Sophie Rosset

LIMSI, CNRS

Dialogue humain-machine

une introduction

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
Campus Universitaire Bât. 508 – F-91405 Orsay cedex
www.limsi.fr

Plan

- Objectifs du cours
- Systèmes de dialogue
- Système orienté tâche
- Système purement conversationnel
- Évaluation : métriques et méthodologies

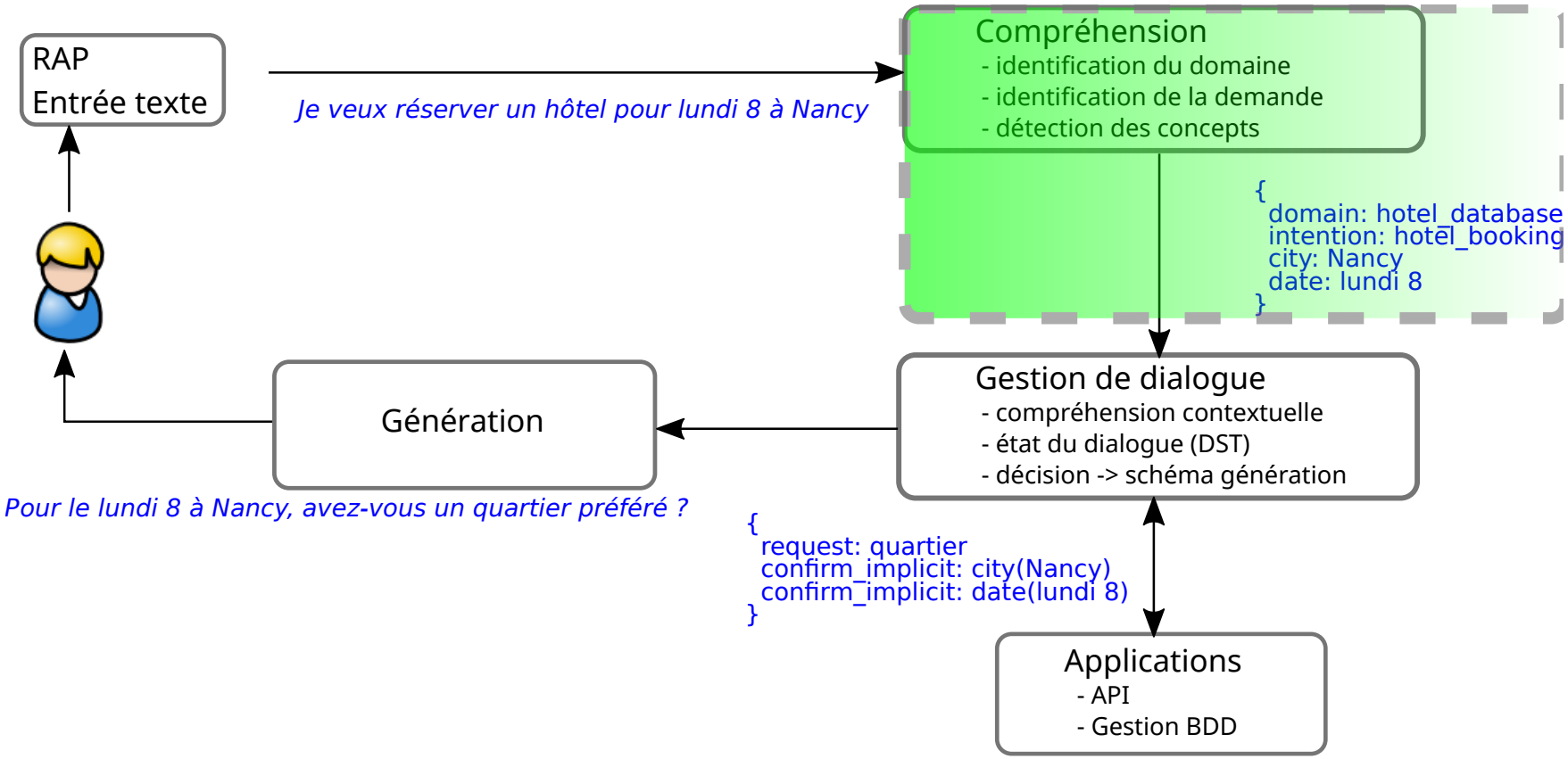
Objectifs du cours (3 séances)

- Brosser un rapide portrait du domaine
  - Définition : qu'entend-ton par système de dialogue, chatbot, etc. ?
  - Historique : quelles évolutions ?
  - Méthodes : quelles sont les principales méthodes ? Pourquoi ?
- Présenter les principales méthodologies de l'évaluation
- Ce que n'est pas ce cours
  - une présentation de chacune des approches !
- **Aujourd'hui :**
  - Rappel/discussion de la séance précédente
  - Présentation des différents « modules »

Plan

- Objectifs du cours
- Systèmes de dialogue
- **Système orienté tâche**
- Système purement conversationnel ou de type *chatbot*
- Évaluation : métriques et méthodologies

Système orienté tâche



Système orienté tâche : compréhension

La compréhension est souvent considérée comme un enchaînement de 2-3 tâches :

- détection/identification :
  - du domaine : si application multi-domaine
  - de l'intention (aka type de demande) : toujours (réserver un hôtel)
- détection des concepts

Découpage apparut dans les années 2000 (lié à l'essor des approches statistiques)
But du découpage : décomposer un problème complexe en plusieurs plus simples
Les approches non statistiques (typiquement avec des règles) tendent à tout réaliser en même temps.

Système orienté tâche : compréhension

Identification domaine/intention

- Apprentissage supervisé : classification
- Formulation du problème :

  un ensemble d'énoncés u_i associés à un label c_i
  D = $(u_1,c_1),...(u_n,c_n)$

  $\rightarrow$ entraîner un modèle pour estimer le label correspondant à un nouvel énoncé u_k

Méthodes : classification

- SVM, MaxEnt, etc.
- Modèles neuronaux variés comme les CNN, DBN (*Deep Belief Networks*), les RNN et LSTM etc. (voir Tuto Chen et al.)

Système orienté tâche : compréhension

Détection de concepts (*slot filling*)

concepts = les entités mentionnées dans l'énoncé (mentions)
slots = les attributs de la tâche et du domaine
slot filling = associer aux slots les mentions (normalisées)

Words	une	chambre	pour	deux	adultes
Slots	B-nb-room	I-nb-room	O	B-nb-pers	I-nb-pers

Exemple tiré de [Bonneau-Maynard et al., 2006]

Méthodes : annotation en séquence

- CRF : [Hahn et al., 2010]
- LSTM : [Yao et al., 2014]
- RNN et RNN-CRF : [Mesnil et al., 2015]
- ... en pleine effervescence !

Système orienté tâche : compréhension

Modèle de compréhension joint

Peut-on faire les deux étapes en une seule fois ? OUI

- Slot filling puis classification : [Guo et al., 2014]
- Représentation commune (GRU) puis parallélisation : [Zhang and Wang, 2016]
- Représentation commune et parallélisation, bi-LSTM + CNN : [Neuraz et al., 2018]

Compréhension : analyse fondées sur des grammaires

- Formalisme logique et λ-calcul [Villaneau and Antoine, 2004]
- Expressions régulières [Galibert, 2009] utilisés dans [Campillos Llanos et al., 2016]
- Context Free Grammar [Glass et al., 1995]

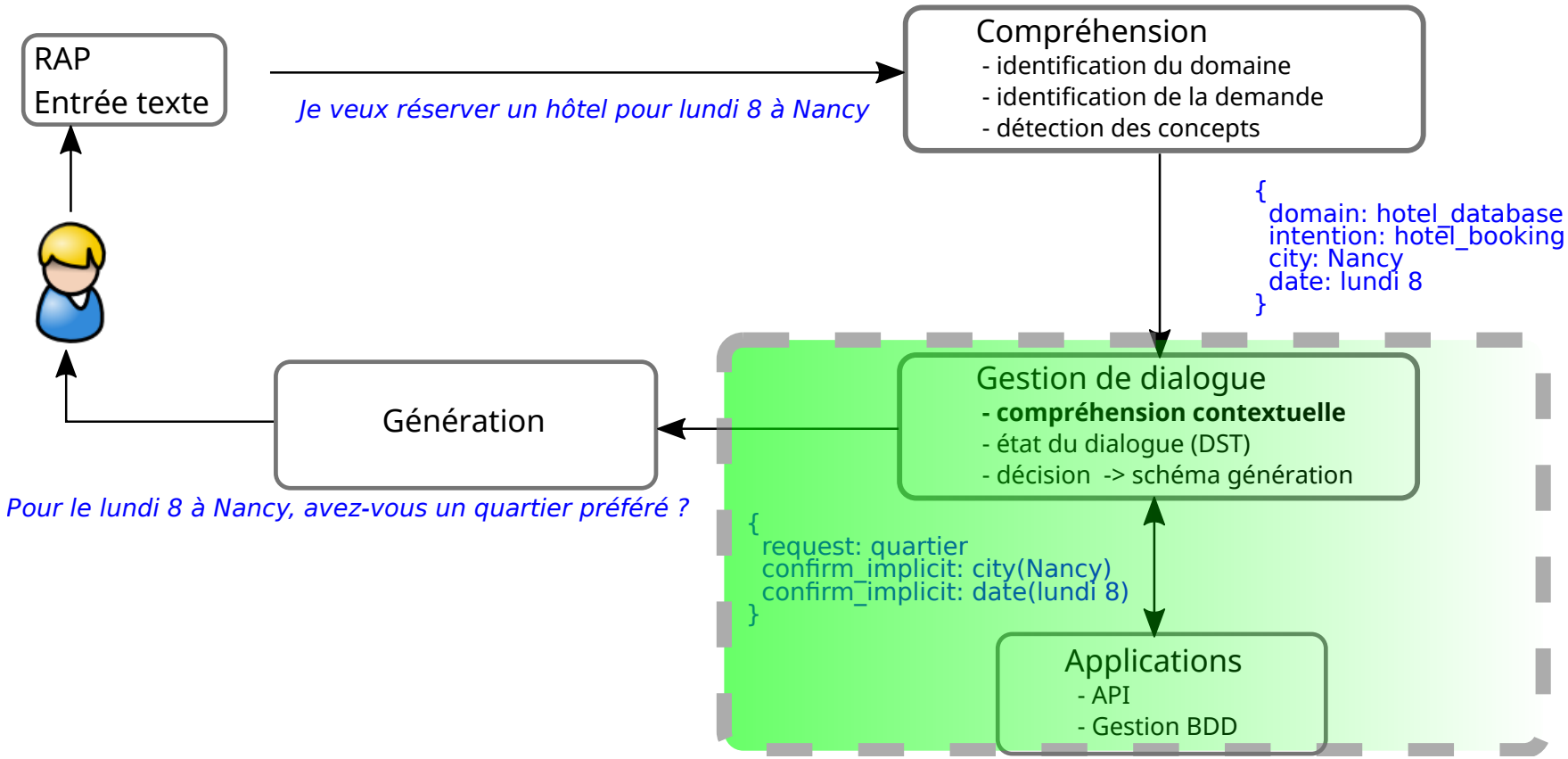
Petite précision

Tout n'est pas si simple

- Il existe un benchmark international, ancien, en anglais : ATIS
  - 2 tâches : SF + Intent
  - corpus considéré comme facile [Tür et al., 2010, Béchet and Raymond, 2018]
  - résultats meilleurs avec DNN, $\approx$ 95%
- il existe un benchmark en français, un peu moins ancien, : MEDIA
  - 1 tâche : SF
  - corpus plus difficile
  - résultats autour de 10-12% CER
  - DNN difficilement meilleure que les CRF (et nécessite beaucoup de « feature engineering » (voir Thèse Edwin Simonnet)

$\rightarrow$ toujours faire attention aux affirmations sur un corpus, une langue... ce n'est pas toujours le cas général (voire jamais) !

Système orienté tâche



Système orienté tâche : gestion dialogue

Compréhension contextuelle

- La langue naturelle, hors contexte, est ambiguë
  U : je veux réserver un hôtel pour lundi 8 à Nancy
  S : pour combien de nuits ?
  U : deux
  $\rightarrow$ amount(2) = nb-night(2)
- Certaines demandes ne se comprennent qu'avec le contexte
  U : je voudrais partir vers 17 heures
  S : vous avez un train à 16 heures 57 ...
  U : et le suivant
  S : le train suivant part à 17 heures 30 ...

Système orienté tâche : gestion dialogue

Compréhension contextuelle : liaison d'entités

Parfois on peut avoir besoin de lier une entité (le contenu du slot) à un identifiant d'une base ou encore faire le lien avec une autre expression de la base

avez-vous une maladie cardiaque ?

→ *hypertension* présent dans la base (ici un dossier patient) et doit matcher sur maladie cardiaque

Système orienté tâche : gestion dialogue

Compréhension contextuelle : méthodes

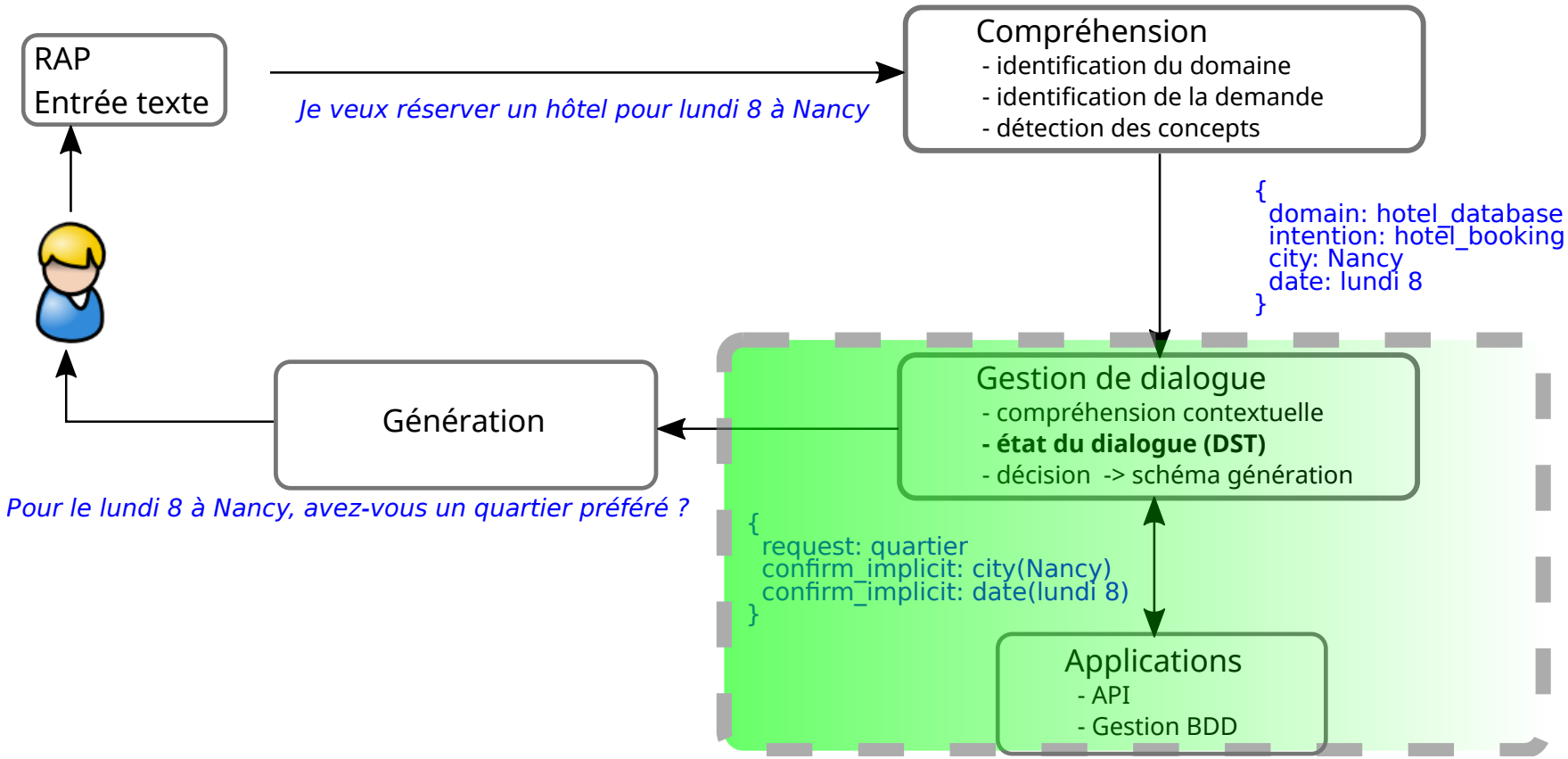
- Le plus souvent, interne à la gestion de dialogue (mise à jour des états, ie des vecteurs qui représentent ce dont on parle + module supplémentaire pour la liaison)
  - parfois simple vérification fondée sur une distance de Levenshtein par exemple)
  - parfois utilisation de ressources et ontologies spécialisées (médical)
  - parfois comparaison de représentation dense de mots (embeddings, voir cours avec Sahar Ghannay)

Système orienté tâche : gestion dialogue

Compréhension contextuelle : méthodes

- Dépend fortement d'une définition du domaine
- Modèle à base de connaissances (frame-based, information state update) [Traum and Larsson, 2003, Campillos Llanos et al., 2016]
- Modèles neuronaux
  - LSTM qui encodent un ou plusieurs énoncés précédents [Hori et al., 2015]
  - end2end memory network [Chen et al., 2016]

Système orienté tâche



Système orienté tâche : gestion du dialogue

Etat du dialogue

- Objectif : fournir une représentation complète de ce que veut l'utilisateur à n'importe quel moment du dialogue [Henderson, 2015]
- état du dialogue représenté par ensemble de slots : paire(attribut,valeur) où des valeurs peuvent être des paires également etc.
- état du dialogue peut également être représenté par un vecteur qui va contenir ces mêmes informations
- état du dialogue bouge avec le temps...

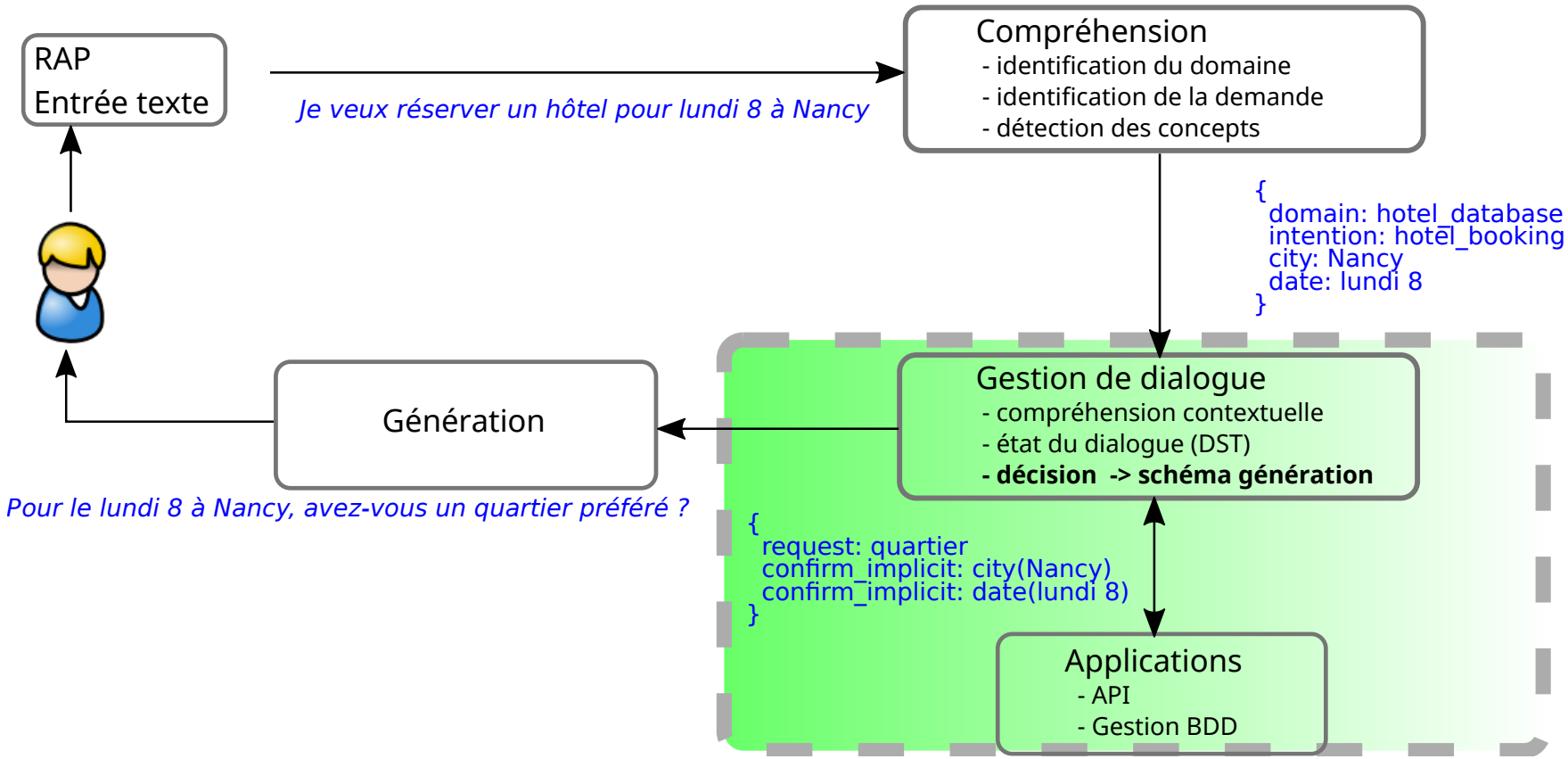
La prise de décision s'appuie sur l'état du dialogue courant et le met à jour...
Benchmark international : DSTC `https://www.microsoft.com/en-us/research/event/dialog-state-tracking-challenge/`

Système orienté tâche : gestion du dialogue

Dialogue state tracking (DST) : méthodes

- Méthodes génératives (et rule based)
  - Rule-based [Larsson and Traum, 2000] mais ne permet pas de prendre en compte des hypothèses multiples de compréhension
  - Réseaux bayésiens dynamiques (DBN) : énumérer tous les possibles peut être explosif, donc soit en maintenant un beam search [Young et al., 2007] soit en supposant une indépendance conditionnelle entre les composants de l'état de dialogue [Thomson and Young, 2010]
  - Inconvénient [Metallinou et al., 2013] : doivent tout modéliser y compris ce qui n'est pas vu dans l'entraînement
- Méthodes discriminantes : modéliser la tâche comme une tâche de classification $\rightarrow (P(s_t|o_0, ..., o_t)$
  - Classifieurs linéaires [Metallinou et al., 2013]
  - Réseaux de neurones [Henderson et al., 2013]

Système orienté tâche



Système orienté tâche : gestion du dialogue

Contrôleur (module de décision) : objectifs

Le contrôleur de dialogue gère la prise de décision. Il doit décider quand et quoi dire à l'utilisateur, quand et quoi rechercher comme information.
Il s'appuie pour cela sur l'état du dialogue.

Différentes approches

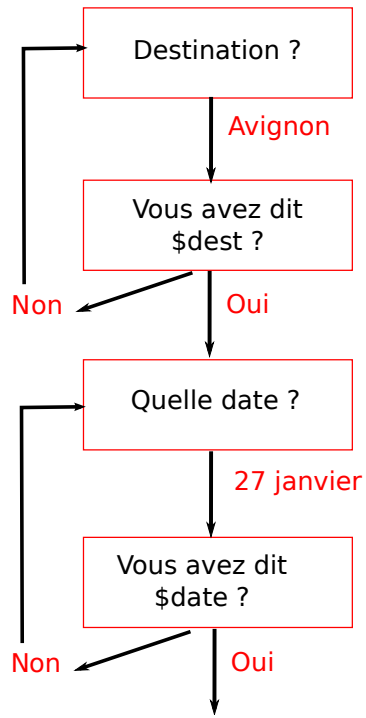
- Les graphes $\rightarrow$ automates à états finis
- Les schémas (frame)
- Les approches statistiques
- Les approches neuronales

Ces approches ne s'appuient pas nécessairement sur un état de dialogue en tant que tel, explicité

Système orienté tâche : gestion du dialogue

Contrôleur : automate à états finis

- Les noeuds représentent les questions du système
- Les arcs représentent les réponses
- Le graphe représente toutes les alternatives possibles (légales)



Système orienté tâche : gestion du dialogue

Contrôleur : automate à états finis

- Les noeuds représentent les questions du système
- Les arcs représentent les réponses
- Le graphe représente toutes les alternatives possibles (légales)

Quelques remarques :

- La gestion du dialogue est très simple tout comme les échanges possibles
- Ne peut être utilisé que dans des tâches très simples et très structurées (slots limités en nombre et valeurs)
- Toujours utile et utilisé
- ne gère aucun état complexe !

Système orienté tâche : gestion du dialogue

Contrôleur (module de décision) : les frame (ou schémas)

- Un schéma est un ensemble de slots ($\rightarrow$ dialogue plus souple car ordre éléments non contraint)
- $\rightarrow$ compréhension plus complexe possible (domaine plus riche)
- Possibilité d'avoir autant de schéma que de (sous-)tâches ou un schéma complexe

Un schéma est une manière plus flexible pour contrôler le dialogue.

- Il représente ce que doit résoudre le système
- Il s'agit d'un ensemble de slots que le système doit remplir au fur et à mesure

Système orienté tâche : gestion du dialogue

Contrôleur (module de décision) : les frame (ou schémas)

Exemple de représentation

Actions	Interprétation
S : Quelle est votre destination ? U : Avignon le 27 janvier	S cherche une destination *Avignon* est une ville et une destination *27 janvier* une date
schéma avant	schéma après
destination : nil départ : nil date : nil	destination : Avignon départ : nil date : 27 janvier

Système orienté tâche : gestion du dialogue

Contrôleur (module de décision) : apprentissage par renforcement

- Proposé par [Young, 2006]
- Gestion du dialogue = prendre une décision
- Apprendre à prendre une décision en fonction d'un état
- Modélisation des états :
  - MDP : Markov Decision Process
  - POMDP : Partially Observable Markov Decision Process
- Apprentissage de la stratégie de dialogue (policy, décision) : Apprentissage par renforcement [Sutton and Barto, 1998]

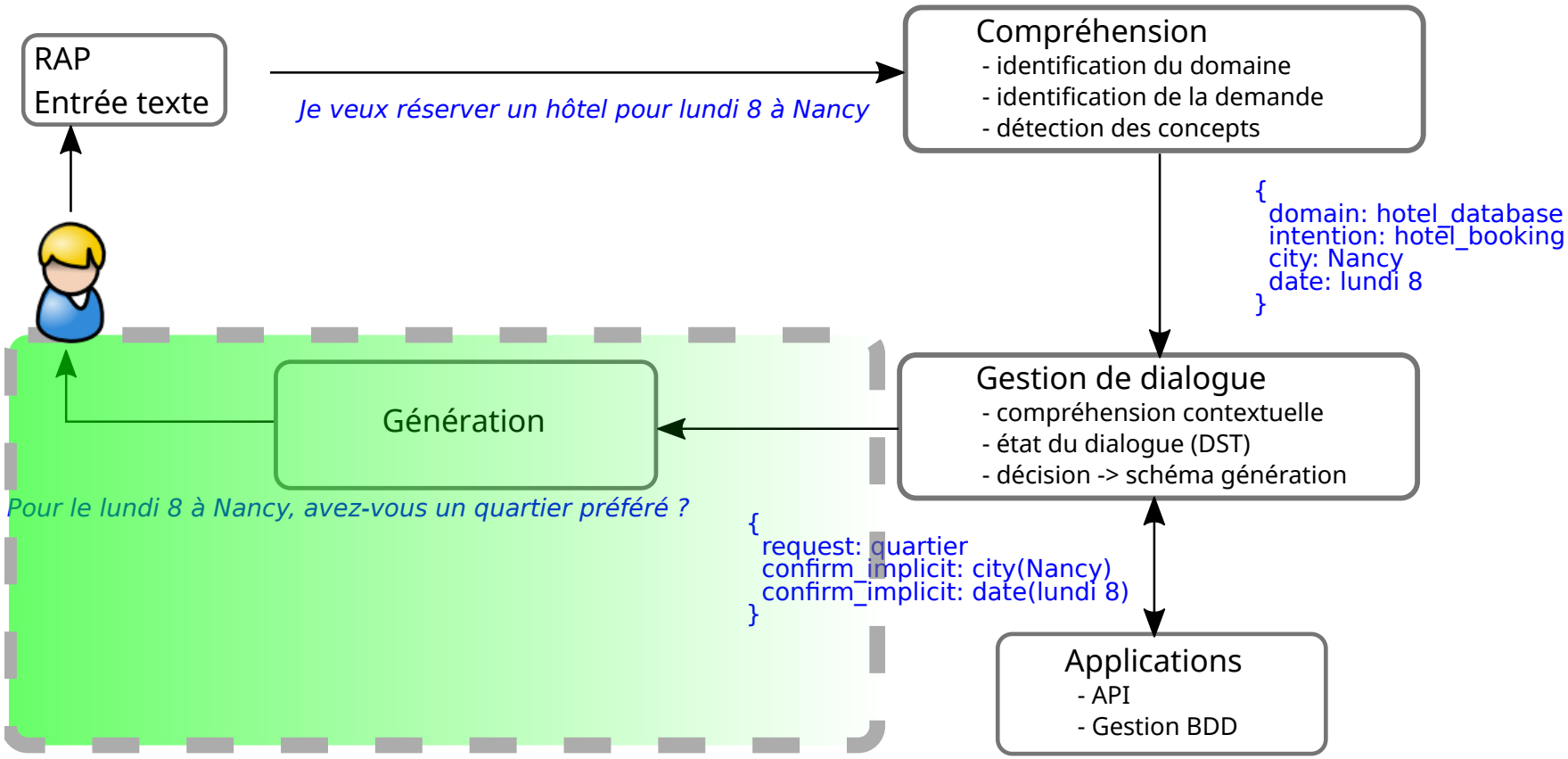
Apprentissage nécessite beaucoup de données et beaucoup d'essais.
$\rightarrow$ simulation d'utilisateur [Schatzmann et al., 2006]

Système orienté tâche : gestion du dialogue

Approches neuronales

- Elles sont récentes [Wen et al., 2017]
- Objectif : apprendre à mapper les schémas du dialogue (les slots) et un historique à une réponse du système de dialogue.
- Des modèles de type encoder-decoder sont utilisés pour l'apprentissage du système.
- Une approche hybride a été récemment proposé permettant dans une architecture neuronale d'intégrer des programmes (ie des règles) [Williams et al., 2017]

Système orienté tâche



Système orienté tâche : Génération

Génération : objectif

Transformer un schéma sémantique en une phrase en langue naturelle

request(quartier)

confirm(city), confirm(date)

$\rightarrow$ S : Pour le lundi 8 à Nancy, avez-vous un quartier préféré ?

- Décider *quoi* dire
- Décider *comment* le dire

Génération : méthodes

- Génération fondée sur des patrons (*template based*)
  - ensemble de paires(phrases à trous,schémas)
- Génération fondée sur des syntagmes (*phrase based*)
  - à partir de là des modèles statistiques (cf. liste des publications sur le site)

Système orienté tâche : Génération

Génération : objectif

Transformer un schéma sémantique en une phrase en langue naturelle

request(quartier)

confirm(city), confirm(date)

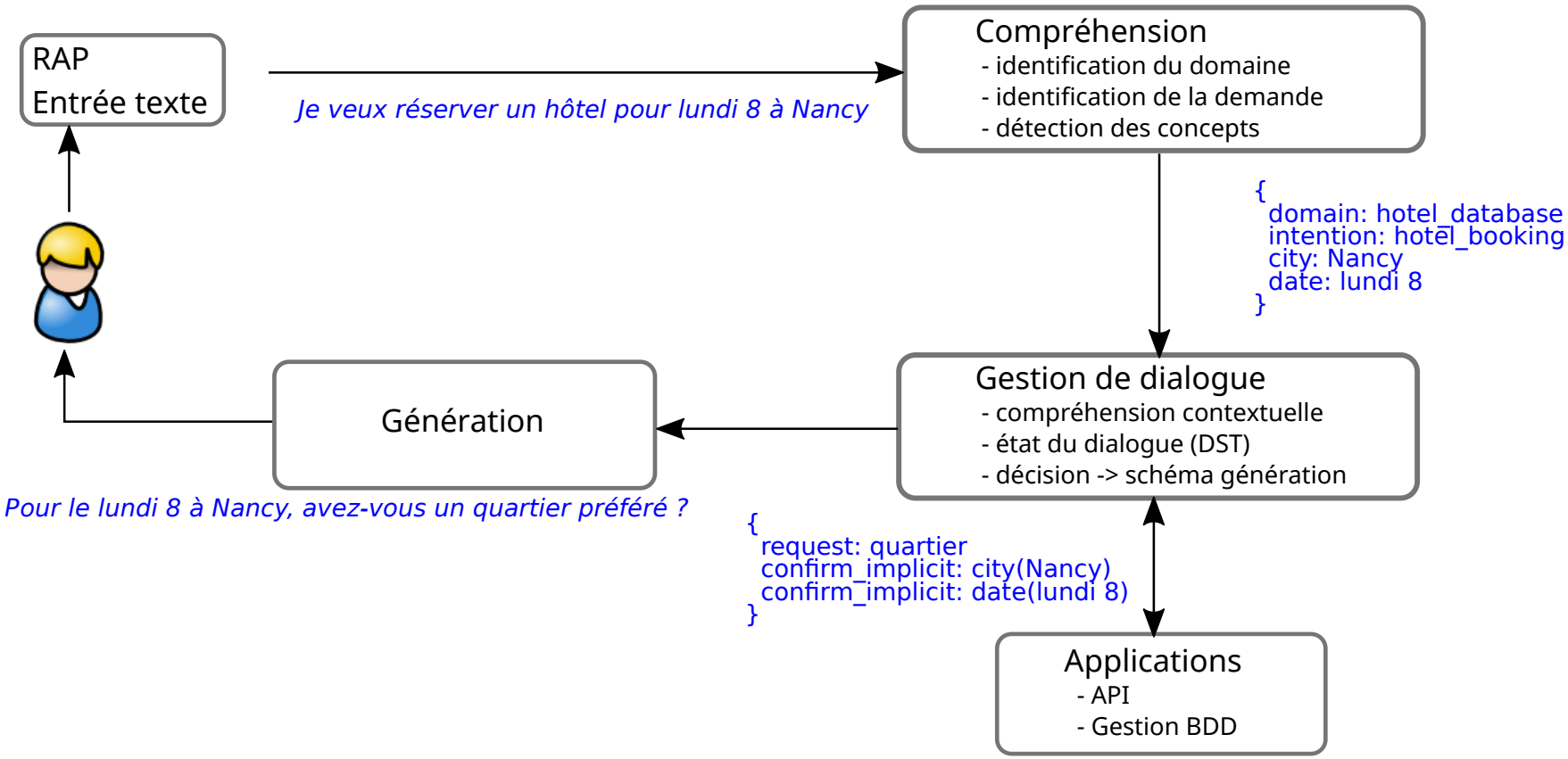
$\rightarrow$ S : Pour le lundi 8 à Nancy, avez-vous un quartier préféré ?

- Décider *quoi* dire
- Décider *comment* le dire

Génération : méthodes

- un challenge : E2E NLG challenge
  `http://www.macs.hw.ac.uk/InteractionLab/E2E/`
- Essor des méthodes à bases de RNN [Dušek et al., 2018] et sur le site du challenge

Système orienté tâche



Système orienté tâche : bilan

Système complexe

Un système de dialogue orienté tâche est un système complexe qui implique plusieurs composants, le plus souvent « pipelinés »

- (Reconnaissance de la parole)
- Compréhension de la langue
- Gestionnaire de dialogue
- Génération en langue
- (synthèse de la parole)

Approches

Les approches sont variées allant du combo règles et connaissances explicites aux approches statistiques et neuronales, et ce quel que soit le module concerné

→ Quelles sont les forces et faiblesses de chacune de ces approches ?

Système orienté tâche : forces et faiblesses...

Règles/connaissances explicites

- Pour
  - facile à interpréter/débuguer
  - nécessite peu de données
  - évolution raisonnablement simple jusqu'à un certain point

  $\rightarrow$ toujours en usage dans les systèmes commerciaux
- Contre
  - difficile à maintenir
  - passage à un nouveau domaine parfois difficile (dépend de l'implémentation)
  - passage à une nouvelle langue parfois difficile (dépend de l'implémentation)
- A voir
  - repose sur une expertise (le développeur doit être expert ? ? ?)

Système orienté tâche : forces et faiblesses...

Statistiques (pomdp, rl)

- Pour
  - données utilisées pour développer (plus proche d'une réalité)
  - pas nécessaire d'écrire/coder des comportements de façon explicite
- Contre
  - passage à nouveau domaine et nouvelles langues difficiles
  - difficile à interpréter/débuguer
  - évolution peu simple (repose sur RL, repose sur utilisateur simulé -> appris sur données disponibles)
- A voir
  - repose sur des données annotées coûteuses à obtenir (expert est présent ici)
  - impossible d'apprendre un modèle qui ferait tout (end-2-end)

Système orienté tâche : forces et faiblesses...

Apprentissage profond (approches neuronales)

- Pour
  - données utilisées pour développer (plus proche d'une réalité)
  - pas nécessaire d'écrire/coder des comportements de façon explicite
  - passage à l'échelle plus simple
  - puissance des représentations
  - end-2-end devient envisageable
- Contre
  - passage à nouveau domaine et nouvelles langues difficiles
  - difficile à interpréter/débuguer
- A voir
  - Comme pour les autres approches, repose sur de très grandes quantités de données, au moins alignement input/output ... comment les obtenir ? peut-on généraliser ? quoi ? comment ?

Système orienté tâche : forces et faiblesses...

Questions à se poser

- données disponibles ? quelles formes ?
- coûts obtention des données *vs* coûts développement
- richesse et complexité du domaine

Quelques pistes

- Génération de données ? voir [Neuraz et al., 2018]
- Autres idées ?

Plan

- Objectifs du cours
- Systèmes de dialogue
- Système orienté tâche
- **Système purement conversationnel** ou de type *chatbot*
- Évaluation : métriques et méthodologies

Système purement conversationnel

Objectifs

- modéliser des capacités conversationnelles tout venant
- fondamentalement : générer la réaction la plus appropriée étant donné un contexte et un énoncé utilisateur

Méthodes

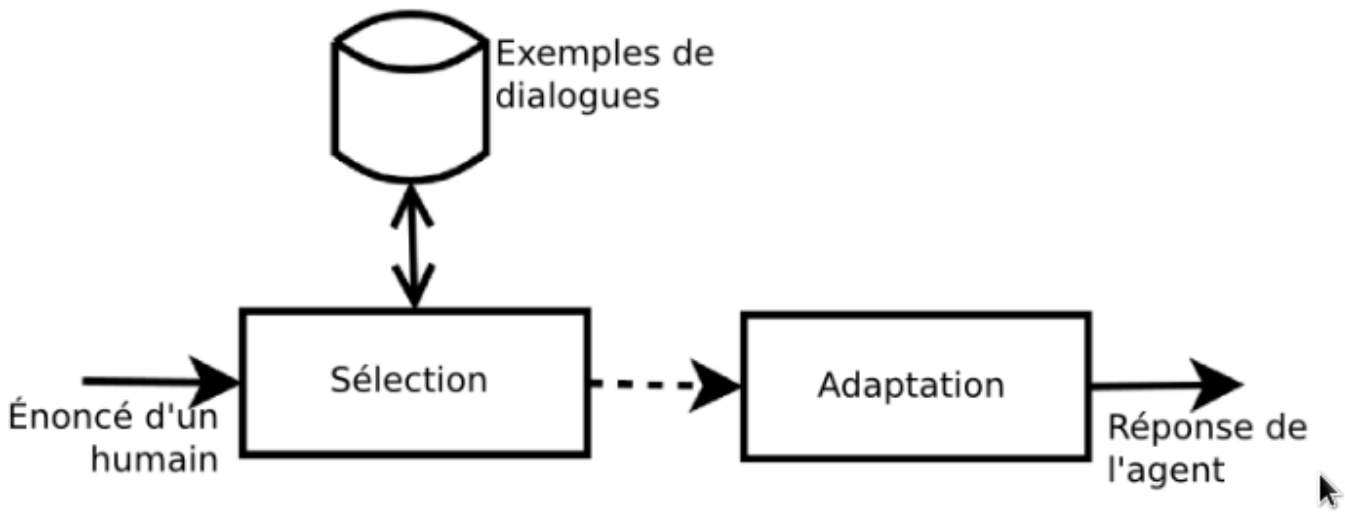
- Hypothèse : dans de grand corpus de conversations, on peut trouver la manière...
- Deux familles d'approches
  - approches de type « recherche d'information »
  - approches génératives

Système purement conversationnel

Méthodes de type RI

Essentiellement il s'agit de trouver une réponse par similarité sémantique

- méthode RI classiques : TF-IDF, BM25, etc.
- calcul de similarité par embeddings de phrase ou énoncé



Plan

- Objectifs du cours
- Systèmes de dialogue
- Système orienté tâche
- Système purement conversationnel ou de type *chatbot*
- **Évaluation : métriques et méthodologies**

Évaluation : métriques et méthodologies

Évaluation « objectives »

- Évaluer automatiquement ce que fait le système
  - compréhension
  - gestion du dialogue
  - génération
- Compare le résultat du système (hypothèse) à ce qui est attendu (référence)
  - Précision, Rappel, F-mesure
  - Concept Error Rate
  - Slot Error Rate
  - score BLEU et ses dérivés (génération de surface)

→ voir eidi-mesures.pdf

Évaluation : métriques et méthodologies

Évaluation « subjectives »

- Évaluer la perception que l'utilisateur a
  - des capacités du système (*est-ce que le système comprend ? répond bien ? ...*)
  - de l'interaction qu'il a avec le système (*est-ce qu'il est agréable ? poli ? ...*)
- Questionnaires, échelles de Likert...

Évaluation : métriques et méthodologies

Questions

- Existe-t-il une corrélation entre la satisfaction utilisateur et les scores objectifs ?
- Peut-on prédire à partir d'indicateurs la satisfaction utilisateur ?

Obtention des données

- En faisant appel à des volontaires
  - annotation des données
  - + possibilité évaluation « subjectives »
- Par simulation
  - pas d'évaluation « subjective »
- En utilisant des corpus existants [Bonneau-Maynard et al., 2006, Henderson, 2015, Williams et al., 2017]
  - peu de domaines et de tâches disponibles
  - inadapté à de nouveaux domaines, langues, tâches

Références

Béchet, F. and Raymond, C. (2018).

Is ATIS too shallow to go deeper for benchmarking Spoken Language Understanding models ?
In *InterSpeech 2018*, pages 1–5, Hyderabad, India.

Bonneau-Maynard, H., Ayache, C., Bechet, F., Denis, A., Kuhn, A., Lefevre, F., Mostefa, D., Quignard, M., Rosset, S., Servan, C., and Villaneau, J. (2006).

Results of the French Evalda-Media evaluation campaign for literal understanding.
In *lrec*, pages 2054–2059, Genoa.

Campillos Llanos, L., Bouamor, D., Zweigenbaum, P., and Rosset, S. (2016).

Managing linguistic and terminological variation in a medical dialogue system.
In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France.

Chen, Y.-N., Hakkani-Tür, D., Tür, G., Gao, J., and Deng, L. (2016).

End-to-end memory networks with knowledge carryover for multi-turn spoken language understanding.
In *INTERSPEECH*, pages 3245–3249.

Dušek, O., Novikova, J., and Rieser, V. (2018).

Findings of the E2E NLG Challenge.
In *Proceedings of the 11th International Conference on Natural Language Generation*, Tilburg, The Netherlands.
arXiv :1810.01170.

Références

Galibert, O. (2009).

Approaches and methodologies for automatic Question-Answering in an open-domain, interactive setup.
Phd dissertation, Université Paris Sud - Paris XI.

Glass, J., Flammia, G., Goodine, D., Phillips, M., Polifroni, J., Sakai, S., Seneff, S., and Zue, V. (1995).
Multilingual spoken-language understanding in the mit voyager system.
Speech communication, 17(1-2) :1–18.

Guo, D., Tur, G., Yih, W.-t., and Zweig, G. (2014).

Joint semantic utterance classification and slot filling with recursive neural networks.
In *Spoken Language Technology Workshop (SLT), 2014 IEEE*, pages 554–559. IEEE.

Hahn, S., Dinarelli, M., Raymond, C., Lefèvre, F., Lehen, P., De Mori, R., Moschitti, A., Ney, H., and Riccardi, G. (2010).
Comparing stochastic approaches to spoken language understanding in multiple languages.
IEEE Transactions on Audio, Speech and Language Processing (TASLP), 16 :1569–1583.

Henderson, M. (2015).

Machine learning for dialog state tracking : A review.
In *Machine Learning in Spoken Language Processing Workshop*.

Références

Henderson, M., Thomson, B., and Young, S. (2013).
Deep neural network approach for the dialog state tracking challenge.
In *Proceedings of the SIGDIAL 2013 Conference*, pages 467–471.

Hori, C., Hori, T., Watanabe, S., and Hershey, J. R. (2015).
Context sensitive spoken language understanding using role dependent lstm layers.
In *Machine Learning for SLU Interaction NIPS 2015 Workshop*.

Larsson, S. and Traum, D. R. (2000).
Information state and dialogue management in the trindi dialogue move engine toolkit.
Natural language engineering, 6(3-4) :323–340.

Mesnil, G., Dauphin, Y., Yao, K., Bengio, Y., Deng, L., Hakkani-Tur, D., He, X., Heck, L., Tur, G., Yu, D., and Zweig, G. (2015).
Using recurrent neural networks for slot filling in spoken language understanding.
IEEE/ACM Trans. Audio, Speech and Lang. Proc., 23(3) :530–539.

Metallinou, A., Bohus, D., and Williams, J. (2013).
Discriminative state tracking for spoken dialog systems.
In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, pages 466–475, Sofia, Bulgaria. Association for Computational Linguistics.

Références

Neuraz, A., Campillos Llanos, L., Burgun, A., and Rosset, S. (2018).

Natural language understanding for task oriented dialog in the biomedical domain in a low ressources context, nips workshop.

In *Machine Learning for Health (ML4H) : Moving beyond supervised learning in healthcare*, Montréal, Québec, Canada.

Schatzmann, J., Weilhammer, K., Stuttle, M., and Young, S. (2006).

A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies.

The knowledge engineering review, 21(2) :97–126.

Sutton, R. S. and Barto, A. G. (1998).

Reinforcement learning : An introduction, volume 1.

MIT press Cambridge.

Thomson, B. and Young, S. (2010).

Bayesian update of dialogue state : A pomdp framework for spoken dialogue systems.

Computer Speech & Language, 24(4) :562–588.

Traum, D. R. and Larsson, S. (2003).

The information state approach to dialogue management.

In *Current and new directions in discourse and dialogue*, pages 325–353. Springer.

Références

Tür, G., Hakkani-Tür, D. Z., and Heck, L. P. (2010).
What is left to be understood in atis ?
pages 19–24.

Villaneau, J. and Antoine, J.-Y. (2004).
Categorials grammars used to partial parsing of spoken language.
In *CG2004*.

Wen, T.-H., Vandyke, D., Mrkšić, N., Gasic, M., Rojas Barahona, L. M., Su, P.-H., Ultes, S., and Young, S. (2017).
A network-based end-to-end trainable task-oriented dialogue system.
In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics : Volume 1, Long Papers*, pages 438–449, Valencia, Spain. Association for Computational Linguistics.

Williams, J. D., Asadi, K., and Zweig, G. (2017).
Hybrid code networks : practical and efficient end-to-end dialog control with supervised and reinforcement learning.
In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, pages 665–677, Vancouver, Canada. Association for Computational Linguistics.

Yao, K., Peng, B., Zhang, Y., Yu, D., Zweig, G., and Shi, Y. (2014).
Spoken language understanding using long short-term memory neural networks.
In *Spoken Language Technology Workshop (SLT), 2014 IEEE*, pages 189–194. IEEE.

Références

Young, S. (2006).

Using pomdps for dialog management.

In *Spoken Language Technology Workshop, 2006. IEEE*, pages 8–13. IEEE.

Young, S., Schatzmann, J., Weilhammer, K., and Ye, H. (2007).

The hidden information state approach to dialog management.

In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV–149. IEEE.

Zhang, X. and Wang, H. (2016).

A joint model of intent determination and slot filling for spoken language understanding.

In *IJCAI*, pages 2993–2999.