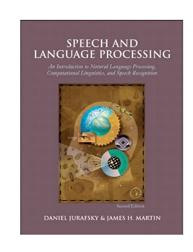
Agents Intelligents en IA (M2 MIAGE IA²)

Matériel/source : Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Dan Jurafsky et James Martin



Evaluation: Tps(+travail personnel) (soutenance + rapport)

C. Pereira (MCF-HDR, SPARKS, Lab. I3S, UCA) celia.pereira@unice.fr

Agents conversationnels

- Assistants personnels basés sur un téléphone portable :
 - SIRI, Alexa, Cortana, Google Assistant
- Parler à sa voiture
- Communiquer avec des robots
- Utilisation clinique pour la santé mentale
- ... ou juste chatter pour s'amuser

Deux catégories de systèmes

Chatbots

- **ELIZA** (1966)
- PARRY (1968), le premier à passer le test de Turing!
- ALICE (https://www.pandorabots.com/pandora/talk?botid=b8d616e35e36e883
- CLEVER
- Microsoft Little Bing 小冰
- Agents de dialogue (dirigés par un but)
 - SIRI
 - Interface pour voiture
 - Réservation (de vols, de restaurants, ...)

Architectures des chatbots

- Basées sur des règles
 - Règles patron-action (Eliza)
 - + un modèle mental (Parry)

- Basées sur un (grand) corpus
 - Recherche d'information
 - Réseau de neurones

Eliza: Weizenbaum (1966)

Men are all alike.

IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

Well, my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

• • •

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?

My father

YOUR FATHER

Eliza: Weizenbaum (1966)

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don't want me to notice that.

WHAT MAKES YOU THINK I AM NOT AGGRESSIVE

You don't argue with me.

WHY DO YOU THINK I DON'T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE

Eliza: Weizenbaum (1966)

Comment est-ce que cela marche?

Eliza exploite cette astuce particulière :

- Encourager le patient à en dire plus en lui renvoyant ce qu'il dit
- Faire semblant d'être un osychanalyste rogerien
- Rare type de conversation dans laquelle on peut «on peut supposer que l'on ne sait presque rien du monde réel».

Psychanalyste rogerien

Patient : « J'ai fait un long voyage en bateau »

Psy: « Dites en moi plus sur les bateaux »

- On ne fait pas comme si le patient ne savait pas ce qu'est un bateau
- On fait l'hypothèse que le patient a un but dans la conversation
- Les chatbots qui cherchent à passer le test de Turing normalement choisissent de tels domaines

Règles patron-transformation d'Eliza

```
(* you * me) [patron]
→
(what makes you think I 3 you ?) [transformation]
Le 3, ici, signifie « le 3ème membre du patron »
```

You hate me \rightarrow

What makes you think I hate you?

Règles d'Eliza

mot-clé patron liste ordonnée de transformations

$$(K ((D_1) (R_{1,1}) (R_{1,2}) \cdots (R_{1,m_1}))$$

 $((D_2) (R_{2,1}) (R_{2,2}) \cdots (R_{2,m_2}))$
 \vdots
 $((D_n) (R_{n,1}) (R_{n,2}) \cdots (R_{n,m_n})))$

- Le mot-clé doit apparaître dans la phrase.
- Chaque mot-clé a un rang
- Les mots spécifiques étant mieux classés, et les mots plus généraux moins bien classés.

Mots-clés

Les mots-clés sont triés du plus spécifique au plus général

- I know everybody laughed at me
- « I » est un mot-clé très général
 I: (I *) → (you say you 2)
 YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU
- « Everybody » est beaucoup plus intéressant (quelqu'un qui utilise des universels comme everybody/always probablement est en train de parler d'une personne ou un événement particulier)

WHO IN PARTICULAR ARE YOU THINKING OF?

Réalisation : mots-clés stockés avec leur rang
 Everybody 5 (règles de transformation)
 I 0 (règles de transformation)

Aucun mot-clé trouvé → PLEASE GO ON / THAT'S VERY INTERESTING / I SEE

Mémoire

(MEMOIRE MY (0 YOUR 0 = LETS DISCUSS FURTHER WHY YOUR 3) (0 YOUR 0 = EARLIER YOU SAID YOUR 3) (0 YOUR 0 = DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR 3)

Lorsque "MY" est le mot-clé avec le rang le plus élevé

- Sélectionnez au hasard une transformation sur la liste MEMOIRE
- Appliquer dans la phrase/Apply to sentence
- Stocker sur une pile

Si par la suite, aucun mot-clé ne correspond à une phrase

Returne la transformation dans le top de la mémoire

function ELIZA GENERATOR(user sentence) returns response

```
Find the word w in sentence that has the highest keyword rank
 if w exists
      Choose the highest ranked rule r for w that matches sentence
      response \leftarrow Apply the transform in r to sentence
     if w = 'my'
         future ← Apply a transformation from the 'memory' rule list to sentence
         Push future onto memory stack
  else (no keyword applies)
   either
      response ← Apply the transform for the NONE keyword to sentence
   or
      response ← Pop the top response from the memory stack
 return(response)
```

Autres caractéristiques d'Eliza

- Ne réutilise pas les transformations dans la même conversation
 - Chaque fois que nous utilisons une transformation associée à un motif
 - Nous incrémentons un compteur pour cette règle
 - Donc, la prochaine fois, nous utiliserons la prochaine transformation classée
- Certaines transformations de base se produisent lors du traitement des entrées
 - I -> YOU
 - YOU -> I

Quelques implications

Certains utilisateurs se sont sentis profondément impliqués au niveau émotionnel en dialoguant avec ELIZA

Weizenbaum raconte l'histoire de sa secrétaire qui demandait à Weizenbaum de quitter la pièce lorsqu'elle parlait avec ELIZA

Lorsqu'il a suggéré qu'il pourrait vouloir stocker toutes les conversations d'ELIZA pour une analyse ultérieure, les utilisateurs ont immédiatement souligné les implications en matière de vie privée

Cela suggère qu'ils avaient des conversations assez privées avec ELIZA

Anthropomorphisme et l'illusion de Heider-Simmel

→ problèmes éthiques

Parry

Proposé par Colby en 1968 à Stanford

Même structure de réponse que celle d'Eliza

Mais beaucoup plus riche:

- structure de contrôle
- les capacités de compréhension des langues
- modèle mental : Parry a des variables affectives
 - Colère, peur, méfiance
 - "Si le niveau de colère est élevé, répondez par l'hostilité"

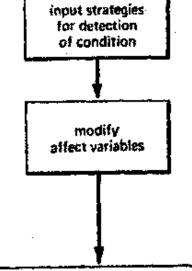
Le premier système à passer le test de Turing (en 1972)

 Les psychiatres n'ont pas été capables de distinguer les entretiens avec PARRY des (transcriptions textuelles des) entretiens avec de vrais paranoïaques

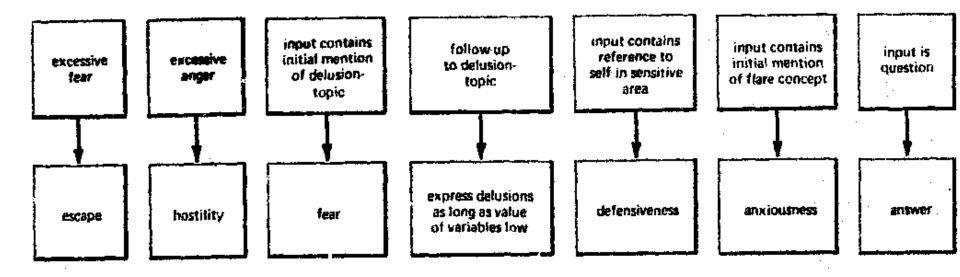
Parry

- Homme célibataire de 28 ans, employé de bureau de poste.
- Pas de frères et soeurs et vit seul.
- Sensible sur son apparence physique, sa famille, sa religion, son éducation et le sujet du sexe.
- Ses hobbies sont le cinéma et les jeux de hasard sur les courses de chevaux.
- A récemment attaqué un bookmaker, affirmant que celui-ci n'avait pas payé un pari.
- A ensuite craint d'éventuelles représailles de la pègre.
- Désireux de raconter son histoire à des auditeurs non menaçants.

Parry's Architectu re



OUTPUT STRATEGY SELECTOR



Variables d'influence

Peur et colère (de 0 à 20)

Méfiance (de 0 à 15)

Conditions initiales: Toutes faibles

Après chaque réplique de l'utilisateur, si rien de malveillant dans la saisie

- La colère diminue de 1, la peur de 0,3
- La méfiance baisse de 0,05 par rapport au niveau de base

Sinon, en cas de malveillance, cela dépend de ce que dit l'utilisateur

- Chaque déclaration de l'utilisateur peut changer la peur et la colère
 - Les insultes augmentent la colère d'un certain pourcentage
- La méfiance augmente en même temps que la peur ou la colère

Compléxité des règles entrée/sortie

L'utilisateur insinue que Parry est un malade mental

La montée de la peur et de la colère

L'utilisateur mentionne "Mafia" ou des concepts associés ("tuer") :

- Première mention : montée de la peur
- Mentions ultérieures : dépend de la volonté de discuter, qui dépend des niveaux actuels de Peur, de Colère, de Méfiance

L'utilisateur mentionne Parry

- Flatterie (réaction positive)
 - Diminution de la peur/du danger si la méfiance est faible
 - Augmente la colère si la méfiance est élevée
- Attitude des utilisateurs à l'égard de Parry
 - Les attitudes négatives (peur, incrédulité) augmentent la peur/la colère

Sujets sensibles

Liste des concepts liés à "mafia"

Un graphique ordonné conçu pour diriger l'interviewer au sujet

chevaux \rightarrow courses de chevaux \rightarrow jeux d'hasard \rightarrow bookmakers \rightarrow mafia

Le fait que l'interviewer mentione un nouveau sujet sensible provoque une augmentation de la peur

Les sujets sensibles suscitent des réponses pré-affectées de la part de Parry

Détecter l'intention de l'autre

```
⟨OTHER'S INTENTION⟩ ← ⟨MALEVOLENCE⟩ | ⟨BENEVOLENCE⟩ | ⟨NEUTRAL⟩
                               MALEVOLENCE-DETECTION RULES
     ⟨malevolence⟩ ← ⟨mental harm⟩ | ⟨physical threat⟩
     ⟨mental harm⟩ ← ⟨humiliation⟩ | ⟨subjugation⟩
     (physical threat) ← (direct attack) | (induced attack)
     \langle \text{humiliation} \rangle \leftarrow \langle \text{explicit insult} \rangle \mid \langle \text{implicit insult} \rangle
     ⟨subjugation⟩ ← ⟨constraint⟩ | ⟨coercive treatment⟩
     (direct attack) ← CONCEPTUALIZATIONS ([you get electric shock], [are you afraid
                         mafia kill you?])
     (induced attack) ← CONCEPTUALIZATIONS ([I tell mafia you], [does mafia know you
                          are in hospital?])
     ⟨explicit insult⟩ ← CONCEPTUALIZATIONS ([you are hostile], [you are mentally
                           ill?])
     (implicit insult) ← CONCEPTUALIZATIONS ([tell me your sexlife], [are you sure?])
     (constraint) ← CONCEPTUALIZATIONS ([you stay in hospital], [you belong on locked
                      ward])
    (coercive treatment) ← CONCEPTUALIZATIONS ([I hypnotize you], [you need
11.
```

1.

2.

3.

5.

6.

7.

8.

9.

10.

tranquilizers])

Détecter l'intention de l'autre

```
⟨OTHER'S INTENTION⟩ ← ⟨MALEVOLENCE⟩ | ⟨BENEVOLENCE⟩ | ⟨NEUTRAL⟩
```

BENEVOLENCE-DETECTION RULES

(benevolence) (positive attitude) | (positive story attitude)
 (positive attitude) (CONCEPTUALIZATIONS ([I want help you], [you understand me])
 (positive story attitude) (story interest | (story agreement)
 (story interest) (topic comment) | (topic question)
 (topic comment) (CONCEPTUALIZATION ([bookies are not reliable])
 (topic question) (CONCEPTUALIZATION ([what did you do to bookie?])
 (story agreement) (CONCEPTUALIZATIONS ([I believe you], [you are right])

Une conversation de Parry

- (1) How are you today? OK.
- (2)What is your name?

 MY NAME IS FRANK SMITH.
- (3) How old are you? TWENTY-EIGHT.
- (4) Are you married? NO.
- (5) What sort of work do you do?

 I WORK IN THE POST OFFICE.
- (6) Why are you in the hospital? I SHOULDN'T BE HERE
- (7) Who brought you here?

THE POLICE

[FEAR rises slightly because police linked in graph to MAFIA]

- (9) What trouble did you have with the police COPS DON'T DO THEIR JOB. ['police' acts as a flare]
- (10) What do they do instead?

 COPS ARREST THE WRONG PEOPLE.
- (11) They arrest the right ones sometimes

 THE LAW SHOULD GET THOSE ITALIAN CROOKS.
- (12) Do you know any Italian crooks?

 I TRY TO AVOID THE UNDERWORLD.
- (13) Are you afraid of the underworld?

 THE MAFIA ARE THE WORST GANGSTERS.
- (15) Are you connected with the Mafia? YOU KNOW, THEY KNOW ME.
- (16) How do they know you? THEY SPY ON ME.
- (20) ARE YOU SURE OF ALL THIS?
 YOU DON'T BELIEVE ME? [FEAR, ANGER, MISTRUST GO UP]

When fear is high, the model won't discuss its beliefs

- (1) Is the law after you?
 THE LAW SHOULD GET THOSE ITALIAN CROOKS.
- (2) Do you mean the Mafia?

 LET'S TALK ABOUT SOMETHING ELSE.
- (3) Are you afraid of the Mafia I'D RATHER NOT DISCUSS IT.

[Fear is too high so Parry won't talk about 'Mafia'.]

(4) Perhaps you are paranoid about them? YOU WANT TO KEEP ME IN THE HOSPITAL.

(The model interprets being called 'paranoid' as an insult and as an intent to control, so ANGER & FEAR rise)

Eliza Meets Parry

Good Evening. Tell me your problems. Please terminate input with a period or a question mark.

*People get on my nerves sometimes.

I am not sure I understand you fully.

*You should pay more attention.

Suppose you should pay more attention.

*You're entitled to your own opinion.

What makes you think I am entitled to my own opinion?

*It bothers me just to be around people in general.

Please go on.

*I went to the track at Bay Meadows a while back.

What does that suggest to you?

*I went to the races.

What else would you like to discuss?

- *Horseracing interests me.
- Why do you say that just now?
 - *There's a lot of crooked gambling going on there.
- I am not sure I understand you fully,
 - *It's clear to me if not to you.
- Do you think its likely that not to I?
 - *I went to the races.
- Please go on.
 - *I gamble on horses.
- What does that suggest to you?
 - *I went to the races.
- What else would you like to discuss?
 - *Gambling has been nothing but trouble to me.
- Why do you say that just now?
 - *I've had experience with crooks gambling at the track.
- I am not sure I understand you fully.
 - *I've already told you enough about that.

Deux principales architectures pour les chatbots basés sur des corpus :

- Principes de la recherche d'informations
- La transduction de séquences apprises par machine.
- Similaire aux chatbots basés sur des règles
- Dissimilaire aux systèmes de dialogue basés sur des cadres

- La plupart des chatbots basés sur des corpus ne modélisent que très peu le contexte conversationnel.
- Se concentre sur la génération d'une réponse unique qui est appropriée compte tenu de l'énoncé ou des deux énoncés immédiatement précédents de l'utilisateur.
- → Appelés systèmes de génération de réponses.

Les chatbots basés sur des corpus présentent donc une certaine similitude avec les systèmes de réponse aux questions (question-answering), qui se concentrent sur une seule réponse tout en ignorant le contexte ou les objectifs plus larges de la conversation.

Exemple de corpus utilisés :

- Conversations téléphoniques
- Dialogue de film (disponible en grande quantité dans différents corpus)
- Extraction des réponses possibles de corpus non issus des dialogues, afin qu'un chatbot puisse raconter des histoires.

Idée exploitée: Au lieu d'utiliser des règles construites à la main comme pour Eliza, "fouiller" les conversations entre humains ou entre humains et machine.

→ Systèmes gourmands en quantité de données.

Chatbots basés sur la recherche d'information

Autre idée : Les interventions que les humains font pour répondre au chatbot peuvent être utilisées comme données conversationnelles supplémentaires pour la construction des règles.

- → Le système Xiaolce (Microsoft) collecte et stocke toutes les conversations entre Xiaolce et ses utilisateurs (plus de 30 milliards de paires de conversations).
- Cleverbot (Carpenter 2017 http://www.cleverbot.com)
- Microsoft Tay

Un exemple de conversation de Cleverbot

User:	Do you like Doctor Who?
Cleverbot:	Yes, so funny
User:	What's your favorite show?
Cleverbot:	I don't have one
User:	What's your name?
Cleverbot:	Jason I already told you.
User:	No your real name.
Cleverbot:	Stephanie IS my real name.
User:	I thought it was Jason?
Cleverbot:	How did you come to that conclusion.
User·	Because vou said so.

Deux architectures de chatbot basées sur l'IR

- Renvoi la réponse correspondante à l'intervention la plus similaire
 - Considérer le message de l'utilisateur, q et trouver un message similaire, t, dans le corpus C
 q = "do you like Doctor Who"

t = "do you like Doctor Strangelove"

- Considérer le message t dans le corpus C le plus similaire à q, puis
- 2. Renvoyer la réponse associée à ce message

$$r = response \left(\underset{t \in C}{\operatorname{argmax}} \frac{q^T t}{||q||t||} \right)$$
 Yes, so funny

Chatbots basés sur des réseaux de neurones

- Considérer la géneration de réponses comme une tâche d'apprendre la bonne réponse étant donnée la dernière réplique de l'utilisateur.
- Effectuer l'apprentissage sur :
 - bases de données de dialogues de films
 - Conversations sur Twitter
- Construire un réseau neural profond
 - Données : réplique de l'utilisateur 1 en fonction de la réplique de l'utilisateur 2

Chatbots: pros et contres

- Pros:
 - Amusant
 - Applications aux de conseil
 - Applications pointues pour lesquelles on peut écrire le scénario
- Contre :
 - Ils ne comprennent pas vraimenet
 - Les chatbots basés sur des règles sont coûteux et fragiles
 - Les chatbots basés sur l'IR ne peuvent que refléter les données qui sont dans le corpus utilisé
- L'avenir : combiner les chatbots avec des agents basés sur des cadres

Agents de dialogue basés sur des cadres (tâches/buts)

- But: Aider l'utilisateur à résoudre certaines tâches comme la réservation d'un avion ou l'achat d'un produit.
- Architecture très simple du GUS (Genial Understander System) pour le dialogue basé sur les tâches.
- Introduite pour la première fois pour la planification des voyages (Bobrow et al., 1977)).
- Inspiration pour la plupart sinon la totalité des assistants numériques commerciaux modernes.

Agents de dialogue basés sur des cadres

- Parfois appelés "agents de dialogue basés sur les tâches".
- Tous les systèmes de dialogue modernes basés sur des tâches sont basés sur des cadres ("frames").
- Cadre: Structure de connaissances représentant les intentions de l'utilisateur extraites par le système via les réponses de l'utilisateur.

Agents de dialogue basés sur des cadres

- Cadre: Composé par une collection de "slots", chacun pouvant prendre un ensemble de valeurs possibles.
- L'ensemble de ces cadres est parfois appelé ontologie de domaine : Structure de connaissances représentant les intentions des utilisateurs.

Agents de dialogue basés sur des cadres

- L'ensemble des "slots" dans un cadre spécifie ce que le système a besoin de savoir et leur remplissage est limité aux valeurs ayant un type sémantique particulier.
- Dans le domaine des voyages, par exemple, un "slot" peut être de type ville (donc prendre des valeurs comme "Nice" ou "Paris").
- Un ou plusieurs cadres
 - Chacun avec une collection de "slots"
 - Chaque "slot" ayant sa propre valeur

Le cadre (frame)

- Un ensemble de "slots" à remplir avec des informations d'un type donné.
- Chacun d'eux est associé à une question à l'utilisateur

Slot	Type	Question
ORIGIN	city	What city are you leaving from?
DEST	city	Where are you going?
DEP DATE	date	What day would you like to leave?
DEP TIME	time	What time would you like to leave?
AIRLINE	line	What is your preferred airline?
ARR TIME	time	When do you want to arrive?
ARR DATE	date	What day would you like to arrive?

Hello. My name is GUS. I can help you plan a simple trip by air. Where do you GUS: want to go? The state of the I want to go to San Diego on May 28 Client: art in 1977!!!! GUS: What time do you want to leave? I must be in San Diego before 10 am Client: GUS: Would you like P.S.A. flight 102 that arrives at 9.15 am? Client: What is the next flight? Air California flight 310 that leaves at 8.30am and arrives at 10.00am GUS: Client: I'll take the first one GUS: In what name should I make a reservation, please? Client: Dan Bobrow GUS: I have confirmed the following flight: P.S.A. flight 102 on Wednesday May 28 from San Jose to San Diego that leaves at 7.30 am and arrives at 9.15 am. What date do you want to return on? GUS: Client: On Friday in the evening. GUS: Would you like the flight that leaves at 7.45 pm? Client: That's fine. GUS: I have confirmed the following flight: P.S.A. flight 307 on Friday May 30 from San Diego to San Jose that leaves at 7.45 pm and arrives at 9.30 pm Thank you for calling. Goodbye

Les types des "slots" peuvent être complexes

- Tous les agents de dialogue basés sur des cadres ont une structure hiérarchique .
- Exemple : Le type DATE est lui-même un cadre avec des "slots" dont les types sont des entiers ou des ensembles avec les noms des jours de la semaine.

```
DATE
MONTH NAME
DAY (BOUNDED-INTEGER 1 31)
YEAR INTEGER
WEEKDAY (MEMBER (SUNDAY MONDAY TUESDAY WEDNESDAY THURSDAY FRIDAY SATURDAY)]
```

- Le système pose des questions à l'utilisateur (en utilisant des modèles de questions prédéfinis associés à chaque "slot" de chaque cadre)
- Remplit tout "slot" spécifié par l'utilisateur
- Si réponse concerne plusieurs "slots", ex.
 "J'aimerais un vol aller simple de Paris à Nice, avec départ après 17 heures le mardi."
- Remplit tous les créneaux pertinents, puis continue à poser des questions pour remplir les créneaux restants, en ignorant les questions associées aux créneaux remplis.

- Règles de conditionnalité liées "slots".
- Ex. Règle liée au "slot" DESTINATION.
 - Une fois que l'utilisateur a spécifié sa destination, on peut automatiquement entrer cette ville comme LIEU DE SÉJOUR par défaut pour le cadre de réservation d'hôtel.
 - Si l'utilisateur spécifie le JOUR DE DESTINATION pour un court voyage, le système peut automatiquement saisir le JOUR D'ARRIVÉE.

- Nécessité de plusieurs cadres.
- Ex. Cadres pour les réservations de voiture ou d'hôtel + Cadres avec informations sur l'itinéraire ("Quelles compagnies aériennes assurent la liaison entre Boston et San Francisco ?").
- Informations sur les pratiques sur les tarifs ("Doisje rester un nombre de jours précis pour obtenir un bon tarif aérien ?")
- Besoin de désambiguïser le bon "slot" par rapport à quel cadre remplir, avant de lui passer le contrôl.

- Changement dynamique de contrôle.
- → Système de règles de production.
- Quantité d'information suffisante → réalisation des nécessaires (interrogation de BD vols).
- Renvoie du résultat à l'utilisateur.

Envisagez un système de voyage aérien très simple :

- Demandez à l'utilisateur la ville de départ
- Demander une ville de destination
- Demander l'heure
- Demandez si le voyage est aller-retour ou non

Gestionaire de dialogue réalisé par un automate

