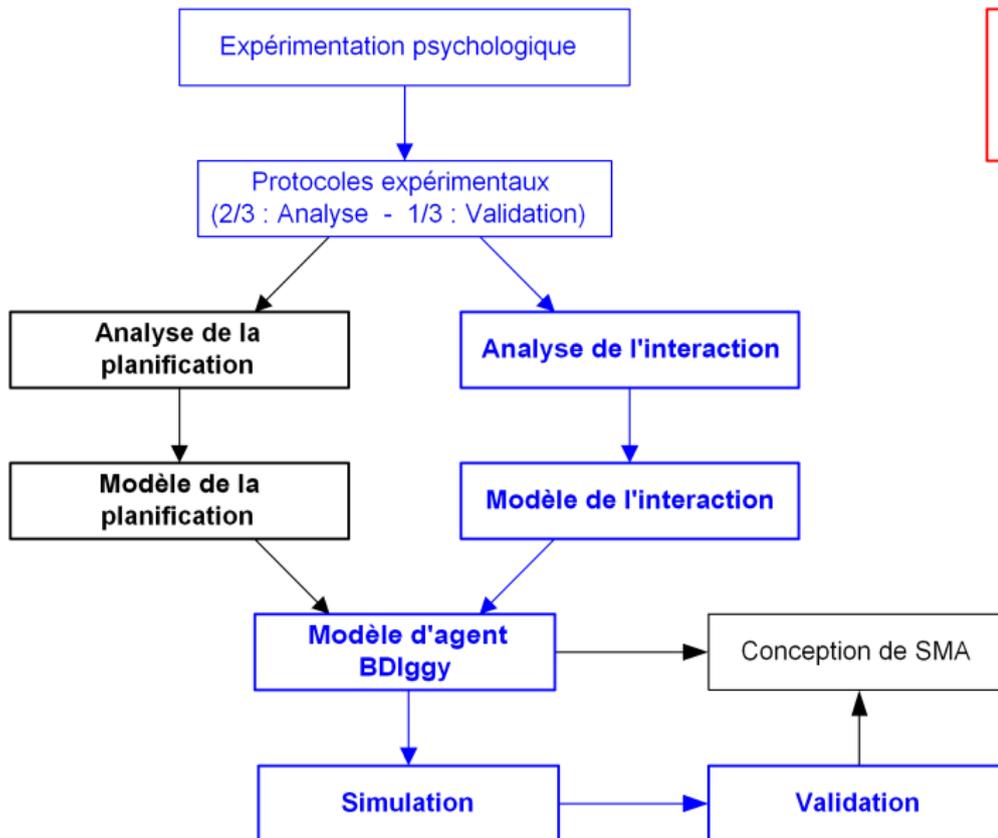
- La modélisation cognitive permet d'étudier les processus humains de communication et de raisonnement afin d'améliorer la conception des SMA.
- Les SMA fournissent des outils de simulation pour valider les modèles cognitifs.

But

analyse, modélisation et **simulation** des capacités humaines de planification et d'interaction dans un cadre de planification multi-agents.







Plan de la présentation

- 1 Positionnement
- 2 Projet de l'agence de voyage
- 3 Analyse des interactions humaines
- 4 Modélisation des interactions humaines
- 5 Modèle d'agent : architecture BDI_{GGY}
- 6 Simulation et validation
- 7 Conclusions et perspectives

- 1 **Positionnement**
- 2 Projet de l'agence de voyage
- 3 Analyse des interactions humaines
- 4 Modélisation des interactions humaines
- 5 Modèle d'agent : architecture BDI_{GGY}
- 6 Simulation et validation
- 7 Conclusions et perspectives

Théorie des actes de langage

- Fondements (Austin70).
- Taxinomie et formalisation en $F(P)$ (Searle72).
- Logique illocutoire (Searle&Vanderveken85).

⇒ **Modèle de l'énoncé.**

Théorie discursive (Vanderveken99)

- Extension du modèle Genevois (Roulet&al85).
⇒ **Structuration des dialogues.**
- Extension des actes de langage.
⇒ **Satisfaction du discours.**

Positionnement (2/3)

Communication entre agents logiciels

Langages de Communication entre Agents (ACL) :
KQML (Finin&al92), FIPA-ACL (Labrou00)

- F(P)=performative(énoncé).
- Communication, message, contenu.
- Sémantique des performatives.

Protocoles de communication FIPA (Labrou00)

Ils définissent le type et l'enchaînement des messages.

- ⇒ **Conçus pour les communications entre agents logiciels.**
- ⇒ **Utilisables pour des dialogues humains ?**

Architectures issues de la modélisation cognitive

- Prédominance des systèmes SOAR (Newel&Laird92) et ACT-R (Anderson98).
- MicroPsi (Bach03) : architecture d'agent modélisant l'interaction entre émotion, planification et cognition.

Architectures issues de la résolution coopérative de problèmes

- (Allen&al02) : Résolution coopérative de problèmes.
- Collagen (Rich&Sidner98) : gestionnaire de collaboration entre un agent humain et un agent logiciel.

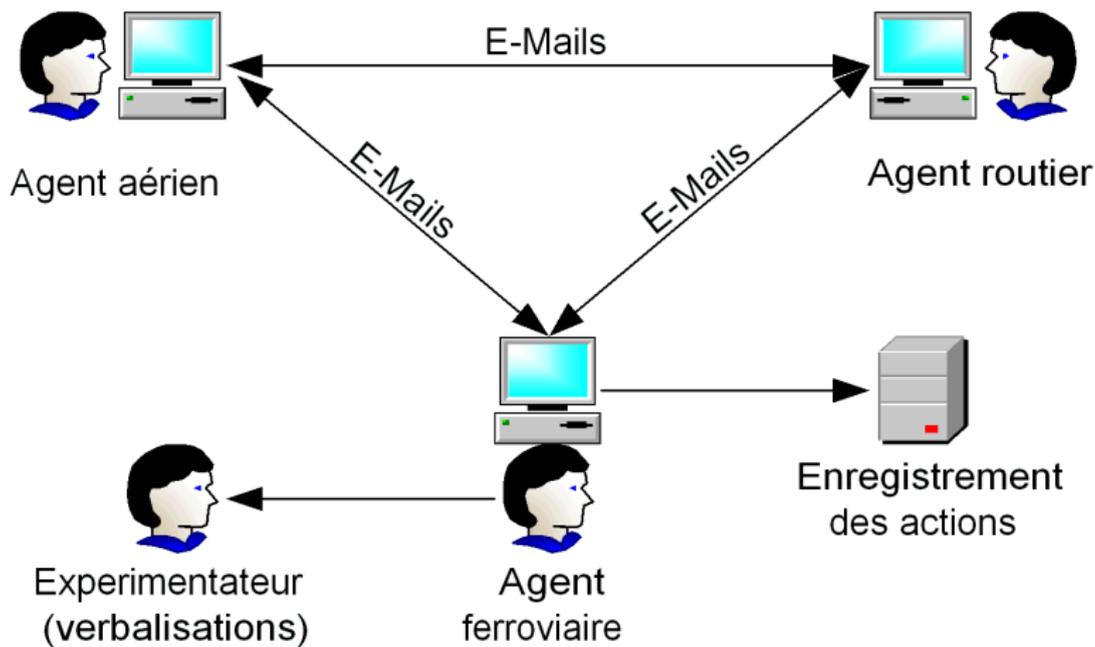
Architectures BDI

- Travaux sur l'intentionnalité (Bratman87)
- PRS (Georgeff87), dMARS (Inverno&al97).

- 1 Positionnement
- 2 **Projet de l'agence de voyage**
- 3 Analyse des interactions humaines
- 4 Modélisation des interactions humaines
- 5 Modèle d'agent : architecture BDI_{GGY}
- 6 Simulation et validation
- 7 Conclusions et perspectives

Projet de l'agence de voyage (1/3)

Expérimentation psychologique



Projet de l'agence de voyage (2/3)

Exemple de protocole expérimental

Enregistrement des actions	Verbalisation concomitante
<p>[09:31:52] Sujet: alexandre [09:31:52] Rôle: Agent aérien [09:31:52] Problème: Aller [09:31:52] Date: Mardi 5 Mai 1998 17:32:43 [09:31:52] 1, Demande horaires Annecy-Annecy (Pach'AIR) [09:31:55] 2, Demande horaires Annecy-Annecy (Pigeon Vol) [09:31:59] 3, Demande horaires Annecy-Brest (Pigeon Vol) [09:32:03] 4, Demande horaires Annecy-Brest (Pach'AIR) [09:33:45] 5, Nouveau message [09:34:21] 6, Reçoit message 1 de Agent ferroviaire [09:34:53] 7, Envoie message 1 De: Agent aérien A: Agent ferroviaire Y a-t-il une gare à Château-Neuf ? [09:34:57] 8, Affiche message numero 1 de Agent ferroviaire De: Agent ferroviaire A: Agent aérien Salut, j'aimerais connaître les prix et les horaires du premier vol NANTES-PARIS pour demain pour 1 personne. merci [09:35:12] 9, Demande horaires Nantes-Paris (Pach'AIR) [09:35:15] 10, Demande horaires Nantes-Paris (Pigeon Vol) [09:35:39] 11, Mémoire horaire [09:35:39] 12, Ajoute horaire dans panneau Mon problème Nantes-Paris (avion Pigeon Vol) Date: Demain -- 1 personne 08h40/09h35 - 595 F [09:35:43] 13, Envoie horaire Nantes-Paris (avion) [09:36:05] 14, Reçoit message 1 de Agent routier [09:36:43] 15, Envoie message 2 De: Agent aérien A: Agent ferroviaire C'est le plus tôt, le dernier arrive à 20h50, tous à 595F =====HORAIRE===== Nantes-Paris (avion Pigeon Vol) Date: Demain -- 1 personne 08h40/09h35 - 595 F ...</p>	<p>"Si, je mets 4 est-ce qu'il va me mettre le prix..... "</p> <p>"Alors Château-Neuf départ et arrivée Maguelone... Donc Château-Neuf/Orléans en.... en... Château-Neuf/Orléans... Donc là... Je me demande si je l'envoie en train directement à Maguelone ou en avion depuis Paris... 3000 F... Donc ils iront en train..." "Ah! Il faut que je demande s'il y a une gare à Château-Neuf, on ne sait jamais.... "</p> <p>"Il est drôle lui... Je vais lui envoyer l'horaire le plus tôt."</p> <p>...</p>

- Planification séquentielle/parallèle.
- Construction de plans de manière : prospective, rétrospective, centrifuge, centripète.
- Modèle de planification humaine (Chaignaud96) réutilisable : phase, état d'esprit, stratégies, tactiques, observations et personnalité.
- Implantation : réutilisation du système IGGY (Chaignaud00).

⇒ **Implémentation du système IGGY2** (analyse complète des protocoles).

- 1 Positionnement
- 2 Projet de l'agence de voyage
- 3 **Analyse des interactions humaines**
- 4 Modélisation des interactions humaines
- 5 Modèle d'agent : architecture BDI_{GGY}
- 6 Simulation et validation
- 7 Conclusions et perspectives

2/3 des protocoles expérimentaux sont analysés

- Envois et réceptions de message conservés.
- Fusion par trinôme.

Au niveau des énoncés

- Liste de performatives pertinentes (FIPA-ACL, KQML).
- Lien $F(P)$: performatives appliquées à des états mentaux (Théorie des actes de langage).

Au niveau du discours

- Regroupement en échanges des performatives thématiquement liées (Vanderveken).

Analyse des interactions humaines (2/3)

Liste des performatives observées (KQML, FIPA-ACL)

Performatives	Actes de langage	Occurrence
<i>query</i>	directif	474 (42,97%)
<i>reply</i>	descriptif	437 (39,62%)
<i>refine</i>	directif	60 (5,44%)
<i>thank</i>	descriptif	30 (2,72%)
<i>propose</i>	engageant	29 (2,63%)
<i>inform</i>	descriptif	26 (2,36%)
<i>cancel</i>	directif	18 (1,63%)
<i>acceptProposal</i>	directif	10 (0,91%)
<i>notUnderstood</i>	descriptif	9 (0,82%)
<i>refuseProposal</i>	directif	3 (0,27%)
total		1103 (100%)

Analyse des interactions humaines (3/3)

Liste des échanges observés

Échanges: séries d'énoncés guidés par un but discursif.

Échange	Intention discursive	Performative d'ouverture	Performative de clôture
<i>Demande d'information</i>	directif	query	reply
<i>Proposition d'information</i>	engageant	propose	reply
<i>Envoi spontané d'information</i>	descriptif	inform	inform
<i>Traitement d'erreur</i>	directif	-	cancel

Plan de la présentation

- 1 Positionnement
- 2 Projet de l'agence de voyage
- 3 Analyse des interactions humaines
- 4 **Modélisation des interactions humaines**
- 5 Modèle d'agent : architecture BDI_{GGY}
- 6 Simulation et validation
- 7 Conclusions et perspectives

Modélisation des interactions humaines (1/5)

Modélisation des énoncés

Théorie des actes de langage (Searle72): $F(P)$
→performative(état mental)

directif(désir du locuteur)

air→railway

query(pD(air pStage(Orleans Paris ? <(10h30) train ? ? false)))

descriptif(croyance)

railway→air

reply(pB(pStage(Orleans Paris 08h25 09h30 train 1 80 false)))

engageant(désir de l'interlocuteur)

railway→air

propose(pD(air pStage(Orleans Paris ? ? train ? ? false)))

Problématique

- Représentation fidèle des échanges observés.
- Type et enchaînement des performatives échangées.
- Modéliser la temporalité des échanges.

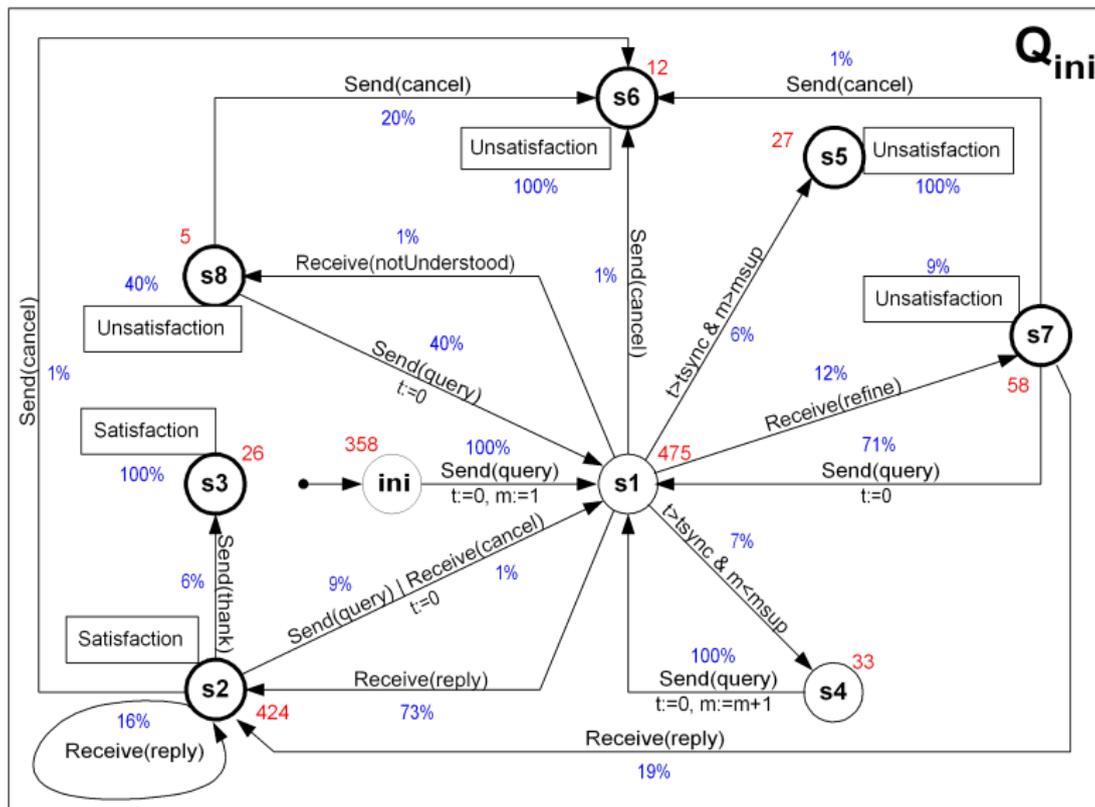
Représentation par automates temporisés (Alur94)

4 paires d'automates temporisés :

- Demande d'information (Q_{ini} et Q_{int}).
- Proposition d'information (P_{ini} et P_{int}).
- Envoi spontané d'information (I_{ini} et I_{int}).
- Traitement des erreurs (E_{ini} et E_{int}).

Modélisation des interactions humaines (3/5)

Exemple d'automate temporisé



- 4 paires d'automates temporisés couvrent l'intégralité des dialogues.
⇒ **Faible complexité.**
- Utilisation possible dans les SMA. Ces automates permettent à l'agent de :
 - générer des messages,
 - interpréter les messages reçus.⇒ **Gestion des attendus des conversations.**

Modélisation des interactions humaines (5/5)

Sémantique des performatives : query

$$M = pMessage(A_S A_R \text{ query } pD(A_S S))$$

$$\left[\begin{array}{l} pD(A_S S) \\ pB(!pMeans(S)) \\ !pB(S) \\ !pB(!S) \end{array} \right] \frac{Q_{ini}(ini, s2, s4, s7, s8) \xrightarrow{Send(query)} Q_{ini}(s1)}{aAdd(pB(pSent(M))); aUpdateTA(M)}$$

$$\frac{Q_{int}(ini, s1, s2, s4, s5) \xrightarrow{Receive(query)} Q_{int}(s1)}{aAdd(pB(pD(A_S S))); aUpdateTA(M)}$$

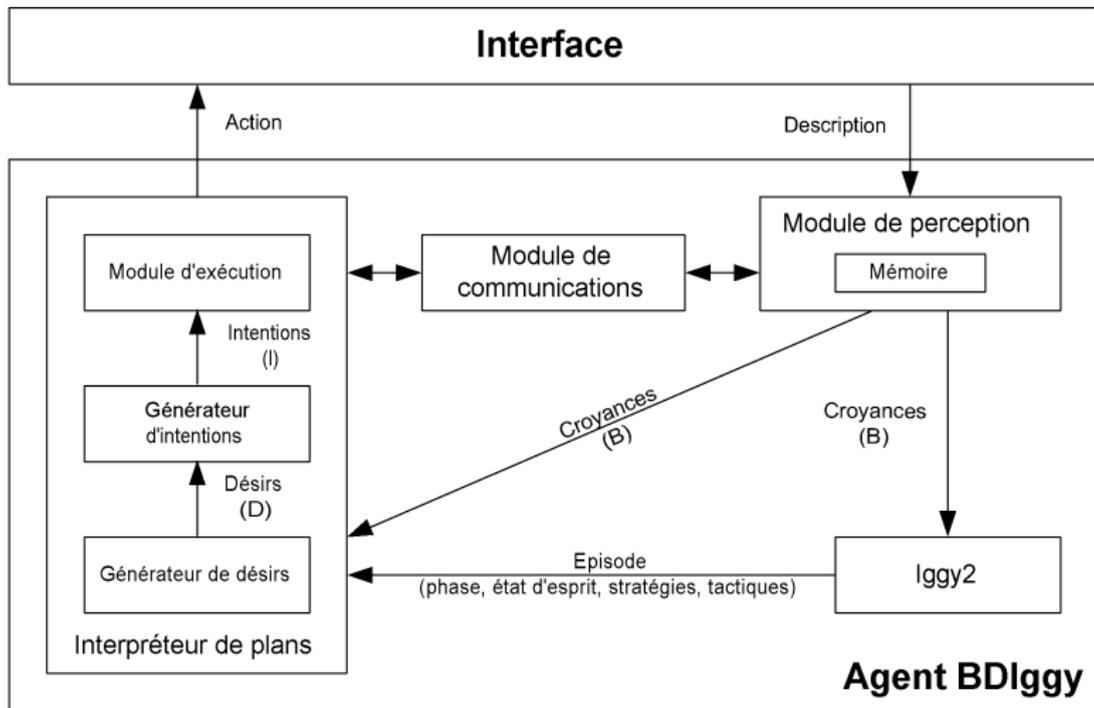
Plan de la présentation

- 1 Positionnement
- 2 Projet de l'agence de voyage
- 3 Analyse des interactions humaines
- 4 Modélisation des interactions humaines
- 5 **Modèle d'agent : architecture BDIggy**
- 6 Simulation et validation
- 7 Conclusions et perspectives

- Modèle de planification (IGGY2).
- Modèles d'interaction :
 - Modèle de l'énoncé (Performatives(états mentaux)).
 - Modèle de la dynamique conversationnelle (automates temporisés).
 - Lien énoncés/discours/états mentaux (sémantique des performatives).
- Intégration homogène : BDI.

Architecture d'agent (2/7)

BDIGGY



Architecture d'agent (3/7)

Le module de perception et la mémoire

Mémoire à court terme	Représentation de l'interface	Mémoire dynamique
	Description des différents panneaux de l'interface	
	Croyances non traitées	
	Croyances sur des désirs à évaluer Croyances sur l'état des échanges en cours	
Mémoire à long terme	Croyances persistantes	Mémoire statique
	Informations obtenues par la base de données Informations obtenues auprès des autres agents	
	Représentation du problème	
	Problème à résoudre Représentation de la carte (villes, distances et transports) Connaissances implicites	

Modèle humain de résolution coopérative de problèmes.

Architecture de tableau noir à contrôle hiérarchique

- Réutilisation du moteur d'IGGY.
- Spécialistes du domaine (dépendant du domaine).
- Interpréteur Lisp en Java.

Intégration à une architecture proche des BDI

- Observations \approx sous-partie des croyances.
- Episode \approx plan abstrait.

Architecture d'agent (5/7)

L'interpréteur de plans

Générateur de désirs

- Interprète un *épisode* comme un plan abstrait.
- Types de désir : *abstrait/élémentaire, instancié/non instancié, satisfait/insatisfait, réussi/non réussi, actif/en attente*.
- Gère la(les) pile(s) de désirs de l'agent (purge, tri, etc.).

Générateur d'intentions

- Traduit les désirs en intentions.
- Types d'intentions : *active/en attente*.
- Gère la(les) pile(s) d'intentions de l'agent (purge, tri, etc.).

Module d'exécution

- Exécute les actions intentionnées directement sur l'interface.
- Fait appel au module de communication.

Architecture d'agent (6/7)

Exemple de fonctionnement de l'interpréteur de plans

Iggy2 : prospective, ... ,(Paris, optimized, cheapest)

Pile de desirs : [
pD(railway pProblem(railway true))
]

Pile d'intentions : []

Architecture d'agent (6/7)

Exemple de fonctionnement de l'interpréteur de plans

Iggy2 : prospective, ..., (Paris, optimized, cheapest)

Pile de desirs :

```
pD(railway pTravel(pStage(Beaupreau Baisieux * ≤(11:00) *  
    1 ≤(1310) true))  
pD(railway pProblem(railway true))  
]
```

Pile d'intentions : []

Architecture d'agent (6/7)

Exemple de fonctionnement de l'interpréteur de plans

Iggy2 : prospective, ..., (Paris, optimized, cheapest)

Pile de desirs :

```
pD(railway pStage(Beaupreau Nantes * * coach * * *))  
pD(railway pTravel(pStage(Beaupreau Nantes * * coach * * *)  
                  pStage(Nantes Baisieux * * * * * *))  
pD(railway pProblem(railway true))  
]
```

Pile d'intentions : []

Architecture d'agent (6/7)

Exemple de fonctionnement de l'interpréteur de plans

Iggy2 : prospective, ..., (Paris, optimized, cheapest)

Pile de desirs :

```
pD(railway pStage(Baisieux Nantes * * coach * * *))
pD(railway pTravel(pStage(Beaupreau Nantes * * coach * * *)
                  pStage(Nantes Baisieux * * * * * *))
pD(railway pProblem(railway true))
]
```

Pile d'intentions :

```
pl(AskForStage,road,pStage(Beaupreau Nantes * * coach * * *))
  — sendMessage,%agent,pD(%role %stage)
  — waitMessage,%agent,pMessage(%agent %role reply pB(%stage))
    OR pMessage(%agent %role reply pB(!%stage))
  — memorize,pMessage(%agent %role reply %stage)
]
```

Architecture d'agent (6/7)

Exemple de fonctionnement de l'interpréteur de plans

Iggy2 : prospective, ..., (Paris, optimized, cheapest)

Pile de desirs : [

pD(railway pStage(Baisieux Nantes * * coach * * *))

pD(railway pTravel(pStage(Beaupreau Nantes * * coach * * *)
pStage(Nantes Baisieux * * * * * *))

pD(railway pProblem(railway true))

]

Pile d'intentions : [

pl(AskForStage,road,pStage(Beaupreau Nantes * * coach * * *)

— waitMessage,%agent,pMessage(%agent %role reply pB(%stage))
OR pMessage(%agent %role reply pB(!%stage))

— memorize,pMessage(%agent %role reply %stage)

]

Il contient

- le modèle des énoncés (performatives(états mentaux)),
- le modèle de la dynamique conversationnelle (automates temporisés),
- la sémantique des performatives.

Il est appelé

- par le module de perception à la réception d'un message,
- par le module d'exécution à l'envoi d'un message.

Plan de la présentation

- 1 Positionnement
- 2 Projet de l'agence de voyage
- 3 Analyse des interactions humaines
- 4 Modélisation des interactions humaines
- 5 Modèle d'agent : architecture BDI GGY
- 6 **Simulation et validation**
- 7 Conclusions et perspectives

Simulation et validation (1/2)

Simulation

Simulation de la résolution du problème de l'agence de voyage

- Lancement de 3 agents BDI_{GGY} simultanément.
- Interfaçage avec l'environnement de l'expérimentation.
- Entrée : 1 personnalité pour chaque agent.
- Sortie : 1 protocole généré par agent.

Réglage du système

- Utilisation des 2/3 protocoles expérimentaux.
- Comparaison entre les protocoles humains (personnalités) et les protocoles générés (mêmes personnalités).
- Nombreux paramètres : délais dans les automates temporisés, proportions d'utilisation de chaque ingrédient d'IGGY2, nombre moyen d'erreur, etc.

Test à la Turing

Trinômes générés
Trinômes expérimentaux

Mélange de Trinômes → Experts ?

	Ensemble 1		Ensemble 2		Total
	Humain	BDIggly	Humain	BDIggly	
Expert 1	2/3	1/3	1/2	2/4	6/12
Expert 2	1/3	1/3	-	-	2/6
Expert 3	-	-	0/2	2/4	2/6
Expert 4	-	-	2/2	2/4	4/6

Plan de la présentation

- 1 Positionnement
- 2 Projet de l'agence de voyage
- 3 Analyse des interactions humaines
- 4 Modélisation des interactions humaines
- 5 Modèle d'agent : architecture BDI_{GGY}
- 6 Simulation et validation
- 7 **Conclusions et perspectives**

- Analyse des protocoles expérimentaux.
- Système de planification coopérative humaine (IGGY2).
- Modèle de l'interaction humaine :
 - Performatives appliquées à des états mentaux.
 - Automates temporisés.
 - Sémantiques des performatives.
- Planification et interaction sont intégrés de façon homogène à une nouvelle architecture d'agent (BDIGGY).
- Validation des modèles cognitifs et de l'architecture.
 - Simulation de la résolution humaine grâce à des agents BDIGGY.
 - Comparaison des protocoles expérimentaux et des protocoles générés (Test à *la Turing*).

- Intégration du traitement automatique de la langue naturelle.
 - Modules de TAL.
 - Actes de langage indirectes (Searle72).
- Intégration de l'apprentissage dans les automates temporisés.
 - Apprentissage non-supervisé.
 - Apprentissage supervisé.
- Extension à la résolution collaborative de problèmes.
- Extension à d'autres modes d'interaction (ACA, interaction multimodale).
- Conception de SMA hétérogènes.

- Pauchet A., Chaignaud N., El Fallah Seghrouchni A. : *A Cognitive Model of Interaction for Software Agents*, IEEE IAT, Compiègne, France, pp 281-284 (2005).
- Pauchet A., Chaignaud N., El Fallah Seghrouchni A., *Un modèle cognitive de l'interaction pour agents logiciels*, MFI, Caen, pp. 183-192, 2005.
- Pauchet A., Chaignaud N., El Fallah Seghrouchni A., *BDIGGY: a Cognitive Architecture for Cooperative Problem Solving*, IEEE IAT, Halifax, Canada, pp. 523-526, 2003.
- Pauchet A., Chaignaud N., El Fallah Seghrouchni A., *Modelling Planning and Interaction in the Framework of Human Cooperative Problem Solving*, AAMAS Workshop on Human and Multi-Agent Systems, Melbourne, Australia, pp. 31-38, 2003.
- Pauchet A., Chaignaud N., El Fallah Seghrouchni A., *Résolution coopérative de problèmes intégrant planification et interaction*, JFSMA, Hammamet, Tunisia, pp. 309-313, 2003.