



# Turn-taking enhancement in spoken dialogue systems with reinforcement learning

Hatim Khouzaimi

## ► To cite this version:

Hatim Khouzaimi. Turn-taking enhancement in spoken dialogue systems with reinforcement learning. Artificial Intelligence [cs.AI]. Université d'Avignon, 2016. English. NNT : 2016AVIG0213 . tel-01498847v2

**HAL Id: tel-01498847**

**<https://theses.hal.science/tel-01498847v2>**

Submitted on 25 May 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ORANGE LABS

ACADÉMIE D'AIX-MARSEILLE  
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE



## PhD THESIS

This dissertation is submitted for the degree of *Doctor of Philosophy*

**SPECIALTY: Computer Science**

Doctoral School 380 «Sciences et Agronomie»

Laboratoire d'Informatique (EA 931)

### *Turn-taking Enhancement in Spoken Dialogue Systems with Reinforcement Learning*

by

**Hatim KHOUZAIMI**

**Publicly defended on June 6<sup>th</sup> 2016 in front of a jury composed of:**

Dr	Sophie ROSSET	LIMSI, Orsay	<i>Rapporteur</i>
Pr	David SCHLANGEN	Bielefeld University	<i>Rapporteur</i>
McF	Bassam JABAIAH	CERI-LIA, Avignon University	<i>Examiner</i>
McF	Filip JURCICEK	UFAL, Charles University, Prague	<i>Examiner</i>
Pr	Olivier PIETQUIN	Google DeepMind	<i>Examiner</i>
Dr	Romain LAROCHE	NADIA Team, Orange Labs	<i>Supervisor</i>
Pr	Fabrice LEFEVRE	CERI-LIA, Avignon University	<i>Supervisor</i>



# Contents

<b>I</b>	<b>State of the art</b>	<b>21</b>
<b>1</b>	<b>Spoken dialogue systems and incremental processing</b>	<b>23</b>
1.1	Human dialogue . . . . .	23
1.1.1	Dialogue acts . . . . .	23
1.1.2	Turn-taking in human dialogue . . . . .	24
1.1.3	Incremental speech processing in human dialogue . . . . .	26
1.2	Spoken dialogue systems . . . . .	27
1.2.1	Towards a human-like interaction . . . . .	28
1.2.2	Spoken dialogue systems architecture . . . . .	29
1.2.3	Spoken dialogue systems evaluation . . . . .	33
1.3	Incremental dialogue systems . . . . .	34
1.3.1	Principles . . . . .	34
1.3.2	Advantages of incremental processing . . . . .	36
1.3.3	New challenges raised by incremental dialogue processing . . . . .	37
1.3.4	Existing architectures . . . . .	39
<b>2</b>	<b>Reinforcement learning in dialogue processing</b>	<b>43</b>
2.1	Reinforcement Learning . . . . .	43
2.1.1	Reinforcement in biology . . . . .	43
2.1.2	Markov Decision Processes . . . . .	44
2.1.3	Reinforcement Learning . . . . .	45
2.2	Reinforcement learning in spoken dialogue systems . . . . .	48
2.2.1	In the litterature . . . . .	48
2.2.2	Spoken dialogue systems at Orange and LIA . . . . .	49
2.2.3	Dialogue simulation . . . . .	50
2.3	Reinforcement learning in incremental dialogue systems . . . . .	51
<b>II</b>	<b>Contributions</b>	<b>55</b>
<b>3</b>	<b>Turn-taking phenomena taxonomy</b>	<b>57</b>
3.1	Introduction . . . . .	57
3.2	Taxonomy presentation . . . . .	58
3.3	Discussion . . . . .	63
3.3.1	Dialogue initialisation . . . . .	64

3.3.2	Negative feedback . . . . .	65
3.3.3	Positive feedback . . . . .	65
3.3.4	Reference . . . . .	66
3.3.5	Ordered dialogue acts . . . . .	66
3.3.6	Synthesis . . . . .	67
3.4	Turn-taking phenomena in dialogue systems . . . . .	67
<b>4</b>	<b>Turn-taking decision module: the Scheduler</b>	<b>75</b>
4.1	Description . . . . .	75
4.1.1	Overview . . . . .	75
4.1.2	Time sharing . . . . .	76
4.1.3	The Scheduler . . . . .	77
4.2	Illustration . . . . .	79
4.2.1	A textual dialogue system: CFAsT . . . . .	80
4.2.2	A spoken dialogue system: Dictanum . . . . .	81
4.3	Discussion . . . . .	83
4.3.1	Levels of incrementality . . . . .	83
4.3.2	Enhancing a traditional dialogue system's turn-taking abilities at a low cost . . . . .	84
<b>5</b>	<b>Dialogue strategies</b>	<b>87</b>
5.1	Utterance representation . . . . .	87
5.2	Elementary tasks . . . . .	88
5.3	Slot-filling strategies . . . . .	89
5.3.1	System initiative strategies . . . . .	90
5.3.2	User initiative strategies . . . . .	91
5.3.3	Mixed initiative strategies . . . . .	92
5.3.4	First efficiency comparison . . . . .	92
5.4	Incremental strategies . . . . .	94
<b>6</b>	<b>Incremental dialogue simulation</b>	<b>99</b>
6.1	Overview . . . . .	99
6.2	Incremental dialogue simulation . . . . .	100
6.2.1	User Simulator . . . . .	101
6.2.2	ASR output simulator . . . . .	102
6.3	Personal Agenda management simulated environment . . . . .	104
6.3.1	The Service: personal agenda assistant . . . . .	105
6.3.2	Simulator instantiation . . . . .	107
6.4	Functioning illustration . . . . .	109
<b>7</b>	<b>Handcrafted strategies for improving dialogue efficiency</b>	<b>111</b>
7.1	User and system initiative . . . . .	111
7.1.1	Strategies . . . . .	111
7.1.2	Experiments . . . . .	116
7.2	Incremental strategies . . . . .	117
7.2.1	TTP implementation . . . . .	117

---

7.2.2	Experiments . . . . .	119
<b>8</b>	<b>Reinforcement learning for turn-taking optimisation</b>	<b>123</b>
8.1	Reinforcement learning Model . . . . .	124
8.1.1	Background . . . . .	124
8.1.2	State representation . . . . .	124
8.1.3	Actions, rewards and episodes . . . . .	128
8.1.4	Fitted-Q Value Iteration . . . . .	129
8.2	Experiment . . . . .	130
8.2.1	Setup . . . . .	130
8.2.2	Results . . . . .	132
<b>9</b>	<b>Experiment with real users</b>	<b>133</b>
9.1	The Majordomo domain . . . . .	133
9.2	Experimental protocol . . . . .	134
9.2.1	Implementation . . . . .	134
9.2.2	Conduct of the dialogue . . . . .	135
9.2.3	Key Performance Indicators . . . . .	135
9.3	Results and discussion . . . . .	137
9.4	Discussion . . . . .	138
	<b>List of Figures</b>	<b>145</b>
	<b>List of Tables</b>	<b>147</b>
	<b>Bibliography</b>	<b>149</b>

---

# Acknowledgement

I would like to thank all the people that contributed, directly or indirectly, to create the rich and interesting environment in which I had the chance to work on this thesis. First of all, I would like to thank my supervisors Romain Laroche and Fabrice Lefèvre for their help and their support. Romain provided valuable ideas and advice that helped a lot along the way, as well as a refreshing point of view whenever I felt stuck. Fabrice was also very encouraging and his experienced eye raised critical questions which always led to improved versions of my publications and my work in general. Also, I would like to thank Sophie Rosset, David Schlangen, Bassam Jabaian, Filip Jurčiček and Olivier Pietquin for being part of my jury.

During these three years, I had the chance to meet several other PhD students with which I developed both professional and personal relationships. I would like to thank Layla El Asri for her important and helpful advice, Tatiana Ekeinhor-Komi for always maintaining the good mood and the fun, Emmanuel Ferreira for his help and for welcoming me during my stay at the LIA, Merwan Barlier for always sharing the novelties in the field of artificial intelligence as well as interesting conferences, Nicolas Carrara for the interesting technical discussions. Also, I want to thank Tianxiang Xia for the great job he did during his internship with me at Orange.

It would not have been possible for me to complete this thesis without the help, the trust and the autonomy I was given both at Orange and the LIA. The NaDia team I was enrolled in at Orange were extremely supportive, very interested in my thesis topic and very helpful and reactive when I needed advice concerning the dialogue solution we are working with: Disserto. On the other hand, the researchers at the LIA were very welcoming and interesting to work with.

Finally, I would like to thank my parents for their support during the course of this thesis.



---

# Abstract

Incremental dialogue systems are able to process the user's speech as it is spoken (without waiting for the end of a sentence before starting to process it). This makes them able to take the floor whenever they decide to (the user can also speak whenever she wants, even if the system is still holding the floor). As a consequence, they are able to perform a richer set of turn-taking behaviours compared to traditional systems. Several contributions are described in this thesis with the aim of showing that dialogue systems' turn-taking capabilities can be automatically improved from data. First, human-human dialogue is analysed and a new taxonomy of turn-taking phenomena in human conversation is established. Based on this work, the different phenomena are analysed and some of them are selected for replication in a human-machine context (the ones that are more likely to improve a dialogue system's efficiency). Then, a new architecture for incremental dialogue systems is introduced with the aim of transforming a traditional dialogue system into an incremental one at a low cost (also separating the turn-taking manager from the dialogue manager). To be able to perform the first tests, a simulated environment has been designed and implemented. It is able to replicate user and ASR behaviour that are specific to incremental processing, unlike existing simulators. Combined together, these contributions led to the establishment of a rule-based incremental dialogue strategy that is shown to improve the dialogue efficiency in a task oriented situation and in simulation.

A new reinforcement learning strategy has also been proposed. It is able to autonomously learn optimal turn-taking behaviour throughout the interactions. The simulated environment has been used for training and for a first evaluation, where the new data-driven strategy is shown to outperform both the non-incremental and rule-based incremental strategies. In order to validate these results in real dialogue conditions, a prototype through which the users can interact in order to control their smart home has been developed. At the beginning of each interaction, the turn-taking strategy is randomly chosen among the non-incremental, the rule-based incremental and the reinforcement learning strategy (learned in simulation). A corpus of 206 dialogues has been collected. The results show that the reinforcement learning strategy significantly improves the dialogue efficiency without hurting the user experience (slightly improving it, in fact).

---

# Résumé

Les systèmes de dialogue incrémentaux sont capables d'entamer le traitement des paroles de l'utilisateur au moment même où il les prononce (sans attendre de signal de fin de phrase tel un long silence par exemple). Ils peuvent ainsi prendre la parole à n'importe quel moment et l'utilisateur peut faire de même (et interrompre le système). De ce fait, ces systèmes permettent d'effectuer une plus large palette de comportements de prise de parole en comparaison avec les systèmes de dialogue traditionnels. Cette thèse s'articule autour de la problématique suivante : est-il possible pour un système de dialogue incrémental d'apprendre une stratégie optimale de prise de parole de façon autonome? Tout d'abord, une analyse des mécanismes sous-jacents à la dynamique de prise de parole dans une conversation homme-homme a permis d'établir une taxonomie de ces phénomènes. Ensuite, une nouvelle architecture permettant de doter les systèmes de dialogues conventionnels de capacités de traitement incrémentales de la parole, à moindre coût, a été proposée. Dans un premier temps, un simulateur de dialogue destiné à répliquer les comportements incrémentaux de l'utilisateur et de la reconnaissance vocale a été développé puis utilisé pour effectuer les premiers tests de stratégies de dialogue incrémentales. Ces dernières ont été développées à base de règles issues de l'analyse effectuée lors de l'établissement de la taxonomie des phénomènes de prise de parole. Les résultats de la simulation montrent que le caractère incrémental permet d'obtenir des interactions plus efficaces. La meilleure stratégie à base de règles a été retenue comme référence pour la suite.

Dans un second temps, une stratégie basée sur l'apprentissage par renforcement a été implémentée. Elle est capable d'apprendre à optimiser ses décisions de prise de parole de façon totalement autonome étant donné une fonction de récompense. Une première comparaison, en simulation, a montré que cette stratégie engendre des résultats encore meilleurs par rapport à la stratégie à base de règles. En guise de validation, une expérience avec des utilisateurs réels a été menée (interactions avec une maison intelligente). Une amélioration significative du taux de complétion de tâche a été constatée dans le cas de la stratégie apprise par renforcement et ce, sans dégradation de l'appréciation globale par les utilisateurs de la qualité du dialogue (en réalité, une légère amélioration a été constatée).

---

# Introduction

## Context and thesis subject

Building machines that are able to vocally communicate with users is driven by the desire to make human-computer interaction as natural and efficient as possible. Implementing the way humans converse in a machine involves many issues: no more need for traditional interaction devices like the keyboard, hands-free and eye-free interaction (useful in many situations like cars for example), new communication paradigms where a real human-human conversation is simulated...These issues are also emphasised by the new technological trends in the modern world: Internet of Things (IoT), smart devices...Also, recent advances (especially in the field of speech recognition) gave birth to vocal agents both in academia and industry, even though this kind of systems belonged exclusively to the science-fiction domain a few years ago. Nowadays, it is possible to check for restaurants around, check one's account or send a search query on the web by uttering a few words only. Also, conversational agents are able to understand more and more vocabulary and language variations.

Nevertheless, the ability of these systems to engage in a real and natural conversation is still very limited. One of the main reasons is the oversimplified turn-taking mechanism they use: when the user speaks, the system waits until the end of the user's utterance before processing the request. This mechanism hurts both the naturalness and the efficiency of the dialogue.

Through several contributions, this thesis addresses the problem of turn-taking capabilities enhancement in dialogue systems. Reinforcement learning is applied to an incremental dialogue system to make it able to optimise turn-taking decisions in an autonomous fashion. First, simple hints and explanations are provided in the following in order to give the reader a first intuition of what reinforcement learning and incremental dialogue systems mean. The rest of this thesis clarifies these ideas by providing more precise definitions and by grounding the manipulated notions in the current existing literature.

Originally, a **dialogue** designates a sequence of communication acts between two or more individuals through natural language, either spoken or written (from Greek, *dia* means *through* and *logos* means *speech*), in order to achieve a task or to find an agreement (this last point separates dialogues from conversations). With the emergence of

---

speech technologies, a research thread (to which this thesis belongs) developed machines that are able to substitute these individuals to a certain extent. They are made of a set of elements that are interacting with each other following precise rules, generally in the purpose of performing a specific task. As a consequence, they are referred to as **dialogue systems**.

*Incrementalism* is a method of work aimed at achieving a given task gradually, step by step. The adjective **incremental** designates any process that advances in that way. At each step, each new laid brick is called an increment. How is that related to dialogue systems? In a nutshell, an utterance is incrementally processed if the listener (the system) does not wait until its end before processing it (understanding it on the fly instead). As a result, these **incremental dialogue systems** can also utter words or sentences while the speaker is still holding the floor. Inversely, the user can interrupt the incremental expression of the machine and the system will know when it was interrupted.

In computer science, **Learning** refers to the field of *Machine Learning* which is the science of building models that will drive algorithms to perform a certain task, and calibrating them automatically from data. **Reinforcement** is borrowed from the field of behavioural psychology. A behaviour can be strengthened in many ways, like being more frequently performed, for longer durations or after shorter delays for example. This is generally due to a *positive stimulus* received by the agent under study, after adopting this behaviour. **Reinforcement Learning** is a mathematical framework with algorithms for solving problems through experience.

So, what is the point of applying reinforcement learning to incremental dialogue systems? What is the problem to solve in such systems? Traditional dialogue systems have only one kind of decision to make: what to say. Incremental dialogue systems, on the other hand, are free to speak whenever they want, which adds an extra dimension to the decisions it should make. In this thesis, decisions about the content of what the system should say are not studied, only timing decisions are focused on. Reinforcement Learning is therefore applied to investigate the following question: can an incremental dialogue system learn the proper timing for speaking in an autonomous fashion, directly from its interactions with users?

## Motivations

According to a study performed by Gartner Consulting, the market of artificial intelligence applications was worth 5 billion dollars in 2014 with an exponential potential growth given the forecasts, reaching a 42 billion dollars size in 2024. As far as personal assistants are concerned, the market size should be multiplied by 13 by then. Therefore, these technologies are on the way of becoming profoundly present in several aspects of our every day life. In order for the vocal modality to be used in such a context, it has to be a robust and comfortable way of human-machine interaction in the sense that:

- 
- Conversational agents should be reactive enough for the conversation to be smooth and not tedious like it is the case in currently deployed dialogue systems.
  - Task-oriented dialogue systems should be efficient in such a way that users prefer talking to their devices instead of using any other kind of human-machine interface.
  - Dialogue strategies should be robust to errors and misunderstandings. They should be able to quickly recover from them in order to avoid desynchronisations between the system and the user.

In traditional dialogue systems, the multiplication of dialogue turns degrades the user experience because of the important silences at each floor transition (from the user to the system and vice-versa). For the same reason, recovering from errors becomes costly which leads to frustrating and tiresome interactions for the user. Moreover, the sooner the error is identified, the lesser there is to unravel. In a world where humans interact more and more with machines several times a day, it is crucial to make the user experience more attractive.

Incremental processing is a powerful tool that enables dialogue systems to fix the current turn-taking limitations. An incremental dialogue system is aimed to process the user's request very quickly as she speaks, which makes it able to provide fast answers and fix errors in a reactive way. The user's and the system's speech are no longer viewed as organised sequences of turns but as a continuous signal where two sources are combined.

Nevertheless, the induced freedom in terms of time sharing between the user and the system can lead to chaotic situations if it is not well managed. Humans use several implicit rules to keep their conversations synchronised, therefore, it is interesting to take a close look towards these mechanisms in order to better understand them. The objective of such a study is to make dialogue systems designers aware of the panorama of turn-taking behaviours that they might want to replicate given the task at hand. This can hopefully lead to a first generation of handcrafted turn-taking strategies designed to improve the dialogue efficiency.

Finally, reinforcement learning has been widely used during the last two decades in order to prevent dialogue systems designers to have to set all the parameters. The idea is to automatically optimise an important part of these parameters directly from data with a twofold objective: the designer's task is simplified and the parameter settings are more accurate since they are estimated from real data. In this thesis, reinforcement learning is not applied to classical dialogue management but it is used to optimise turn-taking decisions. This approach is motivated by the fact that the obtained data-driven strategies have the potential to offer better performances than handcrafted ones.



---

## Objectives and contributions

The ecosystem in which this thesis took place is the product of the collaboration between academia (LIA) and industry (Orange Labs). As a consequence, the motivations behind this work are twofold:

- Bringing theoretical insights which can strengthen our understanding of turn-taking mechanisms in human conversation and help designing new algorithms to enhance incremental dialogue systems turn-taking capabilities.
- Building a prototype using industrial tools in order to provide a proof of concept and to demonstrate the advantages of the proposed methodology through real interactions with users.

To do so, the following contributions have been made:

- New incremental dialogue architecture (Khouzaimi et al., 2014b): it transforms a traditional dialogue system into an incremental one at a low cost. As a proof of concept, it has been implemented in a textual (CFAsT) and a vocal dialogue system (DictaNum, (Khouzaimi et al., 2014a)).
- New turn-taking phenomena taxonomy in human-human dialogue (Khouzaimi et al., 2015c): separating turn-taking behaviours made it possible to make clear choices about which ones should be implemented in an incremental dialogue system in order to improve its efficiency.
- New incremental dialogue simulation framework (personal agenda management domain): it is able to simulate incremental Automatic Speech Recognition (ASR) instability (Khouzaimi et al., 2016a).
- Rule-based incremental strategy proposal (Khouzaimi et al., 2015a): it has been implemented in simulation and combined with several slot-filling strategies. The results show that the incremental strategy improves dialogue efficiency in terms of dialogue duration and task completion.
- Data-driven strategy using reinforcement learning (Khouzaimi et al., 2015b): it is implemented in simulation and shown to achieve better results than the rule-based one.
- Real users experiment (Khouzaimi et al., 2016b): an incremental dialogue system prototype (Majordomo domain) has been developed and used to test the previous strategies with real users.

## Outline

The first part of this thesis presents the current state of the art in the field of dialogue systems, incremental dialogue systems and turn-taking as well as reinforcement learning and its application to human-machine dialogue. Chapter 1 introduces the

---

generic architecture of a dialogue system and recalls a brief history about each of its components. Incremental dialogue systems are introduced and two different design paradigms are presented. Finally, the basic theoretical background behind reinforcement learning is provided.

Then the second part thoroughly describes the different contributions that have been made during the course of the thesis. Chapter 3 introduces a new taxonomy of turn-taking phenomena in human-human conversation. A deep analysis of its different components is made following different study dimensions. Moreover, a discussion is led in order to determine which phenomena are more likely to improve dialogue efficiency if implemented in an incremental dialogue system. Then Chapter 4 presents a new architecture for transforming a traditional dialogue system into an incremental one at a low cost as well as two applications that demonstrate its well functioning. This architecture involves a new module called the Scheduler which is in charge of turn-taking decisions. Combining the insights provided in these two chapters, Chapter 5 sheds light on how the selected phenomena from the taxonomy could be implemented in the Scheduler in a task-oriented slot-filling task (three slot-filling strategies are also presented).

In order to be able to generate important dialogue corpora, Chapter 6 introduces a new framework for incremental dialogue simulation which is able to simulate incremental ASR instability. A User Simulator communicates with an incremental dialogue system (composed of a Service for personal agenda management and a Scheduler) through an ASR Output Simulator. The strategies discussed in Chapter 7 are implemented in this simulated environment and compared. Chapter 8 introduces a new strategy that is learned automatically from simulated dialogues using reinforcement learning. The results are compared to the ones offered by handcrafted strategies and presented in Chapter 7.

Finally, Chapter 9 describes an experiment with real users where interactions take place with a Majordomo agent: the users act as if they were interacting with their smart home. A non-incremental strategy as well as a handcrafted incremental and a data-driven ones are compared in order to validate the results obtained by simulation.

---

# List of publications

- (Khouzaimi et al., 2014a) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2014a. Dictanum : système de dialogue incrémental pour la dictée de numéros. In Proceedings of *the TALN 2014 Conference*.
- (Khouzaimi et al., 2014b) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2014b. An easy method to make dialogue systems incremental. In Proceedings of *the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Khouzaimi et al., 2015a) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2015a. Dialogue efficiency evaluation of turn-taking phenomena in a multi-layer incremental simulated environment. In Proceedings of *the HCI International 2015 Conference*.
- (Khouzaimi et al., 2015b) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2015b. Optimising turn-taking strategies with reinforcement learning. In Proceedings of *the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Khouzaimi et al., 2015c) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2015c. Turn-taking phenomena in incremental dialogue systems. In Proceedings of *the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- (Khouzaimi et al., 2016a) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2016a. Incremental human-machine dialogue simulation. In Proceedings of *the International Workshop on Spoken Dialogue Systems (IWSDS)*.
- (Khouzaimi et al., 2016b) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2016b. Reinforcement learning for turn-taking management in incremental spoken dialogue systems. In Proceedings of *International Joint Conference on Artificial Intelligence (IJCAI)*.

## List of publications

---

## **Part I**

# **State of the art**



# Chapter 1

## Spoken dialogue systems and incremental processing

### 1.1 Human dialogue

#### 1.1.1 Dialogue acts

If I say *This dog is big*, I utter a few sounds that can be cast as words. How are these words related to the real objects I refer to? How comes that a sequence of sounds can have effects on others? I can give orders to somebody and make them perform the actions I want as I can congratulate or insult someone and have an emotional impact on her. Also, how comes an utterance can also be judged as a complete nonsense or as a true or false assertion? These are a few questions raised in the philosophy of language.

In his book called *How To Do Things With Words* (Austin, 1962), J.L. Austin focuses on the concept of *speech act*, the title of another book (Searle, 1969) by John R. Searle, who completed this theory of language. Introducing these concepts is aimed towards bringing answers to the previous questions. Saying *My sister just arrived*, one performs a speech act that can be viewed from different points of view. Suppose that someone is listening as this sentence is being uttered and that person does not speak English, then obviously she only hears a sequence of noises which is the physical, low-level nature of the speech act. When considered from that perspective, the latter is referred to as a *phonetic act*. These sounds can then be put together to form a set of vocables and words which constitute a *phatic act*. When the whole meaning of the sentence is taken into account, the speech act is said to be a *rhetic act*. These three levels of analysis are grouped into a bigger concept, the *locutionary act*.

When one focuses not only on the raw meaning of a speech act but on the message that it contains, whether it is a warning or an encouragement for example, it is said to be an *illocutionary act* as beyond the rhetic aspect, an *illocutionary force* is considered. Nevertheless, in (Searle, 1968), Searle rejects this distinction between the rhetic and the illocutionary act made by Austin. According to him, the limit between the two is not



justified hence rejecting the whole *locutionary act* concept. Instead, he suggests to adopt a modified classification. These considerations are, however, beyond the scope of this thesis as the philosophy of language will only be used as a tool to better analyse turn-taking phenomena (see chapter 3). Therefore, only two concepts will be retained: the phonetic act where the speech act is viewed as a succession of sounds and the *illocutionary act* where the meaning is taken into account.

Moreover, saying something to somebody can have psychological effects on that person: congratulating someone or insulting him can be rewarding or hurting, a strongly grounded speech can be convincing, etc. These are referred to as *perlocutionary acts*.

To build a dialogue system, the traditional approach is to consider the user's and the system's speech acts as illocutionary (and sometimes perlocutionary) acts. When studying incremental dialogue systems, the *phonetic act* point of view comes into play. In this thesis, this distinction will be clarified.

In the field of psycholinguistics, Herbert H. Clark brings another vision of dialogue acts in his book *Using Language* (Clark, 1996). The communication channel is split into two tracks, the *communication track* and the *meta-communication track*. The first one is used when the speaker adds new information about the topic of the information whereas the second one is used when she refers to the dialogue itself. For example, saying *OK*, *I see* or nodding her head are meta-communicative acts.

In order to guide the user throughout the dialogue, correct potential errors and to confirm some pieces of information, spoken dialogue systems use the meta-communication track. In this thesis, it is shown that incremental dialogue opens new possibilities to make an even more subtle use of this track.

### 1.1.2 Turn-taking in human dialogue

Turn-taking is a sociological phenomenon that has been generalised to many different situations: card games, road traffic regulation, CPU resource sharing... The term *turn-taking* has been applied in that context for the first time both in (Yngve, 1970) and in Ervin Goffman's personal communications (June 5<sup>th</sup> 1970) independently. (Duncan, 1972) notices that *beyond considerations of etiquette, it is difficult to maintain adequate mutual comprehensibility when participants in a conversation are talking at the same time*. In (Sacks et al., 1974), Harvey Sacks describes the social organisation of turn-taking as an economy where *turns are valued, sought or avoided*, depending on the situation at hand. Then, in the rest of his paper, he focuses on the case of human conversation. Obviously, this is subject to many contextual and cultural variations but the goal of the paper is to meet the challenge of extracting a set of rules that would ultimately describe the human turn-taking mechanisms in a general fashion.

For about six years, Harvey Sacks has been analysing conversation recordings and he came up with a few rules that characterise human conversation and turn-taking. One of them is particularly interesting given the purpose of this thesis: *Transition (from one*

*turn to a next) with no gap and no overlap are common. Together with transitions characterized by slight gap or slight overlap, they make up the vast majority of transitions.*

In fact, this principle has been brought to light even earlier in (Sullivan, 1947) where it has been noticed that the same turn-taking phenomena still hold over most of the languages. Schegloff (1968) suggests that speaking *one party at a time* is one of the basic rules for conversations. Nevertheless, when analysing human conversation more closely, it is shown that humans interrupt each other very frequently (Beattie, 1982; Strombergsson et al., 2013). In this thesis, this idea is extended to human-machine dialogue and it is shown that it might be interesting to interrupt the speaker in some cases in order to increase dialogue efficiency.

During the following decades, a few other attempts to come up with models and classification of turn-taking phenomena in human dialogue have been made. In (Beattie, 1982), a political interview between Margaret Thatcher and Jim Callaghan has been analysed. As a result, the author introduced a classification of turn-taking phenomena where each category is characterised by the answer to these three questions (resulting in a general categorisation with five broad classes):

1. Is the attempted speaker switch successful?
2. Is there simultaneous speech?
3. Is the first speaker's utterance complete?

If the listener manages to successfully take the floor and there is simultaneous speech present, either the speaker's utterance is complete which leads to an OVERLAP or not which translates into a SIMPLE INTERRUPTION. On the contrary, if there is no simultaneous speech, either the speaker's utterance is complete which leads to a SMOOTH SPEAKER SWITCH, either a SILENT INTERRUPTION is involved. Finally, if the listener does not succeed in taking the floor, only one category is reported: BUTTING-IN INTERRUPTION. In (Gravano and Hirschberg, 2011), this classification is extended by introducing the following question: *Is the listener's utterance in response to the speaker's one and indicates only "I am still here/I hear you and please continue?"*. As a consequence, two new categories are introduced when the answer is yes and given the fact that there is simultaneous speech or not: BACKCHANNEL and BACKCHANNEL WITH OVERLAP.

Optimising turn-taking means taking the floor at the right time. Humans are very good at detecting the cues for these timings (Raux and Eskenazi, 2008; Jonsdottir et al., 2008; Gravano and Hirschberg, 2011): prosodic features, lexical features, semantic features, ...

Actually, humans even tend to start speaking before the end of their interlocutor's utterance, even interrupting him sometimes. A study led in (Strombergsson et al., 2013) reports statistics about times and overlaps in human conversation both directly and over the telephone. Given the type of question that is addressed to the listener, the latter tends to respond more or less quickly, often inferring the end of the question and starts uttering the response before its end.

In reality, human turn-taking involves even more complicated behaviours resulting in intertwined turns between the speakers. [Raux and Eskenazi \(2009\)](#) tries to provide a simple model with very few assumptions. It is a state machine where the following states are considered (initially depicted in ([Jaffe and Feldstein, 1970](#))):

- Only the user speaks.
- Only the system speaks.
- No one speaks because the user stopped talking.
- No one speaks because the system stopped talking.
- Both the user and the system speak after a user barge-in.
- Both the system and the user speak after a system barge-in.

In this framework, four basic turn-taking transitions are introduced: *turn transitions with gap*, *turn transitions with overlap*, *failed interruptions* and *time outs*. This is a very low-level model where only phonetic acts are considered (unlike the classification proposed in ([Beattie, 1982](#)) for example where the intent of the listener when he tries to take the floor is taken into account). A similar approach is used in ([Włodarczak and Wagner, 2013](#)), but the situations identified are different:

- **Within-speaker silence:** The speaker stops for a while and then resumes his utterance.
- **Between-speaker silence:** The most intuitive case of turn taking. The speaker stops and after a moment of silence, the listener takes the floor.
- **Within-speaker overlap:** The listener either takes the floor or performs an intervention that is not meant to interrupt the speaker and the latter keeps the floor.
- **Between-speaker overlap:** The listener starts speaking before the end of the speaker's utterance, hence resulting in an overlap and a turn transition.
- **Solo vocalisation:** One person speaks in an unilateral way.

### 1.1.3 Incremental speech processing in human dialogue

During a conversation, the listener does not wait for the speaker's utterance to end before trying to understand it. It is processed incrementally and on top of the intuition that people have related to this, a few studies provided convincing arguments supporting the idea. The most famous and convincing examples are eye-tracking based studies ([Tanenhaus et al., 1995](#); [Eberhard et al., 1995](#); [Arnold et al., 2000](#)). Subjects are given an image to look at through a head-mounted eye-tracking mechanism that records their eye-gaze at a millisecond scale. Then, ambiguous and unambiguous sentences were uttered, for example:

- **Ambiguous version:** Put the apple on the towel in the box.

- **Unambiguous version:** Put the apple that is on the towel in the box.

For this example, when the users are provided with an image of an apple on a towel, a towel with nothing on it and a box, their eye-gaze shows different patterns depending on which version of the utterance they listen to. In the ambiguous case, they tend to start by looking at the apple, then to the towel with nothing on it, then to the apple again and finally at the box. In the unambiguous case, they directly look at the apple and then to the box. Most importantly, these movements happen as the sentence is uttered.

This is also pretty much related to the garden path sentence effect. Consider the following sentence (taken from the related Wikipedia article): *the government plans to raise taxes were defeated*. Most people feel the need to read it twice or even several times before understanding its meaning. When one starts reading *the government plans to raise taxes*, she automatically understands that taxes are planned to be raised by the government but then comes the disturbing end of the sentence: *were defeated*. The only solution is to parse *plans* as a noun and not as a verb. This is also an argument in favour of human incremental processing.

Finally, when humans are reading a text, they tend to skip a few words with no loss of the meaning. In (Ilkin and Sturt, 2011), the authors show that it is possible to predict a few words while reading given the context and the sentence structure (eye-tracking has also been used here). For example, when the readers reach the sentence *The worker was criticised by his boss*, the word *boss* seems to be guessed ahead of time, again supporting the idea of incremental processing.

## 1.2 Spoken dialogue systems

A spoken dialogue system (SDS) is an automated application that interacts directly with a human being in natural language. Virtual assistants like Siri (Apple) or Cortana (Microsoft) are good examples of SDSs. They are task-oriented since their goal is to help the user achieve some task. There also exist a few SDSs that are only used for tutoring (Jordan et al., 2015) or simply chatting and companionship (Sidner et al., 2013).

Historically, McCarthy and Hayes were the first to suggest the introduction of mental qualities (intentions, belief, ...) in a dialogue system (Mccarthy and Hayes, 1969; Mccarthy, 1979) in the 70's. This started a new research thread aimed at describing the internal state and the system's behaviour in a humanlike fashion (Newell, 1980; Bratman, 1987; Cohen and Levesque, 1990; Sadek, 1991; Konolige and Pollack, 1993). In Orange, a new dialogue system called ARTIMIS has been developed. It was based on a new theory of interaction (Sadek, 1991) using the Belief-Desire-Intentions model introduced in (Bratman et al., 1988). During the following years it has been improved and enhanced with new capabilities: an action plan (Louis, 2002), a preference-based operator (Meyer, 2006) and uncertainty management capacities (Laroche et al., 2008).

However, this approach that uses logic has been abandoned for two main reasons.

Firstly, it does not guarantee the VUI-completeness (Pieraccini and Huerta, 2005; Paek and Pieraccini, 2008): it can lead to system behaviours that have not been specified in advance by the designer, and secondly it requires an expertise in the logic field (which considerably reduces the span of potential dialogue system designers). As a result, like all the major actors in the field (Nuance, AT&T, SpeechCycle, etc.), Orange turned to a new solution where the dialogue system is viewed as an automaton. This solution is called Disserto which has proved to be simple, robust enough to allow the development of widely deployed industrial dialogue systems and flexible enough to allow state-of-the-art research. All the services developed in this thesis have been generated using Disserto which will be presented more deeply later on.

### 1.2.1 Towards a human-like interaction

Traditional computer interfaces are presented in the form of controls in an application (text boxes, buttons) or web pages. They are heavily used and they have been proven to be efficient enough, providing an accurate way of human-machine interaction. So, why building spoken dialogue systems? What are the advantages of the vocal modality?

An obvious motivation for building spoken dialogue systems is the quest for human-likeness. In (Edlund et al., 2008), an interesting analysis of the way humans perceive dialogue systems they are interacting with is performed. These systems are complex and humans keep a biased representation of them during the interaction (called metaphors in the article). Two of them are the most common:

- *Interface metaphor*: The system is viewed as an interface just like non dialogue-based systems. Users adopting this representation of the system tend to use vocal commands instead of complete sentences. Moreover, they tend to remain silent when the system tries to behave like a human, by saying *Hi!* for example.
- *Human metaphor*: In this case, the users tend to view the system as a real human, therefore adopting a more natural way of communication. These users generally have higher expectations of the system's ability to answer their requests as well as its ability to perform human-like turn-taking behaviours.

Nevertheless, it is also legitimate to raise the question whether human-likeness should be a goal in the conception of dialogue systems or not. In (Edlund et al., 2008), four kinds of objections to pursuing that goal are discussed. The first one is the feasibility: is there any hope that someday, systems that behave exactly like humans could be built? This has raised huge debates during the last decades. The second question is utility: do people really need systems that imitate humans? Apart from the fact that it would help us to better understand the way humans communicate, it is an interesting application in the case of some applications like companionship and entertainment. It is less obvious when it comes to task-oriented dialogue. The third point is related to the concept of *uncanny valley* (Mori, 1970): as machine's human-likeness increases, they reach a point where they start behaving like humans without having the same capacities which makes them disturbing for the user. The reason behind that is the concept of symmetry (last brought up question in the paper): if machines have a similar

behaviour to humans, then users are likely to push the human metaphor to the limit by unconsciously thinking that machines really understand what they are saying, thus expecting more complex reactions like the ones due to emotions for example. Nevertheless, by bringing even more improvements, this valley will be hopefully crossed and human-like solutions that no longer have such problems will be built.

The multiplication of services and support platforms in modern society engenders huge costs that dialogue systems can help to reduce. By analysing client paths while requiring assistance from an expert, it is possible to identify commonly asked questions and recurrent patterns. By gathering such information, it is also possible to design dialogue systems that can interact with users and respond to their requests without any human intervention (of course, in the case the interaction fails, the client is redirected to a human-operated service platform). This can dramatically reduce service and support costs and also improve the quality of service as it is accessible any time, with no interruptions (unlike real platforms that are more often open at working time only). Moreover, when a client calls, the answer is immediate and her call is no longer queued causing waiting time (which she may pay for) with sometimes no answer in the end.

With the development of the IoT during the last few years, new human-machine interactions can be imagined. For instance, the concept of *Smart Home* is currently making its entry in the market of Artificial Intelligence through the contributions of several companies and start-ups: Amazon Echo, etc. In this context, the advantage of speech communication is clearly relevant as it is hands- and eyes-free. The user can command her house from any room with no extra device needed. In some situations, her hands are already occupied by some other task. For example, she can be cooking ([Laroche et al., 2013](#)) while asking *What should I put next in my salad?* or *Can you add milk to my shopping list please?*.

Finally, as talking agents and talking robots can also be designed for entertainment and companionship ([Gelin et al., 2010](#); [Sidner et al., 2013](#)) and the vocal modality is the most natural way of interaction, it is very useful in this area.

### 1.2.2 Spoken dialogue systems architecture

The classic architecture of an SDS is made of five main modules (Figure 1.1):

1. Automatic Speech Recognition (ASR): transforms the user's audio speech signal into text.
2. Natural Language Understanding (NLU): outputs a conceptual representation of the user's utterance in text format.
3. Dialogue Manager (DM): given the concepts extracted from the user's request, a response (in a conceptual format too) is computed.
4. Natural Language Generation (NLG): transforms the concepts computed by the DM into text.



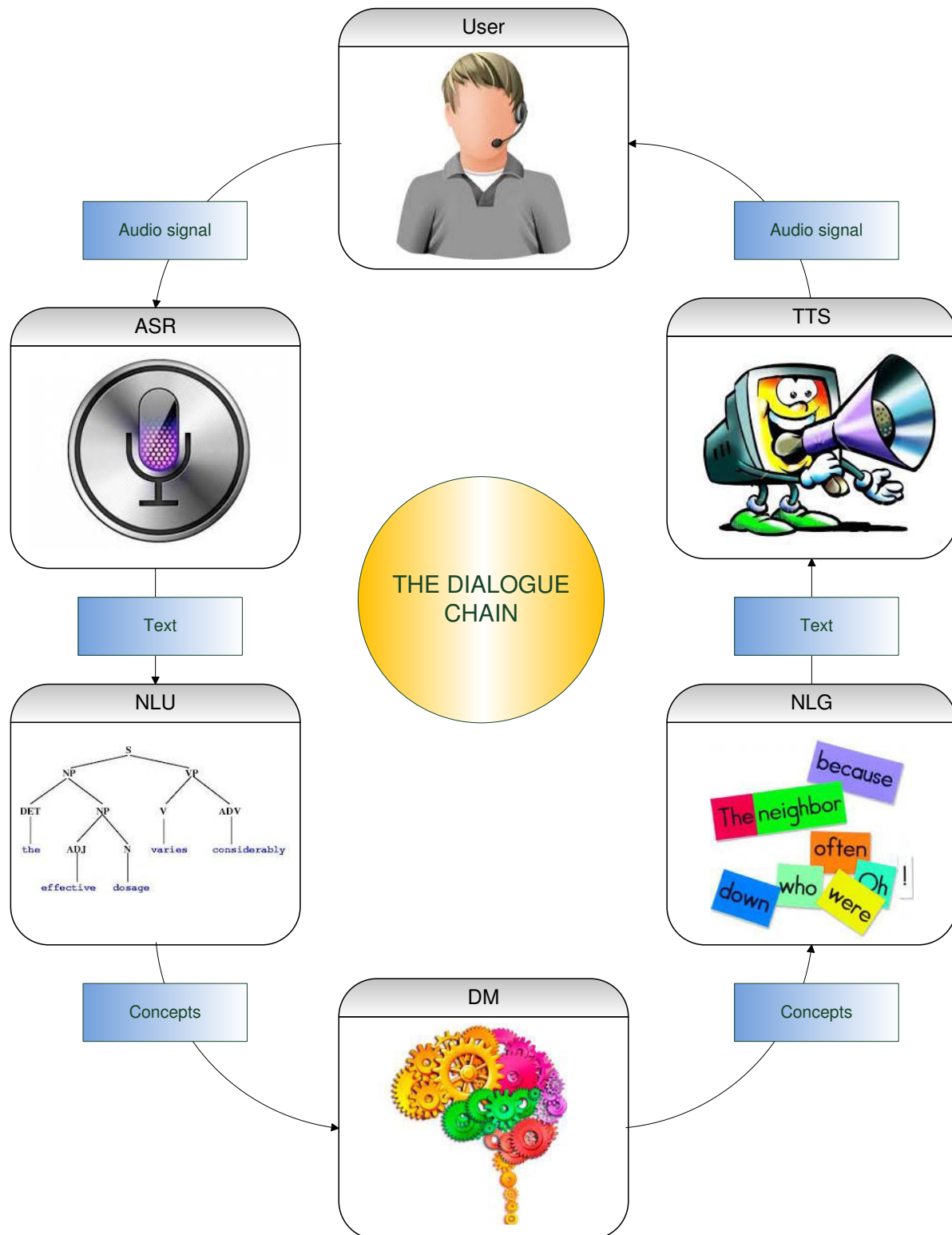


Figure 1.1: The dialogue chain

5-Best	
Hypothesis 1:	I would like to book a flight from New-York to Los Angeles (0.965)
Hypothesis 2:	I could like to book a flight for New-York to Los Angeles (0.931)
Hypothesis 3:	Cold like a book fly New-York to Los Angeles (0.722)
Hypothesis 4:	Cold like look flight pork to Los Angeles (0.570)
Hypothesis 5:	Cold bike look fly New-York to Los Angeles (0.441)

**Figure 1.2:** A 5-Best example corresponding to the sentence “I would like to book a flight from New-York to Los Angeles”.

5. Text-To-Speech (TTS): reads the text outputted by the NLG by using a synthetic voice.

## Automatic Speech Recognition

Speech recognition technology is an old problem with long history. During the 1950s, a group of researchers from Bell Labs developed the first technology that was able to recognise digits from speech (in fact, speech perception has been under study since the early 1930s). Then, during the second half of the last century, new advances have made it possible to build ASR solutions with larger vocabulary and with no dependence on the user. In the 1960s, Hidden Markov Models (HMMs) were proved to be useful for speech recognition ([Gales and Young, 2007](#)) and they were the most popular technique two decades later. Commercial products using ASR technology had to wait until the 1990s to be finally released in the market as they reached an interesting vocabulary scope (even though their accuracy and their delay were far behind the technology available today). Performances kept improving slowly and gradually until 2009, when Deep Learning algorithms were tested ([Mohamed et al., 2009](#); [Deng et al., 2013](#)) introducing huge improvement; the Word Error Rate (WER) decreased by 30%. During the last six years, research continued in that direction giving birth to accurate and reactive speech recognition solutions (Google, Nuance, Sphinx, Kaldi...). These solutions also provide results incrementally in a continuous fashion. Therefore, ASR is less and less considered as a bottleneck in the development of spoken dialogue systems, and as it will be shown later, the delays they offer make it possible to design reactive incremental dialogue systems. Commercial off-the-shelf ASR solutions like Google ASR or Nuance products are able to recognise almost every word in many languages, including named entities. Finally, the ASR output is not only the text that the recognition algorithm figures out to be the best match for the input audio signal, but a list of the N most likely hypotheses and the corresponding confidence scores: it is called the N-Best. For instance, a 5-Best example is represented in Fig. 1.2.



### Natural Language Understanding

NLU is a sub-field of Natural Language Processing (NLP) which scope is wider than the spoken dialogue field. Since the 1950s, researchers have been trying to develop several models and ontologies in order to automatically process natural language with several applications in sight: topic recognition, sentiment analysis, news filtering and analysis, natural speech processing, etc. The ambition manifested during the 1950s and the early 1960s quickly had to face reality as the expected objectives were far from being reached. As a consequence, research in this area was significantly slowed down between the 1960s and the 1980s. During the last decade, NLP research has found a second wind thanks to new Machine Learning techniques, bringing them at the heart of lucrative businesses like recommendation and digital marketing. NLU refers to the set of techniques in order to make the machine understand the underlying structure of a text in natural language. To do so, a lexicon as well as an ontology (concepts and the links between them in a specific domain) should be built. Therefore, earlier NLU solutions are based on a set of handcrafted parsing rules, however, new statistical-based models ([Macherey, 2009](#); [Hahn et al., 2010](#)) have been proven to be more robust and easy to generalise over domains. Also Deep Learning has also been applied to NLP showing interesting performances that NLU modules can benefit from ([Bengio et al., 2003](#); [Collobert et al., 2011](#); [Ferreira et al., 2015a,b, 2016](#)).

### Dialogue Management

As far as Dialogue Management is concerned, two decades ago, for the first time, dialogue has been modeled as Markov Decision Processes (MDPs) problem (see Sect. 2.1), hence being solved using reinforcement learning ([Levin and Pieraccini, 1997](#)). The dialogue state contains all the information needed to determine what is the best action to perform as well as the quality of that state (roughly speaking, to what extent it is desirable for the system to be in that state). The possible actions in each state are the dialogue acts that the system can make while being in that state. In 2007, in order to represent the uncertainty over the user's intent (due to ASR or NLU imperfections), dialogues have been cast as Partially Observable Markov Decision Processes (POMDPs) ([Roy et al., 2000](#); [Williams and Young, 2007](#)). This gave raise to the notion of *belief tracking* which objective is to keep a distribution over the possible user intents. Also, existing dialogue systems are able to interact with the user in the domain they are built for only, however, during the last few years, researchers have been pushing the boundaries of domain extension methods ([Gasic et al., 2013](#)) and open-domain systems ([Pakucs, 2003](#); [Galibert et al., 2005](#); [Ekeinhör-Komi et al., 2014](#); [Wang et al., 2014](#)). Finally, it has been shown that it is possible to use reinforcement learning to continuously adapt to users online ([Ferreira and Lefèvre, 2015](#); [Chandramohan et al., 2012b,a](#)).

### Natural Language Generation

The NLG task is the inverse of the NLU one: it translates dialogue acts concepts into sentences. It started being used in the 1990s for purposes like business and financial news summary (Anand and Kahn, 1992). A few start-ups and big companies also provide automatic text generation solutions that are used to quickly produce reports or official letters. The main challenge for such systems is to be able to generate a variety of different words, expressions and sentences in order for them not to be repetitive and for the result to be as realistic as possible. This is even more crucial when it comes to dialogue systems as they are supposed to simulate real conversations with the user, which is a highly variable. Nevertheless, in practice, template-based generation methods are the most widely used even though data-driven solutions are starting to reach mature development (Mairesse et al., 2010; Mitchell et al., 2014; Manishina et al., 2016).

### Speech Synthesis

During the 1930s, Bell Labs were not only interested in ASR but they also developed a new approach for the reciprocal task: the TTS (also known as *speech synthesis*). The human speech is broken into small acoustic components that are sequentially pronounced by the system. They built the first machine demonstrating this mechanism: the Voder (Dudley, 1939). As far as this task is concerned, the challenge for the system is to sound as human-like as possible, in terms of phoneme transitions, speech rate and prosody. Two research threads tackle this problem in two different manners (Tabet and Boughazi, 2011): the first one uses corpus-based methods where the resulting audio is a combination of pre-recorded phonemes extracted from a corpus and the second one uses HMMs to generate a completely synthetic voice. Even though substantial advances have been accomplished since the Voder, including data-driven approaches (Yu et al., 2011), it is still easy to distinguish between a synthesised and a real human voice.

#### 1.2.3 Spoken dialogue systems evaluation

When building dialogue systems and improving them, it is necessary to determine metrics in order to measure their evolution. What makes a dialogue system better than another one? What metrics should be taken into account while evaluating dialogue systems? What are the most important characteristics of a dialogue system that should be improved?

A distinction can be made between the evaluation of a dialogue system as a whole (*usability* as it is called in (Möller et al., 2007)) and the evaluation of the different components separately. In the second case, more standard metrics exist such as the Word Error Rate (WER) for the ASR or the CER (Concept Error Rate) when it comes to evaluating the NLU. However, all errors are treated the same whereas in reality, some of them have more impact than others. This is a complex problem since even humans disagree when it comes to assessing the *gravity* of each error (Rosset et al., 2013). Also, as far as the

DM is concerned, such metrics do not exist: the usability evaluation is preferred to a local evaluation of the DM since they are very correlated (Dybkaer et al., 2004). As a consequence, many evaluation frameworks for dialogue systems have been developed and used both in academia (Walker et al., 1997; Hone and Graham, 2000; Schmitt et al., 2012) and industry (Evanini et al., 2008; Witt, 2011).

In the evaluation literature, a distinction is made between two kinds of metrics that are usually used for evaluating dialogue systems: objective and subjective metrics. The first category contains all the Key Performance Indicators (KPIs) that can be measured by an algorithm by accessing the dialogue information and meta-information, whereas the second one is made of the user's appreciations of the dialogue quality or specific characteristics like human-likeness or to what extent the user enjoyed the dialogue experience.

Objective metrics that are commonly used are the dialogues' mean duration and the task completion ratio. Generally speaking, these two metrics are correlated as the user gets impatient when the dialogue lasts for too long (the user can also get impatient for other reasons, like the repetition of the same system's dialogue act several times). Moreover, the user's speech rate and the way they communicate introduces some variability when using these metrics. Finally, it is legitimate to ask the question: are shorter dialogues the real desired objective? First, this depends on the type of dialogue system at hand. If it is designed for entertainment or companionship, then there is no need for seeking faster dialogue strategies. However, in the case of task-oriented dialogue, looking for shorter dialogues makes sense as a measure of efficiency. In these situations and especially for daily tasks, an efficiency threshold has to be necessarily reached in order for people to use the dialogue system at hand.

Subjective metrics are generally gathered using a survey at the end of each dialogue or by making experts rate dialogues afterwards. Several metrics can be collected this way: the global quality of the dialogue, naturalness/human-likeness, reactivity, etc. However, this also raises the problem of variability between users. Most often, they are asked to evaluate the system on a Likert scale (1 to 5 or 1 to 10), but a user answering 4 could be equivalent to another user answering 3 or 5. Therefore, the absolute evaluation is less significant than the relative one given a specific user.

## 1.3 Incremental dialogue systems

### 1.3.1 Principles

Most dialogue systems have a simple and rigid way of managing turn-taking. The interaction mode they offer is similar to a walkie-talkie conversation as the system waits for the user to finish her utterance before taking the floor and vice-versa (even though some systems allow the user to interrupt them). Such systems will be referred to as *traditional dialogue systems* in this thesis.

The first idea of incremental systems goes back to incremental compilers (Lock,

1965). An incremental compiler processes each new instruction independently from the previous ones. Therefore, a local modification of the code does not affect the whole result of the compilation. The idea of processing natural language in an incremental way is first introduced in (Wirén, 1992) according to (Kilger and Finkler, 1995). Instead of feeding modules with complete utterances, the input is pushed chunk by chunk (500ms of audio signal, one word of a sentence...etc...) making the output change several times before the end of the user's utterance. Nevertheless, in his book *Speaking: From Intention to Articulation* (Levelt, 1989), Levelt analysed the mechanisms underlying the way humans formulate their ideas in natural language and already reported that the processes involved are incremental. The approach is closer to computational linguistics than psycholinguistics. The second part of the dialogue chain (DM, NLG and TTS) is analysed using a different terminology: the *Conceptualizer*, the *Formulator* and the *Articulator*.

As discussed in Sect. 1.1.3, in human-human conversation, the listener does not wait for the speaker to finish his sentence before processing it; it processes it as it is spoken. As a consequence, human beings perform a large panel of turn-taking behaviours while speaking, like backchanneling (*aha, yeah, ...*) or barging-in.

To replicate these behaviours, a new generation of SDSs has been the focus of research for the last few years. An SDS is said to be incremental when it is able to process the user's speech on the fly. The input signal is divided into small chunks and the growing sentence is reprocessed at each new chunk (Schlangen and Skantze, 2011). Table 1.1 gives an example illustrating the functioning of an incremental NLU module (in a hotel room booking service). For the sake of simplicity, processing delays are neglected.

Time step	NLU input	NLU output
1	I	empty
2	I won't	empty
3	I would	empty
4	I would like to	empty
5	I would like to cook a	empty
6	I would like to book a room	<b>action:</b> BOOK
7	I would like to book a room on May	<b>action:</b> BOOK
8	I would like to book a room on May 7 <sup>th</sup>	<b>action:</b> BOOK <b>date:</b> 05-07
9	I would like to book a room on May 17 <sup>th</sup>	<b>action:</b> BOOK <b>date:</b> 05-17
10	I would like to book a room on May 17 <sup>th</sup> and I will	<b>action:</b> BOOK <b>date:</b> 05-17
11	I would like to book a room on May 17 <sup>th</sup> and I will be driving	<b>action:</b> BOOK <b>date:</b> 05-17 <b>parking:</b> YES

Table 1.1: Example of incremental NLU processing

This mechanism allows the system to react earlier, even before the end of the user's utterance. On the other hand, allowing the user to interrupt the system does not require incremental capabilities; it has been successfully implemented in non-incremental dialogue systems (Lamel et al., 2000; El Asri et al., 2014).

### 1.3.2 Advantages of incremental processing

Before discussing the different reasons why incremental processing should be preferred to rigid turn-taking, it is important to note that a few studies with real users have shown that incremental dialogue systems offer a better user experience. For instance, in (Aist et al., 2007), an ordinal regression has been performed between the user satisfaction and several features with a flag for incremental processing among them. A significant correlation between incremental processing and the global user satisfaction has been found. Other studies also confirm the advantage of incremental speech processing (Skantze and Schlangen, 2009; El Asri et al., 2014; Zhao et al., 2015).

As discussed in Section 1.2.1, human-likeness is a legitimate goal for dialogue systems (at least worth trying). When talking to each other, humans perform a rich set of turn-taking phenomena (see Section 1.1.2) and in spite of the fact that they do not talk in a rigid walkie-talkie manner, they manage to avoid desynchronisations and to keep a conversation that is going forward. Replicating these behaviours from the machine's point of view can therefore be interesting. It might be expected that the user feels more at ease while using a more human-like turn-taking mode hence pushing the human metaphor even further.

The other aspect that is interesting about incremental dialogue is reactivity. As the system processes the user's request before its ends, it is possible to design accurate end-point detection in order to detect the end of this request as soon as possible (Raux and Eskenazi, 2008). Moreover, incremental dialogue systems can interrupt the user to report a problem, like in the following example:

USER: I would like to book a room for tomorrow with ...

SYSTEM: Sorry, we are full tomorrow.

This can help the user get to her goal faster but one should be very careful about the way it is implemented as there is a risk that user interruption actually harms the user experience (even though the intent is to go faster, see Section 1.3.3). Nevertheless, a corpus study led in (Ghigi et al., 2014) showed that when users are interrupted, they tend to adopt a more sober way of expression, hence directly increasing the dialogue efficiency but also indirectly as the risk of misleading off-domain words and expressions is reduced (Zhao et al., 2015).

Early barge-in both from the user and the system's side is also a way to limit desynchronisations. An example of a desynchronised dialogue could be (inspired by the NUMBERS domain described in (Skantze and Schlangen, 2009)):

USER: 01 45 38 37 89

SYSTEM: 01 45 28 37 89

USER: No, not 28 but 38

SYSTEM: Sorry, you mean 28 38

USER: What?

SYSTEM: 28 38 1 (*The system understood “One” instead of “What?”*)

In this example, the user could have reported the mistake earlier if he could barge-in:

USER: 01 45 38 37 89

SYSTEM: 01 45 28...

USER: No, 38

SYSTEM: Ok. 01 45 38

USER: 31 89

SYSTEM: 31 89

Another interesting aspect about incremental dialogue systems is that they can leverage multimodality (Fink et al., 1998). In fact, there are two aspects of multimodality and they can both benefit from incremental processing:

- **Input multimodality:** Most researchers in the community refer to this aspect when talking about multimodal systems. A system input can be multimodal in the sense that it can handle speech input, but also text, gesture or eye-gaze for example. The main challenge faced by this kind of setup is the problem of mixing these inputs in order to infer the correct user intent. In the case of multimodal systems, the world is considered as a flow of information coming from multiple sources (Chao and Thomaz, 2012; Rosenthal et al., 2013). The different modalities are not necessarily sampled with the same rate nor support the same delays, therefore, it is important to find convenient ways to synchronise them.
- **Output multimodality:** The machine can also use different channels of information while communicating (Matthias, 2009). For example, the speech modality can be used at the same time as a moving avatar with facial expressions (the Furhat for example (Skantze and Johansson, 2015)). It can also be coupled with the input multimodality paradigm to create highly interactive interfaces (Johnston et al., 2014).

### 1.3.3 New challenges raised by incremental dialogue processing

The first problem to consider when talking about incremental spoken dialogue systems is the question of ASR *latency*, which is the time needed by the recognition algorithm to provide the text output corresponding to an audio signal. As discussed earlier, the



ASR accuracy has been a bottleneck in the development of spoken dialogue systems for many years but thanks to recent advances in this field, it is no longer the case. Similarly, incremental dialogue systems require quick responses from the ASR but speech recognition modules have been too slow for many years which was a limiting factor in the development of incremental dialogue processing. But, in the last few years, ASR technology has become reactive enough (Breslin et al., 2013; Plátek and Jurčiček, 2014). Still, it is important to be aware that there is a tradeoff between the accuracy, the vocabulary size and the latency. Kaldi, which is an ASR solution designed by researchers (and which is mostly used by them), makes it possible to design one's own acoustic and language model as well as setting one's own parameters in order to control this tradeoff. Off-the-shelf solutions like Google ASR do not give the user such possibilities (however, accuracy and delays in open domain are well balanced for most applications).

If the successive partial results of an incremental ASR module are observed during a user's utterance, it is likely that the progression they follow is not monotonous. In other words, a partial result is not guaranteed to be a prefix of results to come. The following example showing successive ASR results illustrates this phenomenon:

1. Hum
2. I
3. Hum good
4. iPod
5. I would
6. I good bike
7. I would like

This phenomenon is called ASR *instability* (or *stability* depending on the sources) (Selfridge et al., 2011). This factor is also related to the tradeoff between latency and accuracy as preferring fast ASR over accurate ones can lead to very unstable results (the system is not given enough time to compute accurate results most of the time, thus ending up delivering wrong partial results frequently), and vice-versa.

This leads to one of the main challenges raised by incremental processing: the ability to *revise* the current hypothesis. All the modules in the dialogue chain are impacted by this problem. As an illustration, suppose that the user interacts with an incremental personal assistant on her phone and makes the following request: *Please call the number 01 45 80 12 35*. The last digit is first understood as being 30 and then 35, therefore, if the system is too reactive, there is a risk that it starts calling the wrong number and maybe starts uttering the sentence: *Ok, calling 01 45 80 12 30*. Afterwards, the system understands 35 instead of 30 hence needing a correction mechanism in order to stop the TTS, to cancel the call, to perform a new one and to provide a new answer to the user. Nevertheless, even though the system at hand is equipped with such a mechanism, using it very often is not an optimal way of managing incremental input as it causes extra delay as well as non-natural behaviour (stopping the TTS and starting again with

another utterance). This introduces a similar tradeoff to the one discussed for the ASR module but from the DM perspective: if decisions are taken too quickly, some of them may be wrong hence activating the correction mechanism. On the other hand, if the DM is slow to take action, then it lacks reactivity and there is no point for it to be incremental. As a consequence, it is important to determine the right moment to stick with the current partial utterance and to take action based on it (Raux and Eskenazi, 2008; Lu et al., 2011).

Incremental NLG also raises new problems which are illustrated in (Baumann and Schlangen, 2013). In this paper, a system has to describe the trajectory of a car in a virtual world. When the latter approaches an intersection where it has to turn right or left (no road straight ahead), then the system utters something like *The car drives along Main Street and then turns...euh...and then turns right*. In this example, the system is sure that the car is going to turn which makes it possible for it to commit to the first part of the sentence with no risk. However, this is not always the case as a new chunk of information from the user can change the whole system's response. In this thesis, the NLG is not incremental as the DM's response is considered to be computed instantly at each new micro-turn (event though it is not necessarily stable and it can vary from micro-turn to micro-turn). Finally, in purely vocal applications, computing the NLG results incrementally does not make much sense as the user's and the system's utterances do not overlap most of the time (Sacks et al., 1974). However, this is an interesting behaviour as far as multimodal applications are concerned.

Building an incremental TTS module can also be very tricky. In order for the synthetic voice to be the most human-like as possible, prosody should be computed accurately and to do so, the sentence's structure and punctuation have to be taken into account. This information is no longer given in the case of incremental TTS or it arrives too late. Baumann (2014) proposes a method for coping with the problem. It consists in using both low-level and high-level information to predict the right prosody.

### 1.3.4 Existing architectures

#### Sequential paradigm

A general abstract model of incremental dialogue systems has been introduced in (Schlangen and Skantze, 2011). In this approach, the dialogue chain is maintained and each one of the five components is transformed into its incremental version. This view of incremental dialogue systems will be referred to as the *sequential paradigm*.

Each module is composed of three parts, the Left Buffer (LB), the Internal State (IS) and the Right Buffer (RB). As described in Section 1.2.2, each module is also characterised by the type of input it processes as well as the type of output it computes. In incremental dialogue, all these data flows have to be divided into small chunks which are called Incremental Units (IU). For example, the audio signal that is given as an input to the ASR module can be divided into 500ms chunks that are processed one by one. Each IU is first added to the LB, then it is taken by the IS for processing and once a



result is available, a new IU of a new kind is outputted in the RB. The RB of one module is the LB of the following one in the dialogue chain so the data propagation through the dialogue system is insured.

Because of ASR instability, new IU in the LB does not necessarily imply that a new IU will be pushed into the RB on top of the ones that already existed there. An example given in (Schlangen and Skantze, 2011) is the following: suppose the user utters the number *forty* which processed incrementally, then first the ASR outputs *four* and then *forty*. As a consequence, the second hypothesis does not complete the first one but it replaces it in the RB. This phenomenon will be discussed in more details in Chapter 6.

Adopting this paradigm is a natural way of enhancing traditional dialogue systems with incremental capabilities. It is interesting from a computational and design point of view as the different tasks are separated. Therefore, one is able to evaluate the different components independently (Baumann and Schlangen, 2011) and have a global view to determine which area still needs improvement.

### Multi-layer paradigm

The problem of dialogue management in traditional dialogue systems can be formulated as follows: at each dialogue turn, given the dialogue context (including the last user's utterance), what is the right dialogue act to perform? In the incremental frame, this definition no longer holds as dialogue acts are no longer attached to dialogue turns. Therefore, one way to tackle the problem is to split the dialogue management task in two components, the high-level and the low-level handlers. This paradigm is directly motivated by Austin's, Searl's and Clark's contributions discussed in Section 1.1.1 as the high-level module handles illocutionary acts (the communicative track) whereas the low-level one manages phonetic acts (the meta-communicative track).

As reported in (Lemon et al., 2003), this approach is more in alignment with results in the psycholinguistic field. The phenomena observed at the phonetic level are complex, and the interaction happens on multiple levels, not always following the classical dialogue chain. Having a separate module for handling these phenomena is therefore a more natural paradigm.

Switching from the traditional dialogue management approach to the incremental one is also a transition from discrete time to continuous time, from a synchronous to an asynchronous processing (Raux and Eskenazi, 2007). The low-level module is continuously (approximated by a high frequency processing in computers) listening to the outside world and waiting for events that might be interesting to communicate to the high-level handler. In that case, the latter returns actions (dialogue acts) and it is the role of the low-level module to choose whether to retrieve them to the user or not as well as choosing the right moment in case it decides to speak.

Finally, starting from a traditional dialogue system, it is easier and more straightforward to transform it into an incremental one if one adopts this paradigm. Adding an extra low-level module to the dialogue manager is enough (Selfridge et al., 2012).

At each new incremental input, this module sends the whole partial utterance from the beginning of the current turn to the dialogue manager and gets a response. Based on that and eventually some other features, it decides whether to take the floor or not. As most of the requests sent to the dialogue manager are “fake” as they are not meant to be acted on, they should not affect the dialogue context. Therefore, either multiple instances of the dialogue manager are used, either the dialogue context is saved and restored at each new request, unless the low-level module decides to take the floor (see [Chapter 4](#) for additional explanations).



## Chapter 2

# Reinforcement learning in dialogue processing

### 2.1 Reinforcement Learning

#### 2.1.1 Reinforcement in biology

Reinforcement Learning (RL) is a sub-field of machine learning where an agent is put into an environment to interact with, and figures out through the process of *trial and error* what the best actions to take are, given a reward function to maximise (Sutton and Barto, 1998) (see Figure 2.1<sup>1</sup>).

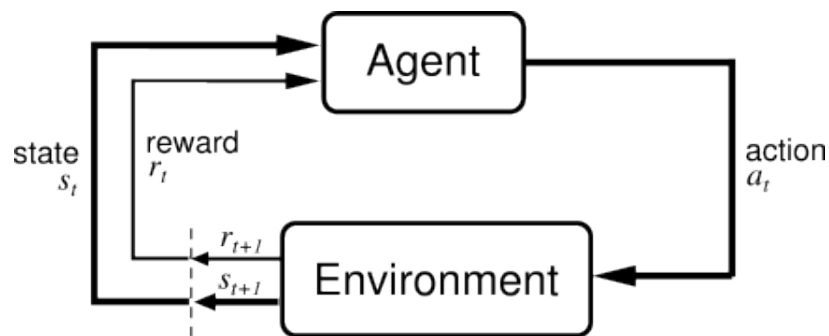


Figure 2.1: The interaction cycle between the agent and the environment in reinforcement learning

It was first inspired by the field of animal psychology where living organisms are considered as the agents. Rewards are associated with stimuli that the agent seeks like food for example. Conversely, punishments are stimuli that it tries to avoid like important heat for instance. In (Thorndike, 1898), hungry animals were put in enclosures where the only way to escape and find food is to perform some simple act (pulling at a loop of cord, pressing a lever, stepping on a platform...). If after a certain period of time

<sup>1</sup>This figure is taken from (Sutton and Barto, 1998).

they were not able to escape, they are taken out of the box without being immediately fed. This experiment showed how animals were able to learn what to do in order to escape from the enclosure.

Similarities between reinforcement learning and neurons behaviours in the brain were also discovered. In (Schultz et al., 1995; Schultz, 1998), similarly to the previous experiment, monkeys were put in situations where the accomplishment of an action is necessary to get food. Then the reaction of their dopamine neurons was analysed. Among many other applications where reinforcement learning meets neuroscience, (Doya, 2007) commented on this experiment: *Although dopamine neurons initially responded to the rewards, when those rewards became fully predictable from preceding sensory cues, such as light and sound, their reward responses went away. Instead, dopamine neurons started to respond to reward-predictive sensory cues. If the reward is omitted after learning, dopamine neuron firing was suppressed at the timing when reward delivery is expected. These are interesting findings on their own, but most exciting for those who are familiar with reinforcement learning theory because it exactly matches what the TD error does.*

### 2.1.2 Markov Decision Processes

The most common model consists in casting the problem as a Markov Decision Process (MDP) which is a quintuple  $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$  where:

- $\mathcal{S}$  is the *state space*. At each time step  $t$ , the agent is in some state  $s_t \in \mathcal{S}$ .
- $\mathcal{A}$  is the *action space*. At each time step  $t$ , the agent decides to take action  $a_t \in \mathcal{A}$ .
- $\mathcal{T}$  is the *transition model* where each  $(s, a, s')$  in  $\mathcal{S} \times \mathcal{A} \times \mathcal{S}$  is associated with a real number in  $[0, 1]$  corresponding to the probability  $\mathbb{P}(s_{t+1} = s' | s_t = s, a_t = a)$ . A more compact notation will be used in the following:  $\mathcal{T}_{ss'}^a = \mathcal{T}(s, a, s')$ .
- $\mathcal{R}$  is the *reward model*. Let  $r$  be the immediate reward due to taking action  $a$  in state  $s$ , then  $\mathcal{R}$  is the set of distributions of  $r$  for every  $(s, a) \in \mathcal{S} \times \mathcal{A}$ . The following notation will be used in the rest of this chapter:  $\mathcal{R}_{ss'}^a = \mathbb{E}[\mathcal{R}(s, a, s') | s, a, s']$ .
- $\gamma \in [0, 1)$  is referred to as the *discount factor*. In the RL framework, the aim of the agent is not to maximise the immediate reward but the *expected return*, where the return  $R_t$  is defined as follows:

$$\begin{aligned} R_t &= r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots \\ &= \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \end{aligned} \tag{2.1}$$

Therefore, when  $\gamma = 0$ , the agent maximises the immediate reward only and when  $\gamma$  tends towards 1, the agent maximises the sum of all the future rewards. In other words, the parameter  $\gamma$  controls how far-sighted is the agent in terms of future rewards.

A *policy*  $\pi : \mathcal{S} \rightarrow \mathcal{A}$  is a mapping between the state space and the action space. An agent is said to follow the policy  $\pi$  when for each time  $t$ , it takes the action  $a_t = \pi(s_t)$ . A policy can also be stochastic, in which case,  $\pi(s, a)$  denotes the probability of choosing action  $a$  when the agent is in state  $s$ . A key aspect of MDPs is the *Markov property*. Being in state  $s$  is the only information available to predict the future, and adding information about what happened during previous time steps has no power of prediction. Therefore, given a policy, each state  $s \in \mathcal{S}$  is associated with a value  $V^\pi(s)$  which is the expected return when in this state and following the policy  $\pi$  afterwards:

$$V^\pi(s) = \mathbb{E}[R_t | s_t = s, \pi] \quad (2.2)$$

Another interesting quantity is the expected return knowing the current state but also the next action, after which  $\pi$  is followed. This is referred to as the Q-function:

$$Q^\pi(s, a) = \mathbb{E}[R_t | s_t = s, a_t = a, \pi] \quad (2.3)$$

Given the definition of  $R_t$ , one can notice that

$$\begin{aligned} V^\pi(s_t) &= \mathbb{E}[R_t | s_t, \pi] \\ &= \mathbb{E}[r_t + \gamma \sum_{k=0}^{\infty} \gamma^k r_{(t+1)+k+1} | s_t, \pi] \\ &= \mathbb{E}[r_t + \gamma R_{t+1} | s_t, \pi] \\ &= \mathbb{E}[r_t + \gamma \mathbb{E}[R_{t+1} | s_{t+1}, \pi] | s_t, \pi] \\ &= \mathbb{E}[r_t + \gamma V^\pi(s_{t+1}) | s_t, \pi] \end{aligned} \quad (2.4)$$

This is known as the Bellman equation for  $V^\pi$  and it can also be written for the Q-function, as follows

$$Q^\pi(s_t, a_t) = \mathbb{E}[r_t + \gamma Q^\pi(s_{t+1}, \pi(s_{t+1})) | s_t, a_t, \pi] \quad (2.5)$$

### 2.1.3 Reinforcement Learning

A natural question that can be asked at this point is: how are these values computed? In reinforcement learning, this is known as the *evaluation problem*. The transition model  $\mathcal{T}$  and the reward model  $\mathcal{R}$  are the elements that define the dynamics of the MDP. If they are known,  $V^\pi$  can be directly computed:

$$\begin{aligned}
 V^\pi(s) &= \mathbb{E}[R_t | s_t = s, \pi] \\
 &= \sum_{a \in \mathcal{A}} \pi(s, a) \mathbb{E}[R_t | s_t = s, a_t = a, \pi] \\
 &= \sum_{a \in \mathcal{A}} \pi(s, a) \mathbb{E}[r_t + \gamma R_{t+1} | s_t = s, a_t = a, \pi] \\
 &= \sum_{a \in \mathcal{A}} \pi(s, a) \sum_{s' \in \mathcal{S}} \mathcal{T}_{ss'}^a (\mathcal{R}_{ss'}^a + \gamma \mathbb{E}[R_{t+1} | s_{t+1} = s', \pi]) \\
 &= \sum_{a \in \mathcal{A}} \pi(s, a) \sum_{s' \in \mathcal{S}} \mathcal{T}_{ss'}^a (\mathcal{R}_{ss'}^a + \gamma V^\pi(s')) \tag{2.6}
 \end{aligned}$$

It is possible to define an order over the policies. Saying that  $\pi_1$  is better than  $\pi_2$  means that for all the states  $s$ ,  $V^{\pi_1}(s) \geq V^{\pi_2}(s)$ . It can be shown that there exists at least one policy that is better than all the others: it is called the *optimal policy* ( $\pi^*$ ).  $V^{\pi^*}$  will be referred to as  $V^*$  (same for all optimal policies) and it is defined as

$$\forall s \in \mathcal{S}, V^*(s) = \max_{\pi} V^\pi(s) \tag{2.7}$$

Similarly, one can define  $Q^*$  as

$$\forall (s, a) \in \mathcal{S} \times \mathcal{A}, Q^*(s, a) = \max_{\pi} Q^\pi(s, a) \tag{2.8}$$

The aim of reinforcement learning is to learn the optimal policy. Similarly to what has been shown for  $V^\pi$ , if the transition and the reward models are known, the Bellman equation corresponding to  $V^*$  (called the *Bellman optimality equation*) can be written with respect to these models (similarly to 2.6):

$$V^*(s) = \max_a \sum_{s' \in \mathcal{S}} \mathcal{T}_{ss'}^a (\mathcal{R}_{ss'}^a + \gamma V^*(s')) \tag{2.9}$$

A similar form can be also shown about the Q-function

$$Q^*(s, a) = \sum_{s' \in \mathcal{S}} \mathcal{T}_{ss'}^a (\mathcal{R}_{ss'}^a + \gamma \max_{a' \in \mathcal{A}} Q^*(s', a')) \tag{2.10}$$

A set of *Dynamic Programming* methods exists in order to efficiently solve these kinds of equations and come up with the optimal policy given the transition and the reward model (knowing  $Q^*$  implies knowing  $\pi^*$  as the latter is the greedy policy with respect to the former Q-function, in the sense that  $\pi^*(s) = \arg\max_a Q^*(s, a)$ ). However, even

though this kind of approaches are theoretically interesting, they only have a few practical applications as most of the times,  $\mathcal{T}$  and  $\mathcal{R}$  are unknown. The agent learns directly from interacting with the environment (*model-free* approach).

It is possible to try to learn  $\mathcal{T}$  and  $\mathcal{R}$  first and then applying a *model-based* algorithm to figure out the optimal policy. Nevertheless, this is not necessary as most algorithms compute the optimal policy by directly estimating the Q-function. This can be done in a straightforward fashion by running several episodes<sup>2</sup>, computing the returns for each state-action couple and for each episode, then using the mean return over all the episodes as an estimate of  $V^\pi$  or  $Q^\pi$ . Algorithms using this kind of approach belong to the category of *Monte-Carlo methods*.

However, as the agent interacts with the environment, it encounters the following dilemma which was originally faced in the bandit problem (Berry and Fristedt, 1985; Bubeck and Cesa-Bianchi, 2012): how to manage the trade-off between *exploration* and *exploitation*. While being at a state  $s \in \mathcal{S}$ , the agent can choose one action among many. Let us say that the Q-function is initialised as a zero function. Therefore, at the beginning the agent has no preference and selects a random action. If this yields a positive reward, then the agent has the choice between these two options to make the next decision:

1. Making the same decision again as it already knows that it is likely to generate a positive reward.
2. Picking another action because it may yield an even greater reward.

In the first case, the agent is exploiting its current knowledge of the environment whereas in the second case, it is said to be exploring as it is increasing its knowledge about the environment (with the risk of generating low or negative rewards in the meanwhile). Because rewards are stochastic, it is not obvious to determine whether sufficient data is available to trust our estimates and start exploiting most of the time. This is a difficult problem and a simple way to deal with it is to use the  $\epsilon$ -greedy approach, where the agent chooses a random action with a probability of  $\epsilon$  and sticks to the greedy action (with respect to the current estimated Q-function) the rest of the time. Nevertheless, more robust solutions have already been suggested like Upper Confidence Bound (UCB) (Auer et al., 2002) for the bandit problem and Upper Confidence Reinforcement Learning (UCRL) (Auer and Ortner, 2005) for reinforcement learning.

On policy reinforcement learning algorithms keep evaluating the current policy and at the same time, altering that policy in order to improve it. A naive approach would be to fix the current policy and to perform as many evaluation iterations as necessary in order to gain a certain confidence over the estimations of  $V$  or  $Q$  and then to derive a new policy to follow, given these values. This is known as *Policy Iteration* but this is not the most efficient way to proceed (so many iterations are needed). In fact, performing only one evaluation iteration before the next policy improvement step can be shown to be enough, keeping the convergence guarantees. This is referred to as *Value Iteration*.

---

<sup>2</sup>To keep things simple in this introduction to reinforcement learning, the MDP is considered to eventually stop.



Also, the notion of iteration can be viewed differently given the approach and the algorithm at hand. In order to refer to the general idea of intertwining evaluation and control, the expression *General Policy Iteration* (GPI) is used.

In fact, it is also possible to evaluate  $V$  or  $Q$  in an even more fine-grained manner. Instead of waiting until the end of the episode to update these values, it is possible to do it after each new decision. That is what *Temporal-difference (TD) methods* do. In comparison with the Monte-Carlo approach, the new sample for  $V^\pi(s)$  or  $Q^\pi(s, a)$  is no longer the real return obtained in the episode but an estimated one using the Bellman equation. In the case of the *sarsa* algorithm<sup>3</sup>, the Q-function is updated as follows<sup>4</sup> ( $\alpha_t$  being a decreasing parameter with time):

$$Q_t(s_t, a_t) = Q_t(s_t, a_t) + \alpha_t[r_t + \gamma Q_t(s_{t+1}, \pi_t(s_{t+1})) - Q_t(s_t, a_t)] \quad (2.11)$$

It is important to notice that  $a_{t+1}$  is the action chosen by following the current estimated policy derived from  $Q$  ( $\epsilon$ -greedy for example) and which will be actually followed in the next step. The sarsa algorithm is therefore called an *on-policy* algorithm. These conditions can be relaxed giving birth to another category of algorithms, the *off-policy* ones. The most famous is *Q-Learning*<sup>5</sup> (Watkins, 1989) where the Q-function is updated as follows:

$$Q_t(s_t, a_t) = Q_t(s_t, a_t) + \alpha_t[r_t + \gamma \max_a Q_t(s_{t+1}, a) - Q_t(s_t, a_t)] \quad (2.12)$$

Here, the policy used for evaluation is not necessarily the one that is followed.

## 2.2 Reinforcement learning in spoken dialogue systems

### 2.2.1 In the literature

Reinforcement learning has been first applied to dialogue systems in (Levin et al., 1997) and since then, it has been the leading machine learning framework in the field. The dialogue state at time  $t$  is generally determined by the history of dialogue acts since the beginning of the dialogue. At each turn, the set of actions is made of all the possible answers at that time.

Dialogue has first been cast as an MDP. For instance, one of the earliest applications of this framework is described in (Singh et al., 1999). The system involved handles a simple slot-filling task where it decides whether to ask for all the slots at once, to

---

<sup>3</sup>See (Sutton and Barto, 1998) for the algorithm description.

<sup>4</sup>At this point,  $V$  will no longer be used, as  $Q$  is the most commonly used in reality.  $V$  is mostly used for pedagogical purposes.

<sup>5</sup>See (Sutton and Barto, 1998) for the algorithm description.

ask for a specific slot or whether to perform a confirmation. In order to handle ASR imperfections, Partially Observable Markov Decision Processes (POMDPs) (Roy et al., 2000; Williams and Young, 2007; Young et al., 2010; Thomson and Young, 2010) can be used. In this framework, the dialogue state is replaced by a distribution over all possible states which is a more natural way of modeling uncertainty, however, they are more complex and more difficult to scale (Lemon and Pietquin, 2007). Another interesting approach that has been applied to dialogue systems is Hierarchical Reinforcement Learning (Cuayáhuitl et al., 2007). It is meant to handle large state spaces with an important dimensionality, which is often the case in dialogue management. It requires the dialogue to be cast as a Semi-Markov Decision Process (SMDP) (Bradtke and Duff, 1994; Barto and Mahadevan, 2003). Huge and complex state spaces are also dealt with by using *summary states* in most cases, which can be built in several ways. Nevertheless, new approaches using Deep Reinforcement Learning methods started being developed very recently (Cuayáhuitl et al., 2015); instead of using handcrafted features for state representation, raw data is directly fed to a deep neural network.

Also, it is noteworthy that even though the main focus in this thesis is dialogue management, reinforcement learning has also been applied to the NLG task in order to optimise information presentation (Walker, 2000; Rieser and Lemon, 2011) and even TTS to decide what kind of prosody should be used (Bretier et al., 2010).

### 2.2.2 Spoken dialogue systems at Orange and LIA

Important research work has been accomplished at Orange during the CLASSiC project. It was mainly focused on the problem of reconciling academic research with industrial activity (Paek, 2007). A new reinforcement model has been developed (in the continuity of work done by (Singh et al., 2002; Williams, 2008)): the Module Variable Decision Process (MVDP) (Laroche, 2010). It has been implemented in an appointment scheduling task hence giving birth to the first dialogue system learning on-line (directly from experience) (Putois et al., 2010). In addition, other research efforts have been made in order to make reinforcement results accessible directly in design mode.

During the course of this thesis, Orange has also focused on other aspects of dialogue such as reinforcement learning convergence speed (El Asri and Laroche, 2013), interaction quality prediction (El Asri et al., 2012, 2014) as a well as reward function inference and state space representation. In (El Asri et al., 2012; Asri et al., 2016), reward shaping is used to learn a reward function directly from a corpus of dialogues with experts' ratings. Moreover, a new framework called Genetic Sparse Distributed Memory for Reinforcement Learning (GSDMRL) has been proposed (El Asri et al., 2016) which purpose is to compute a state representation that is adapted to the utility function to maximise.

Vocal assistants are becoming a part of our everyday life since their introduction in the market. They are able to perform several tasks but they are still static and limited. Therefore, Orange also investigates solutions to make dialogue systems able to manage several tasks by merging dialogue models of different applications which also makes it

easily extensible (Ekeinhor-Komi et al., 2014). Finally, designing systems that can listen to human/human conversations and make decisions based on them is also a topic that is addressed in this lab (Barlier et al., 2015). It is motivated by several applications like connected houses, meeting rooms, call centers...

As far as LIA (Laboratoire Informatique d'Avignon) is concerned, several subjects concerning human-robot interaction (mainly through dialogue) and Natural Language Processing (NLP) are driving the research activity. Among them:

- **Interactive Voice Response (IVR):** The objective of the Port-MEDIA (Lefèvre et al., 2012) project is to design robust multi-language and multi-domain models for IVR, an IVR being the interface between a user and a database (Jabaian et al., 2013, 2016).
- **Human-Robot interaction:** The main objective of this research field is to design adaptive algorithms to improve the interaction between humans and robots. These are mainly reinforcement learning algorithms. The robot performs poorly at an early stage but with experience, through an error-trial process, it learns to better itself and improve the interaction quality (Ferreira and Lefèvre, 2013; Ferreira and Lefèvre, 2015; Ferreira et al., 2015c).
- **Automatic Speech Translation:** French/English and English/French translation algorithms have been built in collaboration with the LIG (Laboratoire Informatique de Grenoble), based on the Mooses Toolkit. In 2001, the French/English algorithm won second place in the international campaign WMT (Potet et al., 2011; Rubino et al., 2012; Huet et al., 2013).

### 2.2.3 Dialogue simulation

A couple of decades ago, with the development of the dialogue systems research field, the need for evaluation means in order to assess their quality started getting more and more important. Therefore, researchers turn to user simulation methods (also referred to as user modeling). In (Eckert et al., 1997), some of the advantages of these techniques are depicted: the possibility to quickly generate corpora for machine learning techniques at a low cost, easy modeling of different user populations and the possibility of using the same user model across different concurrent dialogue systems for comparison. Nevertheless, the authors recognise that user simulation cannot totally replace interactions with real users in the process of designing reliable dialogue systems: *we believe that tests with human users are still vital for verifying the simulation models.*

Simulating users accurately is a challenging task as their behaviours vary considerably from one person to another and the same user can change her preferences over time (concept-drift) (Schatzmann et al., 2006). Evaluating a user simulator and whether it handles such variability or not is a research track in itself (Pietquin and Hastie, 2013) and the qualities required are of different kinds. The trained user simulator should be consistent with the data that has been used for the training and the sequence of dialogue

acts generated should be coherent. In addition, when it is used in turn to train a data-driven dialogue strategy, the quality of the latter is also an evaluation criteria. Also, it is important that the results obtained in simulation give strong indications about the behaviours with real users.

User simulation is useful during the conception phase of a dialogue system. However, training the simulator from data needs the dialogue system to be conceived already. Therefore, trying to come up with a simple model with only a few parameters is not always a bad idea as it has been proven to achieve good results as well (Schatzmann et al., 2007).

User simulator is also quite similar to the dialogue management task. As a consequence, it is legitimate to ask the following question: why not use reinforcement learning to train user simulators? The answer is that in the case of dialogue management, it is easier to come up with a reasonable reward function: task completion, dialogue duration, subjective evaluation...etc... When it comes to user simulation, the objective function is how well a real user is imitated which is difficult to evaluate. Fortunately, there exists a framework where the reward function is automatically inferred from data which is particularly useful here: Inverse Reinforcement Learning (Russell, 1998; Chandramohan et al., 2011; El Asri et al., 2012).

When it comes to incremental dialogue systems, the only existing user simulator in our knowledge is the one described in (Selfridge and Heeman, 2012). Its state is updated every 10 ms. However, the *ASR instability* phenomenon is not replicated, that is to say that the ASR hypothesis construction is monotonous which introduces a certain bias in comparison with incremental dialogue in real conditions: when a new audio signal increment is heard by the ASR, the output can be partially or totally modified. This simulator made it possible to simulate incremental dialogue processing for the first time, nevertheless, only the case where a new increment is added to the output is modeled.

## 2.3 Reinforcement learning in incremental dialogue systems

In the field of incremental dialogue and turn-taking management, supervised learning is common. The main problem tackled by researchers is the identification of the exact moments where the system should take the floor in order to achieve smooth turn-taking (Raux and Eskenazi, 2008; Gravano and Hirschberg, 2011; Meena et al., 2013). Binary classifiers are used and the features they use are of different natures: lexical, semantic, prosodic...etc...However, a few papers tackled this problem by using reinforcement learning.

Jonsdottir et al. (2008) used reinforcement learning while considering prosodic features only. Backchanneling for example can be performed by humans independently from the meaning. The cost function (negative reward) is taken as gaps and overlaps, hence following Sack's principle discussed in Section 1.1.2.

[Dethlefs et al. \(2012\)](#) adopted a complementary approach where only the semantic content of the user's utterance is taken into account (hierarchical reinforcement learning is used). In human conversation, it is more likely for the listener to react right after a relevant information. Similarly, in the case of a restaurant finding spoken dialogue system, the system should react right after understanding the restaurant's type or price range. In this work, the information pertinence is measured by the Information Density (ID). Therefore, the higher ID during system actions, the more reward it gets.

Instead of trying to minimise gaps and overlaps, the reward function can be designed in a way to optimise dialogue duration and task completion as in ([Selfridge and Heeman, 2010](#)). The system in this paper learns optimal initial turn-taking, in the sense that when a silence is detected, the dialogue participant that has the most relevant thing to say takes the floor first (both dialogue participants are modeled). Like in the previous paper, only semantic features are considered.

A third approach to optimise turn-taking in spoken dialogue systems is to directly try to imitate human behaviours. In ([Kim and Banchs, 2014](#)) Inverse Reinforcement Learning is used to infer a reward function directly from user trajectories in a collected dialogue corpus. Therefore, the reward function automatically incorporates objective and subjective dialogue quality criteria. The authors have made the choice not to consider lexical and semantic features, but rather to limit their work to timing and prosody signals.

## State of the art conclusion

The previous two chapters provide an overview of the current state of art related to dialogue systems in general, incremental dialogue systems as well as reinforcement learning and its applications to human-machine dialogue. In Chapter 1, a few generic notions about human dialogue and the philosophy of language are discussed before providing a definition of what is a dialogue system as well as its different components. Incremental dialogue is also defined afterwards and the different advantages, challenges and architectural considerations that are associated with it are depicted. Reinforcement learning is then discussed in Chapter 2 and a state of the art relative to its application to dialogue systems is provided. Both traditional and incremental dialogue systems are covered.

The next chapters are dedicated to the contributions that have been made during the course of this thesis. A turn-taking phenomena taxonomy in human dialogue has been proposed and used to build a rule-based turn-taking strategy. Then, a new architecture aimed at transforming a traditional dialogue system into an incremental one at a low cost has been built. Also, an incremental user simulator has been implemented and used to build an optimal turn-taking strategy using reinforcement learning, which has been shown to outperform its rule-based counterpart. Finally, these results have been validated in a live study with real users.



**Part II**

**Contributions**





## Chapter 3

# Turn-taking phenomena taxonomy

### 3.1 Introduction

In Chapter 1, the reader is given some clues and some previous work references in order to build a first intuition of what turn-taking is. Here, an analysis of turn-taking in human conversation is performed. It is aimed to provide an answer to the four following questions:

1. What phenomena characterise turn-taking in human conversation?
2. How can they be classified in order to clearly identify the similarities and the differences between them?
3. What are the general categories that emerge from the general picture drawn by this classification?
4. What phenomena are worth implementing in dialogue systems and why?

To do so, a new *turn-taking phenomena taxonomy* is introduced (Khouzaimi et al., 2015c). Compared to the existing classifications presented in 1.1.2, it is aimed to go further by using several levels of analysis hence providing a better grasp of the concept of turn-taking by breaking it into small manageable pieces that can be implemented and studied separately.

Each element of the proposed taxonomy will be referred to as a *turn-taking phenomenon* (TTP). The analysis levels laid in the philosophy of language will be used here while discussing the taxonomy: the locutionary, the illocutionary and the perlocutionary paradigms. Nevertheless, the locutionary act concept is subject to a few disagreements between J. L. Austin and his successor, J. R. Searle. Looking deeper into a locutionary act, it can be broken into three sub-levels: the *phonetic*, the *phatic* and the *rhetic* which correspond to the verbal, the syntactic and the semantic levels. In (Searle, 1968), the author argues that there is no distinction possible between the rhetic level and the illocutionary one and therefore refuses the more general distinction between locutionary and illocutionary level. As a result, he suggests to adopt the four layer structure

composed of: *phonetic acts*, *phatic acts*, *propositional acts* (the act of expressing the proposition) and *illocutionary acts*. Nevertheless, the philosophical subtleties brought by these considerations are beyond the scope of this thesis and the objective in this chapter is to provide pragmatic criteria that will make it possible to distinguish the several TTP at hand. Therefore, only the three following analysis levels have been taken from this theory of language (as they are enough for this analysis): the phonetic, the illocutionary and the perlocutionary levels.

Recall that at the perlocutionary level, the impact that a dialogue act is aimed to have is considered, like convincing, congratulating or insulting for example. Here, an extra dimension is also needed: what is the motivation behind a dialogue act? In the taxonomy introduced in this chapter, some TTPs are exactly the same if viewed as phonetic, illocutionary and perlocutionary dialogue acts, but the reasons why they are performed are different. Making this distinction is interesting from a computational point of view as it is directly correlated to the set of features that are considered by the system in order to make turn-taking decisions.

### 3.2 Taxonomy presentation

Let us consider the three following dialogue situations:

#### Dialogue 1

HECTOR: I would like to try some exotic destination this summer where I can ...

TANIA: ... Have you ever been to India?

#### Dialogue 2

HECTOR: First you put the meat in the oven ...

TANIA: ...aha...

HECTOR: ...then you start preparing the salad...

#### Dialogue 3

HECTOR: What time is it please?

TANIA: It is half past two.

In all the dialogues, Hector initially has the floor and then Tania performs a dialogue act. In dialogues 1 and 2, she does not wait for him to finish his utterance before doing so, unlike in the last dialogue. Therefore, Tania can choose the timing of her intervention at different stages in the progression of Hector's utterance. The first criterion used in the taxonomy introduced here corresponds to this decision. Moreover, in dialogues

1 and 3, Tania utters a complete sentence unlike in dialogue 2. The second criterion is aimed to make the distinction between these kinds of behaviours performed by Tania.

More formally, turn-taking in dialogue refers to the act of taking the floor by one participant (Tania in the previous examples), here called the Taker (T). Two cases can be distinguished; either the other participant, here called the Holder (H), is already speaking or not (the denomination Holder is more adapted to the case where it has the floor, but it is kept as a convention for the other case).

The taxonomy introduced here is mainly aimed to enhance turn-taking behaviours of dialogue systems and compared to several previous studies, the semantic level is given more importance. Therefore, two dimensions are considered:

1. **The quantity of information that H has already injected in the dialogue from T's perspective<sup>1</sup>:** This measures how early in H's utterance T chooses to perform her dialogue act.
2. **The quantity of information that T tries to inject by taking the floor:** T's dialogue act can consist on some implicit reaction (gestures, sounds like *aha*), a complete utterance or something in between.

The different levels of information for each dimension are described in Table 3.1.

<b>H_NONE</b>	No information given <sup>2</sup>
<b>H_FAIL</b>	Failure to deliver any information
<b>H_INCOHERENT</b>	Incoherent information
<b>H_INCOMPLETE</b>	Incomplete information
<b>H_SUFFICIENT</b>	Sufficient information
<b>H_COMPLETE</b>	Complete utterance
<b>T_REF_IMPL</b>	Implicit ref. to H's utterance
<b>T_REF_RAW</b>	Raw ref. to H's utterance
<b>T_REF_INTERP</b>	Reference with interpretation
<b>T_MOVE</b>	Dialogue move (with improvement)

*Table 3.1: Taxonomy labels*

Table 3.2 describes the taxonomy where turn-taking phenomena (TTP) are depicted. The rows correspond to the levels of information added by H and the columns to the information that T intends to add. In order to describe each one of them in details, they are discussed row by row. Moreover, these TTP are shown to fit into several categories which are represented by different colors in the table and explained in Section 3.3.

**H\_NONE:** H does not have the floor, therefore, T takes the floor for the first time in the dialogue. This can be done implicitly by performing some gesture to catch H's attention or by clearing her throat for instance (FLOOR\_TAKING\_IMPL) or more explicitly either by providing a raw message (FLOOR\_TAKING\_RAW) like *I would like to talk to you* or by providing more details about the reasons why T wants to start a conversation (FLOOR\_TAKING\_INTERP): *I am leaving tomorrow and I really have to talk*

<sup>1</sup>Since T is making the turn-taking decisions, the analysis is performed from her perspective.

	T_REF_IMPL	T_REF_RAW	T_REF_INTERP	T_MOVE
H_NONE	FLOOR_TAKING_IMPL	FLOOR_TAKING_RAW	FLOOR_TAKING_INTERP	INIT_DIALOGUE
H_FAIL	FAIL_IMPL	FAIL_RAW	FAIL_INTERP	FAIL_MOVE
H_INCOHERENCE	INCOHERENCE_IMPL	INCOHERENCE_RAW	INCOHERENCE_INTERP	INCOHERENCE_MOVE
H_INCOMPLETE	BACKCHANNEL	FEEDBACK_RAW	FEEDBACK_INTERP	BARGE_IN_CHANGE
H_SUFFICIENT	REF_IMPL	REF_RAW	REF_INTERP	BARGE_IN_RESP
H_COMPLETE	REKINDLE_IMPL	REKINDLE_RAW	REKINDLE_INTERP	END_POINT

**Table 3.2:** Turn-taking phenomena taxonomy. The rows/columns correspond to the levels of information added by the floor holder/taker.

to you about the problem with my insurance. On the other hand, she can start speaking normally (INIT\_DIALOGUE).

**H\_FAIL:** H takes the floor for long enough to deliver a message (or at least a chunk of information) but T does not understand anything. This can be due to noise or to the fact that the words and expressions are unknown to T (other language, unknown cultural reference, unknown vocabulary...). T can interrupt H before the end of his utterance as she estimates that letting him finish it is useless. This can be done implicitly (FAIL\_IMPL) using a facial expression (frowning), a gesture or uttering a sound:

H: Chaque heure passée ici...

T: ...what?

It can also be done by explicitly uttering that H's utterance is not clear so far (FAIL\_RAW):

H: <noise> has been <noise> from...

T: ...sorry, I can't hear you very well! What did you say?

Moreover, T can interrupt H by trying to provide a justification to the fact that H needs to repeat, reformulate or add complementary information in his sentence (FAIL\_INTERP). For example:

H: Freddy was at the concert and ...

T: ...who is Freddy?

Finally, T can also decide to move the dialogue forward without understanding H's utterance (FAIL\_MOVE). This situation happens when T thinks that H's utterance has little chance to be relevant for her or when she thinks it is time to change the discussion topic. For instance:

H: <noise> <noise>...

T: Well, I am not sure I understood your point but frankly speaking, I think we should talk about these details later and focus more on the main problem.

**H\_INCOHERENCE:** T understands H's message and detects an incoherence in it, or between that message and the dialogue context. H can make a mistake like *I went swimming from 10 am until 9 am*, or, *First, go to Los Angeles, then go south to San Francisco...* He can also be unaware of the dialogue context: *You should take metro line A...* when metro line A is closed that day. Again, this can be done implicitly (INCOHERENCE\_IMPL) by adopting the same behaviours as in the case of H\_FAIL, or explicitly (INCOHERENCE\_RAW).

H: Investing in risk-free instruments like stocks is one of the ...

T: ...that is nonsense.

T can also explain the reasons she thinks this is not coherent (INCOHERENCE\_INTERP):

H: I will visit you on Sunday and then ...

T: ...but you are supposed to be traveling by then!

Finally, after an incoherence is detected, instead of waiting for H to correct the mistake, T can take the lead and propose another solution (or another topic) that seems more relevant for her:

H: I think we should call him and...

T: No, we cannot call him because he left his phone here. Do you have his email address, I am going to write a quick message for him.

**H\_INCOMPLETE:** H's utterance is still incomplete (and H is still holding the floor) but all the information given so far is coherent. T can perform a backchannel by nodding her head for example or by saying *Aha* or *Ok* for example (BACKCHANNEL). This gives H a signal that he is being understood and followed, thus encouraging him to keep on speaking. T can also choose to repeat a part of H's sentence for confirmation (FEEDBACK\_RAW). If this part is correct, H continues to speak normally (or sometimes explicitly confirms by adding a *yes* to his sentence), otherwise he declares that he disagrees with T's feedback. Here is an illustration of this mechanism taken from ([Khouzaimi et al., 2014b](#)):

H: 01 45

T: 01 45

H: 65 79

T: 67 79

H: No, 65 79

T: Sorry, 65 79

H: 98

T: 98

H: ...

T: The dictated number is: 01 45 65 79 98. Is that correct?

H: Yes

Another kind of feedback is by adding some related information to H's incomplete utterance (FEEDBACK\_INTERP), for example:

H: I went to see the football game yesterday...

T: ...yeah, disappointing

H: ...with a friend, but we did not stay until the end.

Also, an element in H's partial utterance can make T react immediately and change the conversation topic (BARGE\_IN\_CHANGE):

H: We went to this new restaurant down the street and...

T: Oh yeah. I have heard about it. Is it true that they make the best tapas in town.

**H\_SUFFICIENT:** H has not finished talking, yet, all the information that T needs to answer has been conveyed. If H is listing a few options, T can perform a gesture meaning that she is interested in the last option uttered (REF\_IMPL). She can also do it explicitly (REF\_RAW) (see (El Asri et al., 2014)):

H: Maybe we can schedule an appointment on Monday afternoon?...Tuesday morning?...Wednesday afternoon?...

T: Ok. Fine.

T can also add comments related to her choice, once selecting an option (REF\_INTERP):

H: We have apple juice...tomato juice...

T: Oh Yeah! That is my favorite, plus, my doctor advised me to have some every day.

In the case of goal-oriented dialogues, H keeps talking even though he conveyed all the necessary information for T to formulate an answer. T can choose to interrupt him (BARGE\_IN\_RESP) thus making the dialogue shorter (this can be viewed as a rude move in some cases):

H: I want to book a six-person table tomorrow at 6 please, I hope it is possible since I have ...

T: Sure, no problem. Can I have your name please?

**H\_COMPLETE:** H has finished his utterance. If T thinks that some more information needs to be provided, she can perform a gesture or adopt a facial expression to communicate that (REKINDLE\_IMPL), by explicitly warning him about the problem (REKINDLE\_RAW) by saying *So? Please tell me more.* or by being more specific about the missing piece of information (REKINDLE\_INTERP): *Ok, and what is the destination of the flight?*. If all the necessary information has been provided by H, T can provide new information to make the dialogue progress (END\_POINT). The latter is the most intuitive TTP that people have in mind when trying to model turn-taking.

H: How many friends of yours are coming with us tomorrow?

T: Two, hopefully.

### 3.3 Discussion

This taxonomy is aimed to clarify the notion of turn-taking, hence making it easier to analyse such behaviours in the context of vocal dialogue systems. In human-human conversation, this translates into a rich set of behaviours that are depicted and classified given two criteria. Compared to existing classifications of turn-taking behaviours, the focus is on the semantic content of H's and T's utterances (and other cues like gestures and facial expressions) as well as the reasons that pushed T to take the floor given this information.

As far as replicating TTPs in human-machine interactions is concerned, a big part of research in incremental dialogue systems and turn-taking optimisation has mainly focused on endpoint detection (Raux and Eskenazi, 2008) and smooth turn-taking. Therefore, their main objective is to replicate the phenomenon labeled here as BARGE\_IN\_RESP. Some other studies focus on backchanneling and feedback, often neglecting the semantic part of the dialogue participants utterances and focusing exclusively on prosody and acoustic features (Baumann, 2008; Jonsdottir et al., 2008).

Beyond the fact that the taxonomy provided here is aimed to provide a more exhaustive list than existing classifications, it is also a tool to analyse the similarities and differences between the TTP. The identified TTP can be classified in five categories according to the type of situation in which they apply (referenced by different colors in Table 3.2):

1. Dialogue initialisation (grey)
2. Negative feedback (red)
3. Positive feedback (blue)
4. Reference (yellow)
5. Ordered complete dialogue acts (green)



In the following, each category is discussed separately. Before starting the analysis, let us briefly recall the four different levels of analysis used here (see Chapter 1 for a more complete review):

1. **Phonetic level:** The dialogue acts are considered as a sequence of sounds. This is what a Voice Activity Detection (VAD) module would detect.
2. **Illocutionary level:** The message that can be extracted from this sequence of sound is the focus here.
3. **Perlocutionary level:** Refers to the effect that the dialogue act is supposed to have on the listener.
4. **Reason behind the dialogue act:** The reason that pushed T to perform this dialogue act (this notion have been added here to the traditional dimensions in the philosophy of language, its relevance will be demonstrated during the following analysis).

The following analysis is synthesised in Table 3.3, 3.4 and 3.5.

### 3.3.1 Dialogue initialisation

Four TTPs constitute this category: FLOOR\_TAKING\_IMPL, FLOOR\_TAKING\_RAW, FLOOR\_TAKING\_INTERP and INIT\_DIALOGUE. They should be distinguished from REKINDLE TTPs as they take place at the very beginning of the dialogue or when the dialogue participants stopped to interact for a long while (so that it is legitimate to consider that they are engaging in a new interaction). Viewed as phonetic dialogue acts, FLOOR\_TAKING\_IMPL involves implicit gestures and short sounds whereas FLOOR\_TAKING\_RAW involves longer sentences. In the case of FLOOR\_TAKING\_INTERP and DIALOGUE\_INIT, utterances are even longer in general. However, in some cases, it is possible to inject new information in very short sentences, therefore the length of the sentence cannot be used as a criterion to distinguish between these TTPs, except between FLOOR\_TAKING\_IMPL and the rest. As an illustration, imagine that you observe people interacting using a language that is unknown to you. Imagine, that they are silent and suddenly one of them utters a short sound. In that case, it is not obvious whether he just called his interlocutor's name or whether he actually uttered some new piece of information.

Considering the illocutionary level, FLOOR\_TAKING\_IMPL and FLOOR\_TAKING\_RAW introduce no new information apart from the fact that T wants to take the floor (however, they are different at the phonetic level). FLOOR\_TAKING\_INTERP, on the other hand, adds new information but which is not relevant to the main topic of the conversation unlike INIT\_DIALOGUE. As a consequence, viewed as a perlocutionary act, INIT\_DIALOGUE plays a double role: making H aware that T is starting an interaction (shared with FLOOR\_TAKING\_IMPL) and adding new information at the same time. Finally, the reason why H performs FLOOR\_TAKING\_IMPL is the desire to start a new interaction. INIT\_DIALOGUE is also performed for the same reason but in ad-

dition it is also aimed to announce the conversation subject and/or to make H aware of some new piece of information.

### 3.3.2 Negative feedback

Negative feedback is communicated through one of the six following phenomena: `FAIL_IMPL`, `FAIL_RAW`, `FAIL_INTERP`, `INCOHERENCE_IMPL`, `INCOHERENCE_RAW` and `INCOHERENCE_INTERP`. They all suggest that both participants have to take a step back in order to clarify or to correct something in the dialogue. From a phonetic point of view, as in 3.3.1, there is no rigorous distinction between these TTPs (unless the implicit ones are only gestures or facial expressions), even though the implicit ones are generally shorter than the explicit ones, which in turn are generally shorter than the interpreted ones.

It is interesting to notice that, viewed as a phonetic and an illocutionary dialogue act, `FAIL_IMPL` and `INCOHERENCE_IMPL` are the same or at least extremely hard to distinguish. This is also true for `FAIL_INTERP` and `INCOHERENCE_INTERP`. These dialogue acts translate into exactly the same signal sent by T, and only the dialogue context makes it possible to separate them. At the perlocutionary level, they are quite similar as they are both making H stop and take a step back in the dialogue. However, they are different as in the case of `FAIL` TTPs, where the goal is to make H repeat the same sentence again (or rephrase it while keeping the same meaning), whereas in the case of `INCOHERENCE` TTPs, T wants H to change his sentence and its meaning because it is problematic.

Actually, the strong difference between `FAIL` and `INCOHERENCE` TTPs comes from the fourth level that has been added to the analysis: the motivation behind behaving as such. These kind of differences is what motivated adding this dimension of analysis. As said earlier, the behaviour can be identical between the two categories (frowning, or saying *What?* for example), but the core difference between them comes from the fact that what pushes T to perform a `FAIL` TTP is the fact that she does not understand what has been said by H so far, and she does not want to lose track of the conversation, whereas in the case of an `INCOHERENCE`, she feels the need to notify a problem.

### 3.3.3 Positive feedback

`BACKCHANNEL`, `FEEDBACK_RAW`, `FEEDBACK_INTERP` and `REKINDLE` are aimed to give H a positive feedback in the sense that, unlike negative feedback, he is encouraged to keep the floor and to keep injecting new information. `BACKCHANNEL` and `REKINDLE` are generally shorter from a phonetic point of view (they can be also be gestures) but they are different as `BACKCHANNEL` involves an overlap whereas `REKINDLE` is performed after H releases the floor. `FEEDBACK_RAW` and `FEEDBACK_INTERP` are usually longer but there is no difference between them at this level of analysis.

At the illocutionary level, BACKCHANNEL, FEEDBACK\_RAW and REKINDLE are all close to the dialogue act *I understand what you said so far, please continue* (with a few subtle differences, though). On the other hand, FEEDBACK\_INTERP is richer as new information is injected. At the perlocutionary level, T wants to have a double effect on H:

1. Reassure him that his message has been understood so far.
2. Encouraging him to go on and add more information.

Finally, when considering the reasons that pushed T to perform these TTPs, REKINDLE\_IMPL, REKINDLE\_RAW and REKINDLE\_INTERP are different since T is surprised that H's utterance has already stopped. As a consequence, she feels the urge to ask for more. The same distinction between IMPL, RAW and INTERP as in previous TTPs applies here.

### 3.3.4 Reference

REF\_IMPL, REF\_RAW and REF\_INTERP are the TTPs that constitute this group. The phonetic analysis does not provide interesting insights apart from the fact that REF\_IMPL can be a gesture or a shorter speech act than REF\_RAW and REF\_INTERP. This category is interesting from an illocutionary point of view as the message that T tries to send is not present in her utterance but in H's one.

From a perlocutionary perspective, these TTPs are aimed to make H stop and understand T's answer from his own sentence. Finally, what pushes T to act this way is to avoid repetition and be more efficient.

### 3.3.5 Ordered dialogue acts

In this last section, FAIL\_MOVE, INCOHERENCE\_MOVE, BARGE\_IN\_CHANGE, BARGE\_IN\_RESP and END\_POINT are discussed. The simplest way of viewing dialogue is by adopting the walkie-talkie paradigm. Time is shared between participants in a sequential way where each one of them takes the floor and then releases it for the other to speak. As described previously, this corresponds to the END\_POINT phenomenon. At the phonetic level, it is characterised by the absence of overlap (or very small overlaps) and even a gap most of the time. On the other hand, no gaps are involved in FAIL\_MOVE, INCOHERENCE\_MOVE, BARGE\_IN\_CHANGE and BARGE\_IN\_RESP and overlaps are frequently observed.

From the illocutionary and perlocutionary point of view, FAIL\_MOVE and INCOHERENCE\_MOVE, BARGE\_IN\_CHANGE on the one hand and BARGE\_IN\_RESP and END\_POINT on the other hand form two groups. In the first one, while interrupting H, T suggests to consider a new idea or a new topic therefore making H shift its attention towards it. In the second group, the same conversation topic is developed. The phenomena of the first group can be distinguished following the illocutionary dimension

since in the case of `FAIL_MOVE`, a misunderstanding is communicated whereas in `INCOHERENCE_MOVE`, T communicates that H's utterance is problematic and finally, none of these two problems is communicated during a `BARGE_IN_CHANGE`. As far as the second TTP group is concerned, there is no difference between its two composing phenomena at the illocutionary level, however, from the perlocutionary point of view, T tries to inject new information in both cases but `BARGE_IN_RESP` comes with the additional intent of making H stop talking. T is pushed to act as such whenever she thinks that she has enough information to start uttering her next dialogue act. Therefore, the motivation behind such a behaviour is to increase efficiency by suppressing an unnecessary part of H's utterance hence gaining time. However, in some situations, barge-in cannot be performed either because of real constraints (a real walkie-talkie conversation for example) or because of social codes (politeness, timing allowed during an official meeting or a hearing, etc.).

#### 3.3.6 Synthesis

In Tables 3.3, 3.4 and 3.4, the previous analysis is synthesised in the form of a table. A phonetic profile is associated with each phenomenon (it is not accurate nor it is always respected, it is only aimed to give a general idea about how the TTP takes place in time between H and T), as well as a description of the illocutionary and perlocutionary levels. The elements motivating each phenomenon also appear in the table (x and y are placeholders that are used to refer to the content of the dialogue acts).

## 3.4 Turn-taking phenomena in dialogue systems

In this section, the TTP that exist in traditional dialogue systems are first identified. Incremental dialogue systems can extend this list to the other TTP in the taxonomy, however, some of them are very complex to execute and some are unlikely to improve the dialogue efficiency. The objective of this section is to discuss each of the previous TTP in order to determine which of them are worth implementing in an incremental dialogue system.

`INIT_DIALOGUE` is a TTP that is involved in every dialogue system, including traditional ones. `FLOOR_TAKING` TTPs are of limited interest when it comes to task-oriented dialogue, therefore, they are not implemented in task-oriented dialogue systems. There are two ways of initialising the dialogue, the user initiative one (the user starts speaking first like this is the case for Siri for instance) and the system initiative way (the system delivers an initial prompt, when calling an IVR for example). `END_POINT` is also necessary for any kind of dialogue, however, the way the dialogue participants exchange turns is not always the same given the situation at hand. Humans are very good at detecting end of utterance clues beforehand, making them achieve smooth turn-taking. Traditional dialogue systems, on the other hand, rely on long enough silences as markers of end of turn. A research thread is dedicated to studying methods of reducing these silences by considering different clues (prosodic, lexical

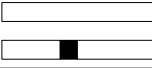
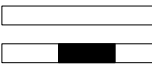
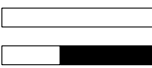
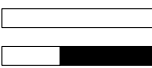




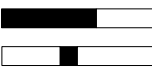

TTP	Phonetic act	Illocutionary act	Perlocutionary act	Motivations
FLOOR_TAKING_IMPL		I need your full attention	<ul style="list-style-type: none"> <li>Shift H's attention towards T</li> </ul>	<ul style="list-style-type: none"> <li>Desire to start a conversation</li> </ul>
FLOOR_TAKING_RAW		I need your full attention	<ul style="list-style-type: none"> <li>Shift H's attention towards T</li> </ul>	<ul style="list-style-type: none"> <li>Desire to start a conversation</li> </ul>
FLOOR_TAKING_INTERP		I need your full attention because of x	<ul style="list-style-type: none"> <li>Shift H's attention towards T</li> </ul>	<ul style="list-style-type: none"> <li>Desire to start a conversation</li> <li>More efficiency by providing a justification</li> </ul>
DIALOGUE_INIT		I start this conversation and I inform you that x	<ul style="list-style-type: none"> <li>Shift H's attention towards T</li> <li>Make H aware of the dialogue topic (x)</li> </ul>	<ul style="list-style-type: none"> <li>Desire to start a conversation about x</li> </ul>
FAIL_IMPL		I don't understand what you are talking about	<ul style="list-style-type: none"> <li>Make H stop</li> <li>Make H repeat or reformulate</li> </ul>	<ul style="list-style-type: none"> <li>Fix desynchronisation</li> </ul>
FAIL_RAW		I don't understand what you are talking about	<ul style="list-style-type: none"> <li>Make H stop</li> <li>Make H repeat or reformulate</li> </ul>	<ul style="list-style-type: none"> <li>Fix desynchronisation</li> </ul>
FAIL_INTERP		I don't understand what you are talking about because of x	<ul style="list-style-type: none"> <li>Make H stop</li> <li>Make H repeat or reformulate</li> <li>Making H aware of what is preventing T from understanding</li> </ul>	<ul style="list-style-type: none"> <li>Fix desynchronisation</li> <li>More efficiency by providing more precision about the problem</li> </ul>
FAIL_MOVE		I don't understand what you are talking about, consider x instead	<ul style="list-style-type: none"> <li>Make H stop</li> <li>Shift H's focus conversation towards x</li> </ul>	<ul style="list-style-type: none"> <li>Fix desynchronisation</li> <li>Moving the conversation forward</li> </ul>
INCOHERENCE_IMPL		What you just said is problematic	<ul style="list-style-type: none"> <li>Make H stop</li> <li>Make H reconsider what he just said</li> </ul>	<ul style="list-style-type: none"> <li>Fix desynchronisation</li> </ul>
INCOHERENCE_RAW		What you just said is problematic	<ul style="list-style-type: none"> <li>Make H stop</li> <li>Make H reconsider what he just said</li> </ul>	<ul style="list-style-type: none"> <li>Fix desynchronisation</li> </ul>

Table 3.3: Taxonomy labels (1/3)

### 3.4. Turn-taking phenomena in dialogue systems



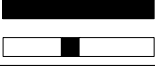




TTP	Phonetic act	Illocutionary act	Perlocutionary act	Motivations
INCOHERENCE_INTERP		What you just said is problematic because of x	<ul style="list-style-type: none"> <li>• Make H stop</li> <li>• Make H reconsider what he just said</li> <li>• Making H aware of the problem in his utterance</li> </ul>	<ul style="list-style-type: none"> <li>• Fix desynchronisation</li> <li>• More efficiency by providing more precision about the problem</li> </ul>
INCOHERENCE_MOVE		What you just said is problematic because of x, consider y instead	<ul style="list-style-type: none"> <li>• Make H stop</li> <li>• Make H reconsider what he just said</li> <li>• Making H aware of the problem in his utterance</li> <li>• Shift H's focus towards y</li> </ul>	<ul style="list-style-type: none"> <li>• Fix desynchronisation</li> <li>• More efficiency by providing more precision about the problem</li> <li>• Moving the conversation forward</li> </ul>
BACKCHANNEL		I understand (and sometimes: I agree)	<ul style="list-style-type: none"> <li>• Make H continue</li> </ul>	<ul style="list-style-type: none"> <li>• More information from H</li> </ul>
FEEDBACK_RAW		I understood that you said x	<ul style="list-style-type: none"> <li>• Make H continue</li> <li>• Make H correct in case of a misunderstanding or not</li> </ul>	<ul style="list-style-type: none"> <li>• More information from H</li> <li>• Desire to confirm that H's utterance has been well understood</li> </ul>
FEEDBACK_INTERP		I understood that you said x that is related to y	<ul style="list-style-type: none"> <li>• Make H continue</li> <li>• Make H correct in case of a misunderstanding or not</li> </ul>	<ul style="list-style-type: none"> <li>• More information from H</li> <li>• Desire to confirm that H's utterance has been well understood</li> <li>• Stronger confirmation by adding related information y</li> </ul>
BARGE_IN_CHANGE		I understood that you said x, tell me more about it	<ul style="list-style-type: none"> <li>• Shift H's focus towards x</li> </ul>	<ul style="list-style-type: none"> <li>• More information about x</li> </ul>
REF_IMPL		Yes, x	<ul style="list-style-type: none"> <li>• Make H stop and understand T is referring to the last element he uttered (x)</li> </ul>	<ul style="list-style-type: none"> <li>• Selecting an option</li> <li>• Less effort as x has already been uttered by H</li> </ul>

Table 3.4: Taxonomy labels (2/3)





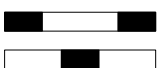

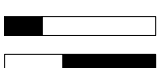
TTP	Phonetic act	Illocutionary act	Perlocutionary act	Motivations
REF_RAW		Yes, x	<ul style="list-style-type: none"> <li>• Make H stop and understand T is referring to the last element he uttered (x)</li> <li>• Make H correct in case of a misunderstanding</li> </ul>	<ul style="list-style-type: none"> <li>• Selecting an option</li> <li>• Less effort as x has already been uttered by H</li> <li>• Desire to confirm that the last option has been well understood</li> </ul>
REF_INTERP		Yes, x that is related to y	<ul style="list-style-type: none"> <li>• Make H stop and understand T is referring to the last element he uttered (x)</li> <li>• Make H correct in case of a misunderstanding</li> </ul>	<ul style="list-style-type: none"> <li>• Selecting an option</li> <li>• Less effort as x has already been uttered by H</li> <li>• Desire to confirm that the last option has been well understood</li> <li>• Stronger confirmation by adding related information y</li> </ul>
BARGE_IN_RESP		x	<ul style="list-style-type: none"> <li>• Make H aware of x</li> </ul>	<ul style="list-style-type: none"> <li>• Desire to move the dialogue forward</li> <li>• More efficiency because enough information has been provided before the end of the utterance</li> </ul>
REKINDLE_IMPL		The information you gave is not enough, please provide more information	<ul style="list-style-type: none"> <li>• Make H resume</li> </ul>	<ul style="list-style-type: none"> <li>• Fix desynchronisation</li> </ul>
REKINDLE_RAW		The information you gave is not enough, please provide more information	<ul style="list-style-type: none"> <li>• Make H resume</li> </ul>	<ul style="list-style-type: none"> <li>• Fix desynchronisation</li> </ul>
REKINDLE_INTERP		The information you gave is not enough, please provide information x	<ul style="list-style-type: none"> <li>• Make H provide information x</li> </ul>	<ul style="list-style-type: none"> <li>• Complete understanding of H's utterance</li> </ul>
END_POINT		x	<ul style="list-style-type: none"> <li>• Make H aware of x</li> </ul>	<ul style="list-style-type: none"> <li>• Desire to move the dialogue forward</li> </ul>

Table 3.5: Taxonomy labels (3/3)

and semantic) aiming for smoother turn exchange (Raux and Eskenazi, 2008; Gravano and Hirschberg, 2011). However, even though they have been under study for many years, there is still room for improvement. In the rest of this section, the remaining TTPs are discussed from an implementation point of view. Our goal is to come up with a list of TTPs that are the most likely to improve task-oriented dialogue. Similarly, the REKINDLE TTPs correspond to dialogue management strategies that can be implemented in any traditional dialogue system and as a consequence, it will not be considered here. Moreover, since this thesis focuses on task oriented dialogue, FAIL\_MOVE, INCOHERENCE\_MOVE and BARGE\_IN\_CHANGE are not discussed neither since they are aimed to change the discussion topic.

Before leading this discussion, it is important to notice that each TTP has two symmetric versions when it comes to human-machine dialogue: the one where H is the user and T is the machine and the opposite case. Here, the goal is to study system decisions therefore only one side is studied (even though BARGE\_IN\_RESP is implemented from both sides). In order for both cases to be implemented, the incremental dialogue system at hand should always be listening to the user, even though it has the floor (hence being able to be interrupted). As a technical side note, current incremental dialogue systems are used with a headphone for that reason: as the system keeps listening all the time, it is a convenient way to prevent it from hearing itself while speaking (considering its own sentence as a user input). In order to make them useful outside of labs, in more realistic situations, it is necessary to build algorithms that suppress the TTS result from the ASR input before feeding it to the latter (which raises a new challenge as it must also be done incrementally). Moreover, in this thesis, the focus is on the vocal modality. Therefore, no TTP based on gestures is considered for implementation.

Implementing a mechanism that mimics the FAIL TTPs is an interesting idea to explore. Users frequently use off-domain words and expressions (Chigi et al., 2014) and they are also often misunderstood by the ASR. As a consequence, making the system barge-in when it does not understand the user's partial utterance might have a positive impact on the dialogue efficiency. It might be less tiresome for the user as she wouldn't have to repeat her whole sentence several times. Moreover, this reduces the dialogue duration. FAIL\_RAW is not easy to replicate by a machine and it is most of the time performed with gestures and facial expressions at the same time. FAIL\_INTERP is not easy to implement either as giving the accurate reason why it did not manage to understand the user's utterance so far is not an obvious task. Implementing FAIL\_RAW, on the other hand, is much more realistic: when the system has no clue about the user's utterance after a long enough period of time, it simply declares that fact in a straightforward fashion.

In some cases, the user is likely to utter a sentence that is not coherent for the system or that is in contradiction with some data accessible by the latter (like when trying to buy a movie ticket when all the seats are already sold). By definition, the INCOHERENCE TTPs can help manage this case in a more efficient way. For the same reasons as FAIL\_RAW, INCOHERENCE\_RAW is not easy to implement. Unlike FAIL TTPs, it is not natural and more difficult to declare an incoherence without explaining the underlying reasons. Therefore, INCOHERENCE\_INTERP is more interesting TTP to



implement.

BACKCHANNEL has already been implemented in a few incremental dialogue systems (Meena et al., 2013; Hastie et al., 2013) with the aim of increasing its grounding capacities and its naturalness. This thesis focuses more on the efficiency aspect of incremental dialogue than the human-likeness side of the problem. These are somehow correlated, but as this is not a TTP that directly makes the dialogue more efficient (by preventing and fixing errors), it will not be implemented. Moreover, as the first step of the approach followed here is to use simulated dialogues, it is hard to evaluate this TTP in such conditions. On the contrary, FEEDBACK\_RAW provides a concrete opportunity to correct errors. When T tries to repeat a part of H's utterance and she succeeds, this gives H a proof that T heard his sentence (even though, this does not necessarily mean that T has correctly understood the message). If she fails, it is also interesting as H can repeat or reformulate his utterance, hence avoiding a desynchronisation. This is clearly interesting to be implemented in a dialogue system, yet, it can be very challenging. The involved turn-taking mechanism is difficult to manage in the sense that the user should not interpret the system's intervention as a barge-in hence being interrupted. Moreover, the system should be able to recognise whether the user ignored the feedback or tried to correct its content. Therefore, this TTP has been implemented in simulation only. Finally, FEEDBACK\_INTERP requires high natural language language processing capabilities and access to an important knowledge base. Moreover, it is more likely to complicate the dialogue and lead to errors compared to FEEDBACK\_RAW. Therefore, it has not been implemented here.

In (El Asri et al., 2014), REF\_RAW has been implemented from the user's point of view: the system enumerates a list of alternatives and the user barges-in to select one of them. This has been shown to significantly increase the dialogue quality (El Asri et al., 2014). However, implementing it requires changing the dialogue management strategy whereas, as discussed in Chapter 4, this thesis focuses on the impact of adding a turn-taking layer on top of pre-existing dialogue management strategies. As a consequence, REF\_IMPL, REF\_RAW and REF\_INTERP are discarded in our study.

Finally, BARGE\_IN\_RESP is clearly worth implementing from both sides. From the system's perspective, taking the floor as soon as it has enough information to do so can directly increase dialogue efficiency by reducing dialogue duration but also indirectly by preventing the user from adding new misleading information (Ghigi et al., 2014). From the user's point of view, being able to take the floor before the end of the system's utterance can make the dialogue less tiresome. This is especially true for users that are familiar with the system and as a consequence, they are able to predict the rest of the systems dialogue acts ahead of time.

To summarise, four TTP requiring incremental dialogue processing have been selected for rule-based<sup>3</sup> implementation (one of them from both sides: system and user):

- FAIL\_RAW (System side<sup>4</sup>)

---

<sup>3</sup>Interestingly, the reinforcement learning based strategy proposed in Chapter 8 leads to similar behaviours in an optimised fashion.

<sup>4</sup>System side means that the turn-taking decision is made by the system. In other words, T is the system.

- INCOHERENCE\_INTEPR (System side)
- FEEDBACK\_RAW (System side)
- BARGE\_IN\_RESP (System and user sides)

In Chapter 7, the details of the implementation, the rules chosen as well as a comparative study in a simulated environment are provided.

---

*User side* designates the opposite situation.



## Chapter 4

# Turn-taking decision module: the Scheduler

### 4.1 Description

#### 4.1.1 Overview

A new incremental dialogue system architecture is introduced in this chapter. The five modules forming the dialogue chain (see Chapter 1) are split in two groups: those forming the *client* and those constituting the *service*. The ASR and the TTS are necessarily included in the client and the DM in the service. The NLU and the NLG can fit in both categories. This terminology is borrowed from the computer network field (Israel and Mitchell, 1978) where the client can refer to the user and to the application that interacts directly with the user in order to gather useful data for the interaction at the same time. Similarly, the server refers to the application that is in charge of handling user's requests, as well as the remote machine it is deployed on. In the case of dialogue systems, both parts can be embedded in the same device and they can also be distributed across two different machines.

Viewing traditional dialogue systems from this point of view translates into a ping-pong game, where the client sends a request which is processed by the service, and the latter sends a response back. The question tackled here is how to break this rigid mechanism in order to make the system able to process the user's speech incrementally. This chapter shows how, by starting from this new view of dialogue systems instead of the sequential one (dialogue chain), an incremental dialogue system can be derived from a traditional one at minimal cost. In the resulting architecture, the turn-taking decision maker is separated from the DM allowing an autonomous control of the nature and timing of the incremental behaviour.

As illustrated in Fig. 4.1, a new interface is inserted between the client and the service (Khouzaimi et al., 2014b). This new module is called the *Scheduler* (this denomination is borrowed from (Laroche, 2010)). It can be deployed on the same machine

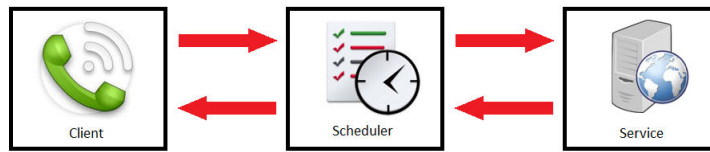


Figure 4.1: The Scheduler: an interface between the client and the service

as the client, as the service or in a dedicated server. The objective is to make the set {Scheduler+Service} behave like an incremental dialogue system from the client's point of view, without modifying the initial functioning of the service. Therefore, this provides a framework that can transform any dialogue system in its incremental version by adding a new layer.

The most classic approach of designing incremental dialogue systems consists in transforming each module (or only some of them depending on the situation at hand) into its incremental version (Schlangen and Skantze, 2011). The alternative approach presented here has the theoretical advantage of clearly separating turn-taking management from the rest. Turn-taking strategies are conceived and formalised independently from the task at hand: they can be reused as they are for different tasks. They can also be manipulated separately and combined in order to form new complex strategies given specific rules. As it will be seen along this thesis, turn-taking strategies will be implemented exclusively in the Scheduler (the rest of the system remaining the same). Our ultimate goal is to make this module learn optimal turn-taking behaviours by itself. Nevertheless, this approach is not fully incremental compared to the one described in (Schlangen and Skantze, 2011) as it will be shown later, which may lead to lower performances in case the DM involves heavy computational operations.

### 4.1.2 Time sharing

In traditional dialogue systems, time is shared in an ordered and clear manner. The dialogue is a simple sequence of turns  $T^1, T^2, \dots$  a turn being the time interval in which a user's utterance followed by the system's corresponding response takes place, or the opposite (depending whether the system adopts a user initiative or a system initiative strategy at each turn). For illustration and to simplify the notation, the system used here is supposed to belong to the first category, therefore, each turn is divided into two smaller time intervals, the user turn  $T^{k,U}$  and the system turn  $T^{k,S}$ :  $T^k = T^{k,U} \cup T^{k,S}$  (the union of the time intervals corresponding to consecutive user and system turns, Figure 4.2).

In this chapter, a few conditions are defined to precisely describe time allocation between the system and the user. The *activation time* of a condition refers to the exact moment when it goes from false to true. *EndTurnCond* is the condition that ends a user turn, it is generally assimilated to a long silence (Raux and Eskenazi, 2008; Włodarczak and Wagner, 2013).

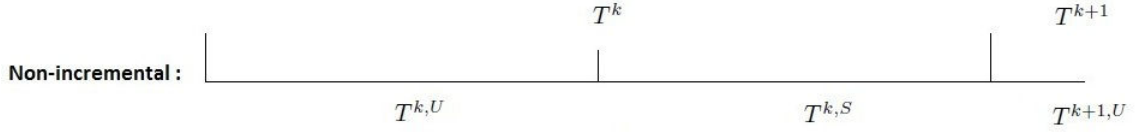


Figure 4.2: Time sharing in traditional settings

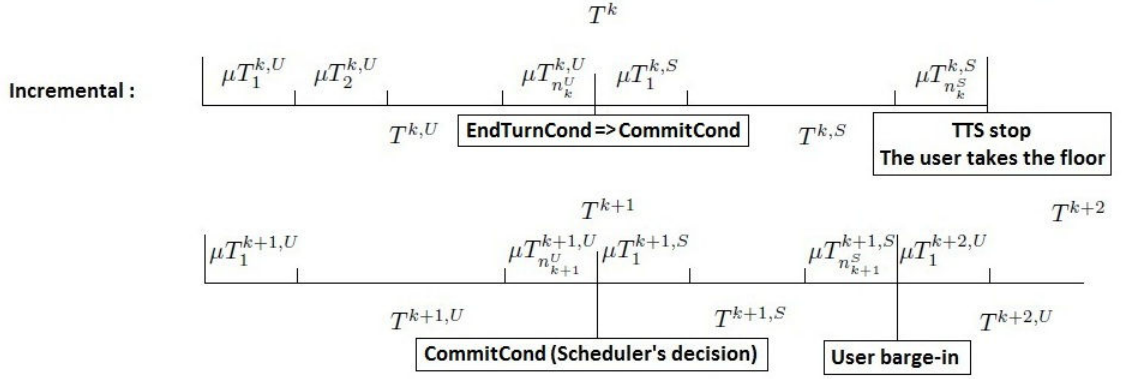


Figure 4.3: Time sharing in incremental settings

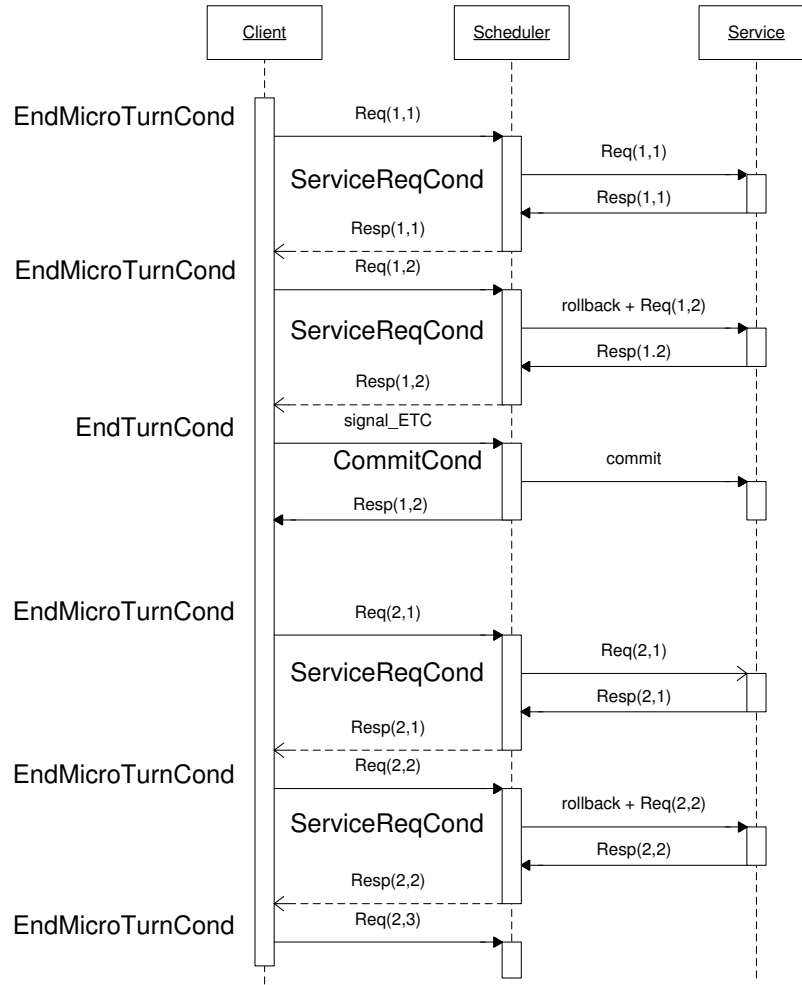
In incremental settings, this time sharing formalism does not hold anymore and a new condition should be defined: *EndMicroTurnCond* (with  $EndTurnCond \Rightarrow EndMicroTurnCond$ ). The time interval separating two activation times of *EndMicroTurnCond* is called a *micro-turn*. As a consequence, the turn  $T^{k,U}$  can be divided into  $n^{k,U}$  micro-turns  $\mu T_i^{k,U}$ :  $T^{k,U} = \bigcup_{i=1}^{n^{k,U}} \mu T_i^{k,U}$ . The  $p^{th}$  sub-turn of turn  $T^{k,U}$  is defined as  $T_p^{k,U} = \bigcup_{i=1}^p \mu T_i^{k,U}$  (Figure 4.3).

The request that the user makes during  $T^{k,U}$  is referred to as  $Req^k$  and the corresponding response is  $Resp^k$ . This architecture does not process incremental units like in (Schlangen and Skantze, 2011), instead, at each new micro-turn, it will take the whole information available since the beginning of the turn<sup>1</sup> (at the  $p^{th}$  micro-turn, all what the user uttered during  $T_p^{k,U}$ ). This *partial request* is called  $Req_p^k$ .

#### 4.1.3 The Scheduler

During the  $p^{th}$  micro-turn of the  $k^{th}$  user turn, the client sends  $Req_p^k$  to the Scheduler. The latter has to decide whether to send it to the service or not and the corresponding condition is called *ServiceReqCond*. A good example is  $ServiceReqCond \leftarrow (Req_p^k \neq Req_{p-1}^k)$  as sending the same brequst twice is useless. Then, the service provides the corresponding response  $Resp_p^k$  and the Scheduler stores it. The key idea of this architecture is that the Scheduler decides whether to retrieve this response to the

<sup>1</sup>This way of managing incremental dialogue is called *restart incremental* in (Schlangen and Skantze, 2011).



**Figure 4.4:** Incremental request processing with the Scheduler: the conditions on the left side trigger the Client to send partial requests

Turn	User sub-turn	Input	Real context	Simulation context
$T^1$	$T_1^{1,U}$	$Req_1^1$	$ctxt(T^0)$	$ctxt(T^0 + T_1^{1,U})$
	$T_2^{1,U}$	$Req_2^1$	$ctxt(T^0)$	$ctxt(T^0 + T_2^{1,U})$
	...	...	$ctxt(T^0)$	...
	$T_{n^{1,U}}^{1,U}$	$Req_{n^{1,U}}^1$	$ctxt(T^0)$	$ctxt(T^0 + T_{n^{1,U}}^{1,U})$
	<b>COMMIT:</b> $ctxt(T^1) = ctxt(T^0 + T_{n^{1,U}}^{1,U})$			
$T^2$	$T_1^{2,U}$	$Req_1^2$	$ctxt(T^1)$	$ctxt(T^1 + T_1^{2,U})$
	...	...	$ctxt(T^1)$	...

**Table 4.1:** A double context: the real context and the simulation context.

client (making it take the floor through the TTS) or not (waiting for more information to come from the client). This decision can also be forced by the client when sending an end of turn signal *signal\_ETC*, like a long enough silence for instance. It is important to note that the Scheduler is able to decide when to take the floor without waiting for *signal\_ETC*, the corresponding condition is called *CommitCond*. The Scheduler functioning over time is illustrated in Fig. 4.4 (dashed arrows represent the Scheduler's response to the client since they may or may not take place, given the Scheduler's decision to commit or not).

Incremental dialogue systems process growing utterances, and new coming information can complete or even modify the meaning of the current sentence. Therefore, they must be able to revoke the current hypothesis and create a new one when necessary (Schlangen and Skantze, 2011). Here, most of the requests that are made to the service are only aimed at seeing what would be its response for certain partial utterances and they are discarded right after. However, they might modify the dialogue state in the service which is a side effect to be avoided. As a consequence, two dialogue contexts are maintained:

- **The real context:** The dialogue context as traditionally used in dialogue systems. Contains the data and the variables that are aimed to last and be used in the rest of the dialogue.
- **The simulated context:** A copy of the real context, at the  $p^{th}$  micro-turn,  $Resp_p^k$  could be useful for the dialogue or not. Therefore, only this context is modified at the first place, the Scheduler decides later whether to keep the changes in the real context or not.

These dialogue contexts are managed by two actions performed by the Scheduler:

- **Commit:** The Scheduler commits to a partial request and the corresponding response when it decides to deliver the latter to the client, hence taking the floor immediately and not waiting for any further information. In that case, the simulated context is saved into the real context (thus becoming the new reference).
- **Rollback (or cancel):** The scheduler cancels the context changes when it decides to discard the last response obtained from the service and that a new - potentially more complete - partial request is received from the client (as shown in Figure 4.4). In that case, the real context is copied into the simulated one, rollbacking it to its original state.

The way the real and the simulated context are managed through the commit and the cancel actions is illustrated in Table 4.1.

## 4.2 Illustration

As a proof of concept, this section describes two instantiations of the previous abstract architecture, both in the case of a textual and a spoken dialogue system.



### 4.2.1 A textual dialogue system: CFAsT



**Figure 4.5:** The incremental version of the CFAsT project. The traditional view is represented on the left and the new incremental one is depicted on the right.

CFAsT stands for Content Finder Assistant. This application, developed at Orange Labs (Laroche, 2014, 2015), is aimed at automatically generating virtual assistants that help the user efficiently find specific content in databases (given as input). At each dialogue turn, the user provides some new information about his target and by using a keyword spotting algorithm, the system keeps narrowing the set of possibilities. The interface is made of a text box with a *validate* button. The dialogue service is deployed as a web service on a servlet container and the client is a javascript web page loaded on the user's browser.

In order to make it incremental, a Scheduler has been deployed as a servlet on the same container as the service and the javascript requests have been re-targeted towards it (the service is no longer directly visible by the client). Moreover, as discussed earlier, the dialogue context in the service have been duplicated. In this setup, *EndTurnCond* corresponds to the event of clicking on *validate* (or hitting the carriage key). A word-based micro-turn configuration has been implemented, therefore, *EndMicroTurnCond* is the action of hitting the space bar (when hit several consecutive times, only the first time is taken into account). Each request being different from the previous one, they should all be sent to the service, thus, *ServiceReqCond* is true all the time.

This implementation is aimed to provide a first proof of concept of the methodology introduced here. However, `signal_ETC` is equivalent to *CommitCond* which in turn is the same as *EndTurnCond* (the Scheduler never takes commit decisions by itself). The resulting behaviour presents a noteworthy similarity with autocompletion in search engines. The most straightforward way this mechanism is implemented is by continuously proposing several hypotheses regarding what would the complete request be, regarding the partial request typed so far. This does not require any request to the search engine itself (which can be viewed as the Service in the architecture introduced here). Another implementation (which lately started to be deployed) is by continuously updating the search result which is quite similar to what the transformed CFAsT achieves. By doing so, the system is more interactive and engaging for the user. It is also more efficient since the user does not have to determine when to stop typing the request since she can stop as soon as she spots interesting results.

#### 4.2.2 A spoken dialogue system: Dictanum

Another proof of concept of the proposed architecture has been implemented using speech. Inspired by (Skantze and Schlangen, 2009), it enables the user to perform number dictation. This task can be encountered in several kinds of applicative domains: phone numbers, credit cards,...Analysing human conversations in such situations shows that the number at hand is often broken into small pieces (a few digits each) and a confirmation follows each one of them. Therefore, the number is incrementally grounded so that in the case of a misunderstanding, only the last dictated part has to be repeated. Traditional dialogue systems behave differently as they ask the user to enter the whole number at once. This translates into long and tedious dialogues as only one minor mistake is enough to make the user repeat the number from the beginning, often for several times. For example:

SYSTEM: Welcome to the Dictanum service. Please dictate your number.

USER: 01 45 65 79 98.

SYSTEM: The dictated number is 01 45 67 79 98. Is that correct?

USER: No.

SYSTEM: Sorry. Let us try again. Please dictate your number.

USER: 01 45 65 79 98.

SYSTEM: The dictated number is 01 47 65 79 98. Is that correct?

USER: No.

SYSTEM: Sorry. Let us try again. Please dictate your number.

[...] (*given that the user is patient enough not to hang up*)

SYSTEM: The dictated number is 01 45 65 79 98. Is that correct?

USER: Yes.

SYSTEM: Thank you for using Dictanum. Bye.

With incremental dialogue processing, it is possible to design a strategy that is closer to the human behaviour. To show that, a number dictation dialogue system called Dictanum (described in (Khouzaimi et al., 2014a)) has been implemented using Disserto. It performs turn-based dialogue hence asking the user for the whole number at once. Here, the client is a web page displayed on the browser that uses Google API<sup>2</sup> for both the ASR and the TTS tasks and the service is a servlet deployed on a servlet container. Solutions like Kaldi (Povey et al., 2011) and their adaptation to incremental processing (Plátek and Jurčiček, 2014) offer more flexibility and can be better tuned according to the task at hand, nevertheless, this requires building one's own acoustic and language models. Since ASR is not the main focus here, Google API has been used as a good off-the-shelf solution which provides low recognition latencies and low word error rates across a very large vocabulary.

Afterwards, the approach introduced here has been used to build the incremental version of the system (like in the case of the CFAsT application, the Scheduler has been deployed as a servlet on the same container as the service). To do so, two silence duration thresholds have been defined: the short silence threshold  $\delta_s$  and the long one  $\Delta_s$  ( $\delta_s \leq \Delta_s$ ). *EndMicroTurnCond* is triggered when a short silence is detected and similarly, *EndTurnCond* corresponds to long silences. A feedback is delivered to the user in the case of a short silence (repeating the last 4 digits). If the user ignores the feedback and keeps dictating his number, the system keeps on adding digits to its list, however, if the user starts his next utterance with *No*, the feedback content is deleted from the number. Here is a dialogue example:

SYSTEM: Welcome to the Dictanum service. Please dictate your number.

USER: 01 45 (*short silence*)

SYSTEM: 01 45

USER: 65 79 (*short silence*)

SYSTEM: 67 79

USER: No, 65 79 (*long silence*)

SYSTEM: Sorry, 65 79

USER: 98 (*short silence*)

SYSTEM: 98

USER: ... (*long silence*)

SYSTEM: The dictated number is 01 45 65 79 98. Is that correct?

USER: Yes.

SYSTEM: Thank you for using Dictanum. Bye.

---

<sup>2</sup><https://cloud.google.com/speech/>

It is interesting to notice that if  $\delta_s = \Delta_s$ , the user is likely to dictate his number in one shot, never waiting for a feedback. Therefore, moving  $\delta_s$  between 0 and  $\Delta_s$  creates a continuum between the traditional and the incremental version.

Dictanum also offers the possibility for the user to interrupt the system during the final feedback, in order to make local corrections. To do that, this feedback is sent to the TTS in the following format: *The dictated number is 01 <sep> 45 <sep> 65 <sep> 79 <sep> 98. Is that correct?*. The latter pronounces the sentence chunk after chunk (chunks are delimited using the separator <sep>), each chunk lasting for the same number of micro-turns. This leads to the following kind of strategy (<sp> designates a short pause):

SYSTEM: The dictated number is: 01 <sp> 45 <sp> 67...

USER: No, 65.

SYSTEM: Sorry. The dictated number is 01 <sp> 45 <sp> 65 <sp> 79 <sp> 98. Is that right?

USER: Yes.

SYSTEM: Thank you for using Dictanum. Bye.

To conclude, this section presents two situations where the architecture introduced in this chapter have been implemented. They play the role of a proof of concept both for textual and spoken dialogue systems. In the following, a more theoretical discussion is led.

## 4.3 Discussion

### 4.3.1 Levels of incrementality

It appears that dialogue systems can be classified into four categories given the way they integrate incremental behaviour. The first category is made of traditional systems (Laroche et al., 2011). Then comes the second category where traditional systems locally implements a few incremental behaviours. For instance, in (El Asri et al., 2014), the system enumerates a list of options and the user selects the one that fits him best by uttering *Yes* or *Ok* for example (REF\_RAW in the taxonomy introduced in Chap. 3). The architecture introduced in this thesis belongs to the third category where incremental behaviour is obtained based on modules that are innately non-incremental (the service in our case). Other examples are described in (Selfridge et al., 2012) and (Hastie et al., 2013). Finally, the fourth category is made of incremental dialogue systems that are constituted of fully-incremental modules. In (Schlangen and Skantze, 2011), an abstract model for incremental architectures is presented where all the categories can fit, but the work that has been pursued by the authors and their research groups later on goes along with the spirit of this last category.

Categories 2, 3 and 4 embed different features related to incremental behaviour (summarised in Fig. 4.2):

- **TTS interruption after input analysis:** The user has the ability to interrupt the system (BARGE\_IN\_RESP from the user's side) but also to perform brief feedback (BACKCHANNEL or FEEDBACK TTP from the user's side) without interrupting the system's sentence. All categories except the first one can easily embed this feature.
- **Link interruption time with TTS:** Useful for simulating REF TTP (REF\_IMPL, REF\_RAW and REF\_INTERP from the user's side). This has been successfully implemented in a system that belongs to the second category in (El Asri et al., 2014). Therefore, it can also be implemented in systems with a higher degree of incrementality (categories 3 and 4).
- **User interruption by the system:** As it will be shown in the rest of this thesis, interrupting the user can improve the dialogue efficiency in some cases (BARGE\_IN\_RESP from the system's side). To do so, the system at hand must offer real incremental capacities which is the case for categories 3 and 4 only.
- **Better reactivity:** One of the main advantages of incremental processing is delivering responses in a quicker fashion since the processing of the user's request starts earlier. Again, real incremental abilities are required which makes it a property that is specific to categories 3 and 4 exclusively. This feature is particularly useful to achieve accurate end point detection (END\_POINT from the system's side).
- **Optimal processing cost:** The third category processes the user's request in a restart incremental way (sending the whole partial utterance at each new micro-turn). This is not optimal as it is possible to process it chunk by chunk. Therefore, this is an advantage that category 4 offers over all the others.

### 4.3.2 Enhancing a traditional dialogue system's turn-taking abilities at a low cost

Adopting the sequential paradigm described in Chapter 1 is a natural way of designing incremental dialogue systems (Schlangen and Skantze, 2011). The dialogue chain is kept unchanged, however, a substantial amount of work has to be done in order to design an incremental version of each one of the modules. The approach introduced in this chapter makes it possible to build an incremental dialogue system starting from a traditional one instead of starting from scratch. Therefore, the development cost is significantly reduced and moreover, the resulting incremental dialogue system benefits from all the experience and the adjustments embedded in the original dialogue system. In the following, the differences between the two approaches are reviewed as well as the elements that are simplified with the new approach and the price one has to pay to adopt it.

Incremental ASR is a prerequisite for the implementation of a Scheduler-based architecture. Therefore, it is not simplified by this approach and an inaccurate, slow or unstable ASR module still hurts the dialogue quality in the same way. In terms of NLU,

two cases have to be distinguished: putting the Scheduler before the NLU or after it. In the first case, it does not make sense to use an incremental NLU as the Scheduler proceeds on a restart incremental fashion (sending the whole user's partial utterance at each micro-turn). However, in the second case, it is possible to benefit from the advantages of incremental NLU (forming concepts in a really incremental way resulting in a more efficient processing). In that case, the Scheduler receives a sequence of potentially unstable sets of concepts.

The core difference between both approaches resides in the DM task. In a full-incremental architecture, dialogue act and turn-taking decisions are intertwined. The DM receives the input concepts chunk by chunk, and at each micro-turn, the new information can be viewed as the continuity of what has been understood so far, or as a signal driving the DM to revoke the current hypothesis before taking a new action. In the Scheduler-based approach, as the restart incremental paradigm is adopted, the revoke mechanism is intrinsically implemented as it is performed beforehand (the ASR changing its best hypothesis is a case of revoke).

Incremental generation and synthesis (Baumann and Schlangen, 2013) bring other challenges that are beyond the scope of this thesis. Nevertheless, the Scheduler-based architecture could be extended in order to support these features. As discussed in Chapter 1, since overlaps are generally limited in dialogue, incremental NLG and TTS are interesting when some other source of information is involved, other than the user's speech. In a multi-modal situation for instance, the system can change its mind while it is uttering a sentence based on some new data like a user's gesture. Another example is the case where a database access takes time and it is modified by a third agent while the system speaks; this could lead to a system utterance like *Well...let's see...it costs 100\$...oh, no sorry, it is even cheaper, 80\$ only....* In (Baumann and Schlangen, 2013), the system is observing a car trajectory and uttering comments as it moves; once a turn is expected, it starts announcing it but it has to wait until the car actually turns before specifying the direction (left/right): *The car drives along Main Street and then turns <hesitation> right.* In this case, the Scheduler should be given the capability to be triggered not only by new ASR results, but it should react to new independent DM outputs also.

Features	Category 1	Category 2	Category 3	Category 4
TTS interruption after input analysis	-	+	+	+
Link interruption time with TTS	-	+	+	+
User interruption by the system	-	-	+	+
Better reactivity	-	-	+	+
Optimal processing cost	-	-	-	+

**Table 4.2:** Available features for dialogue systems given the way they integrate incrementality

In summary, a new architecture has been presented in this chapter. It makes it possible to easily transform an existing traditional dialogue system into an incremental one. Also, it comes with the advantage of clearly separating traditional dialogue management from turn-taking management. This architecture will be used in the rest of this thesis with the objective of designing strategies that will be embedded in the Scheduler in order for it to make optimal turn-taking decisions.



## Chapter 5

# Dialogue strategies

This chapter sheds light on the most common way task-oriented dialogue systems manipulate concepts: slot-filling. Then several strategies that are used to get the right information from the user are presented and discussed. Finally, based on the TTP taxonomy from Chapter 3, an abstract turn-taking strategy is introduced. It will be instantiated later on in the thesis.

### 5.1 Utterance representation

When interacting with a dialogue system, the user's utterance can be represented in different manners. A simple way of encoding the different chunks of information that it contains is by using a slot representation. For example, if the user says *I would like to buy the book Dracula by Bram Stoker*, the NLU can output the following matrix:

ACTION:	Buy
OBJECT:	Book
TITLE:	Dracula
WRITER:	Bram Stoker

Each entry is called a *slot* and it represents a single concept in the utterance. It has two attributes: the *slot name* (on the left) and the *slot value* (on the right). Of course, depending on the application at hand, this matrix could be different; some slots could be considered as non relevant hence being discarded, others could be combined and new slots could be added. Therefore, while designing a dialogue system, a representation is chosen and one has to stick to it. This representation is widely used when it comes to dialogue systems design since it is a simple and natural way of representing information. It is well adapted to most tasks where the dialogue system plays the role of an interface between a user and a database. The system expects a user's request with predefined slots and depending on the latter and the database content, a response is



computed and returned. This is the representation used in this thesis and Section 5.2 defines it more formally.

## 5.2 Elementary tasks

A *task* is associated with an objective that the user wants to achieve. For example:

- Checking a bank account
- Scheduling an appointment
- Finding a restaurant nearby
- Booking a hotel room
- Modifying a flight reservation

Some tasks involve retrieving information from a database and others make the system modify the outside world (database modification, robot movement...). Moreover, a task can involve several *elementary tasks*. An elementary task is defined as an atomic action performed by the system, after the user has made a request with all the necessary information. The task of scheduling an appointment could involve several elementary tasks, for example:

1. Proposing a first time window (refused by the system).
2. Refusing a second time window that is proposed by the system.
3. Proposing a third time window (accepted by the system).

The way a task is organised in elementary tasks might affect the efficiency of its execution. However, this is very dependent on the task at hand and as a consequence, it is out of the scope of this thesis. The latter focuses on the way elementary tasks are handled: the optimisation of elementary tasks using different dialogue strategies.

To complete an elementary task, the system requires information that can be represented as a slot matrix:

$$\begin{bmatrix} slot_1 & x_1 \\ slot_2 & x_2 \\ \dots & \dots \\ slot_n & x_n \end{bmatrix}$$

In this framework, a dialogue system is viewed as an automaton that is able to perform different types of elementary tasks. Hence, a dialogue is a sequence of elementary tasks:  $ET_1, \dots, ET_k, \dots$ . More precisely, before performing  $ET_k$ , the system has to understand the user's request  $X_k = (x_1, x_2, \dots)$  which is a vector containing the slot values provided in her utterance (the size of the vector depends on the type of elementary task). Let  $C_k$  be the dialogue context when it is time to perform  $ET_k$ , then

$ET_k = f(X_k, C_k)$ ,  $f$  being defined in the dialogue system.  $(X_k, C_k)$  is commonly referred to as the *dialogue state*.

Moreover, a distinction can be made between two kinds of slots:

- **Constrained slots:** Some user's inputs are considered valid (but not necessarily correct) and others are not. If the slot at hand is a date and the user responds something which is not a date, this response is not valid in which case it is sure that the slot value is wrong. However, the validity of the response does not guarantee its correctness but it is noteworthy that if a generic ASR with an open grammar domain is used, valid responses that are incorrect are quite rare. On the other hand, if the ASR uses a grammar that is specific to the task at hand, they are more likely to happen.
- **Open slots:** Every input from the user can be considered as a valid value. For example, if the user is asked to utter a message that will be sent to some of his friends later on, he can utter anything. Therefore, unless the system asks for a confirmation, it cannot determine whether the user's input is right or wrong.

The way these slots are communicated depends on the dialogue strategy, the way the requests are formulated, the noise level and the NLU's ability to recognise a large variety of words and expressions. In Section 5.3, three different slot-filling strategies are introduced and discussed.

## 5.3 Slot-filling strategies

Depending on the task, the dialogue system and the user, slots can be filled in different ways in order to complete an elementary task. Three generic strategies are presented here but before delving into that, the notion of *initiative* is introduced. Consider the following dialogue between a customer that has just arrived to a hotel and the receptionist:

CUSTOMER: Hi. I would like to book a room for tonight please.

RECEPTIONIST: Sure. Do you need a parking lot?

CUSTOMER: No. I came here by train.

RECEPTIONIST: All right. Would you like a smoking room?

CUSTOMER: No, I don't smoke.

RECEPTIONIST: Ok. What about breakfast?

CUSTOMER: Yes, please. I will have it here.

RECEPTIONIST: Great, all set then! Let me get your key...

CUSTOMER: Thanks. When should I check out tomorrow?

RECEPTIONIST: Checkout is before 11:30am.

CUSTOMER: Ok. My train is leaving at 6:00pm, would it be possible to leave my bag here and get it back by then?

RECEPTIONIST: Absolutely sir. No problem.

This dialogue can be split into three phases. First, the customer starts the conversation by making a request. It is a result of his own initiative and he is setting the subject of the conversation. Then the receptionist provides an answer and takes the initiative right after by starting to ask specific questions. When, all the necessary information for the reservation is provided, the customer takes the initiative again to get some additional clarifications.

Such a distinction can also be made in the case of dialogue systems (Ferguson and Allen, 2007). When using a *system initiative* strategy, the latter makes requests and asks questions that the user should respond to in order to move the dialogue forward. Symmetrical strategies are called *user initiative* (the user leads the course of the dialogue). Finally, strategies that involve both dialogue modes are called *mixed initiative* strategies. These three dialogue strategies are presented and discussed in the following.

### 5.3.1 System initiative strategies

System initiative strategies can be compared to form-filling. The user is asked to fill several slots in a progressive way. Slot values can be asked for one by one or subset by subset. However, the *one by one* case is the most common and the most simple, therefore, it will be implied when talking about system initiative strategies in the rest of this thesis.

Formally, consider an elementary task  $ET$  that involves  $n$  slots:  $slot_1, \dots, slot_n$ . Completing  $ET$  using the system initiative strategies consists on a dialogue with  $n$  system questions (dialogue acts) called  $ASK(slot_1), \dots, ASK(slot_n)$  followed by the  $n$  corresponding answers containing the required slot values  $x_1, \dots, x_n$ . However, the way errors are managed is different whether the slot at hand is open or constrained. First, it is important to note that in this thesis, the ASR is supposed to run using an open grammar (which means that most of the words in the ASR language can be recognised). So, in the case of a constrained slot, errors are easier to spot since they generally generate an invalid slot value (an information that is not of the expected type, e.g. a date when expecting a time window, a name when expecting a number, etc.). For example, if the system asks for a date and the user responds *May 4<sup>th</sup>* but an ASR error occurs, then this response is more likely to be understood as some utterance which has nothing to do with a date like *Mayflower* or *Make lower* instead of another valid date like *May 5<sup>th</sup>*. Therefore, errors are easy to spot in the case of a constrained slot (assuming that an open grammar is used for the ASR). On the other hand, as far as open slots are concerned (small message dictation, event description...), once the user provides an answer, the system has no mean of checking whether an error occurred or not (since every input is valid). As a consequence, after the user provides an open slot  $slot_k$ , the system immediately asks for a confirmation  $CONFIRM(slot_k)$  (relative to that slot only: *Did you say <slot>?*). This is not the case for constrained slots and the system moves to the

next question as soon as a valid answer has been provided (the only confirmation is the final one, concerning the whole elementary task).

The example below shows a subdialogue corresponding to an elementary task with three slots,  $slot_1$  and  $slot_3$  are constrained whereas  $slot_2$  is open (*Confirmation message* refers to the last confirmation request made by the system once it collected all the necessary slot values, like *Ok, so you want to book a non-smoking room for 2 nights starting from tomorrow* or *You want to schedule an appointment with John on March 3<sup>rd</sup> at 2pm. Is that right?* for example):

SYSTEM:  $ASK(slot_1)$

USER: <noise>

SYSTEM: Sorry, I don't understand.  $ASK(slot_1)$

USER:  $x_1$

SYSTEM:  $ASK(slot_2)$

USER:  $\tilde{x}_2$  (altered version of  $x_2$  due to ASR imperfections)

SYSTEM: Sorry.  $ASK(slot_2)$

USER:  $x_2$

SYSTEM:  $ASK(slot_3)$

USER:  $x_3$

SYSTEM: Confirmation message

USER: Yes

### 5.3.2 User initiative strategies

In this thesis, *user initiative* refers to the following strategy: in order to complete an elementary task  $ET$  (involving  $n$  slots  $slot_1, \dots, slot_n$ ), the user is supposed to provide all the slot values in a complete utterance. If there are missing slots in his request, he is asked to repeat (or reformulate) it. The dialogue then looks like this:

SYSTEM: What can I do for you?

USER:  $x_1, <noise>, \dots, x_n$

SYSTEM: Sorry, I don't understand. What can I do for you?

USER:  $x_1, x_2, \dots, x_n$

SYSTEM: Confirmation message

### 5.3.3 Mixed initiative strategies

In noisy environments, the user initiative strategy can be very tiring to the user, especially when the number of slots is important. Another way to deal with incomplete requests is to switch to the system initiative strategy to gather the missing slots, which is somehow similar to the strategy described in (Lamel et al., 2000). Suppose that the elementary task at hand  $ET$  involves  $n = 5$  slots:  $slot_1, \dots, slot_5$ . A mixed initiative strategy dialogue looks like the following:

SYSTEM: What can I do for you?

USER:  $x_1, <\text{noise}>, x_3, <\text{noise}>, x_n$

SYSTEM:  $ASK(slot_2)$

USER:  $x_2$

SYSTEM:  $ASK(slot_4)$

USER:  $x_4$

SYSTEM: Confirmation message

As a side note, one can notice how the initiative strategy influences the way the information is grounded. The whole request is always grounded at the end but the system initiative and potentially the mixed initiative strategies perform separate slot grounding.

### 5.3.4 First efficiency comparison

Let  $N$  be the number of turns in the dialogue that are performed to complete the elementary task  $ET$  (recall that a dialogue turn is made of a system and a user turn). Also, suppose that for  $ET$  to be completed,  $n_s$  slots have to be specified. The objective of this preliminary study is to make a rough comparison between the previous strategies in a simple fashion, therefore, the following simplifying assumptions are made:

- All the slots have the same probability of not being understood:  $p_{err}$ .
- The fixed dialogue turns dedicated to greeting or saying bye for example are irrelevant for the comparison and are therefore neglected.
- Only constrained slots are considered (open slots are more rare and necessitate a special treatment).
- As a consequence, errors are mostly due to invalid input (asked for again right after the request), so the answer to the final confirmation is considered to be always positive.

**System initiative:** The  $i^{th}$  slot requires  $N_i = 1 + n_{err}^i$  dialogue turns,  $n_{err}^i$  being the number of errors that occurred before the system considers that the slot value has been understood. Therefore,  $N_i$  follows a geometric distribution with parameter  $1 - p_{err}$ :

$$\mathbb{P}(N_i = k) = p_{err}^{k-1}(1 - p_{err}), \forall k \in \mathbb{N}^* \quad (5.1)$$

As a consequence, since  $N = \sum_{i=1}^{n_s} N_i$ :

$$\mathbb{E}[N] = \frac{n_s}{1 - p_{err}} \quad (5.2)$$

**User initiative:** Here,  $N = 1 + n_{err}$  but  $n_{err}$  corresponds to the number of user utterances where at least one slot has been misunderstood. Therefore,  $N$  follows a geometric distribution with parameter  $(1 - p_{err})^{n_s}$  which leads to:

$$\mathbb{E}[N] = \frac{1}{(1 - p_{err})^{n_s}} \quad (5.3)$$

**Mixed initiative:** Let  $n_{mis}$  be the number of missing (misunderstood) slots. Then by reasoning similarly as in the user initiative case,

$$\mathbb{E}[N|n_{mis}] = 1 + \frac{n_{mis}}{1 - p_{err}} \quad (5.4)$$

Moreover,  $n_{mis}$  follows a binomial distribution with parameters  $p_{err}$  and  $n_s$ :

$$\mathbb{P}(n_{mis} = k) = \binom{n_s}{k} p_{err}^k (1 - p_{err})^{n_s - k} \quad (5.5)$$

Consequently,

$$\mathbb{E}[N] = 1 + \frac{p_{err} n_s}{1 - p_{err}} \quad (5.6)$$

Figure 5.1 compares the efficiency of these three slot-filling strategies with five slots and under different noise conditions.  $\mathbb{E}[N]$  is used as a proxy for efficiency and  $p_{err}$  represents the level of noise. It appears that the system initiative strategy is less efficient when the noise level is low compared to the user initiative one. However, as the  $p_{err}$  grows, gathering the slots one by one appears to perform better. Finally, the mixed initiative strategy performs best both in low and high noise situations.

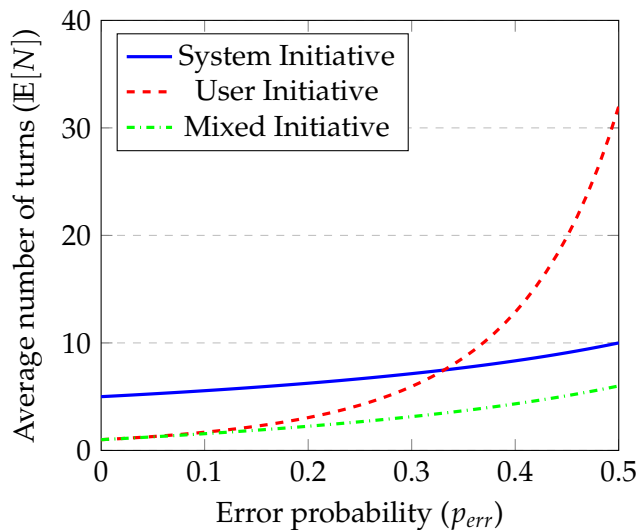


Figure 5.1: Slot-filling strategies efficiency comparison ( $n_s = 5$ )

## 5.4 Incremental strategies

Incremental dialogue processing brings a new dimension (a new degree of freedom in the decision) that can be exploited in order to improve the dialogue efficiency. In Chapter 3, the following TTP have been selected for implementation: FAIL\_RAW, IN-COHERENCE\_INTERP, FEEDBACK\_RAW and BARGE\_IN\_RESP from the user's and the system's perspective. In the following, the ways they can be implemented are presented.

The concept of elementary task has been proven to be useful for slot-filling strategies analysis. As far as incremental strategies are concerned, analysis is made at a more atomic level since only one dialogue turn is considered. In the following, the user is supposed to have the floor and the system is waiting for her to provide  $n_s$  slot. The DM either uses the user initiative or the mixed initiative strategy; in the system initiative strategy, the user's utterances are too short for incremental processing to be relevant.

The reader should keep in mind that the architecture used here is the one introduced in Chapter 4, therefore, the Scheduler module is in charge of making turn-taking decisions. Moreover, as discussed in Chapter 1, another important aspect to consider is *ASR instability* which is one of the major difficulties that incremental processing brings (recall that as the user speaks, his current partial utterance is not necessarily a prefix of the partial utterances to come). Nevertheless, early words in the user's utterance are more likely to stay unchanged than later ones (McGraw and Gruenstein, 2012). As a consequence, making the Scheduler take decisions based on the whole current user's utterance is risky since it is likely to change. Therefore, a *stability margin* (SM) is taken into account. It corresponds to the last part of the utterance that the Scheduler has to discard before making decisions. SM can be expressed in time units (discarding the last second for example) or a number of words or phonemes. The user's request without

SM is called the *last stable utterance*.

The Scheduler can perform three kinds of actions:

- **WAIT:** Performed most of the time, it is chosen when the Scheduler decides not to retrieve the service's response at a certain micro-turn, hence waiting for the user to provide more information during the following micro-turns.
- **SPEAK:** The Scheduler decides to commit to the current user's utterance and to provide the corresponding service's response right away. This either translates into a barge-in or an accurate end point detection.
- **REPEAT:** The Scheduler does not retrieve the last service's utterance but the last pronounced of the last user's stable utterance only<sup>1</sup>. The objective is not to barge-in, instead, the system performs a feedback that might trigger a reaction from the user, but not necessarily.

Based on these three actions, the TTP selected in Chapter 3 have been implemented as follows:

- **FAIL\_RAW:** If the user has been holding the floor for too long without providing a single slot value, it might be interesting for the system to interrupt her asking for a reformulation or a repeat. This situation can happen for two main reasons: the user uses off-domain words or expressions (mainly because she is not familiar with the system) or her utterance has been altered because of noise and ASR imperfections. The system can use several criteria in order to decide whether to interrupt the user or not. For example, it can rely on a time threshold: if the user speaks for a period that is larger than that threshold without providing any slot value, then it performs a SPEAK. This duration can be replaced by a number of words or phonemes for example. Consider the following dialogue where the user tries to check his bank account in a noisy environment:

USER: <noise> like to <noise> number 58 45...

SYSTEM: ...Sorry, I don't understand. What can I do for you?

USER: Check account number <noise> I repeat, check account 58 45 18 A.

SYSTEM: All right, you want to check the account number 58 45 18 A right?

USER: Yes.

The user is interrupted by the system at the first dialogue turn. If the latter didn't make such a decision, its response would have been the same (*Sorry, I don't understand. What can I do for you?*) since an important part of the request was lost. Therefore, the system barge-in spared the user some time and energy. In this example, the second time the user tries to formulate his request, he does it in a more concise way and repeats it as an effort to make himself clearer. This kind of behaviour has been noticed as a result of a corpus study led in (Ghigi et al., 2014).

<sup>1</sup>This is a simple way of simulating feedback. In reality, any part of the last stable utterance could be repeated.



- **INCOHERENCE\_INTERP:** Some user's requests can be problematic for the system because they are not coherent with the current dialogue context. The user might ask for a non existing information or try to perform some forbidden modification in the database for example. This can also be due to a transmission error (noise, ASR imperfection...) or simply because the user is not well aware of the dialogue context and the system's constraints. However, if an open grammar domain is used, the first reason is less likely to be the cause of the problem (a transmission error more often leads to a non understandable utterance by the system). Consider the following example:

USER: I would like to book a room for tomorrow and I will be ...

SYSTEM: ...Sorry, we are full tomorrow.

In this situation, it is legitimate to interrupt the user as the system is sure that the response would be the same even if it waits for the end of the request to take the floor. Implementing this behaviour is very dependent on the domain. Given the latter, a list of dialogue acts that indicate an incoherence should be identified and as soon as the service's response corresponding to the last stable utterance belongs to that list, a SPEAK should be performed.

- **FEEDBACK\_RAW:** Providing vocal feedback while the user is speaking can be very useful in some situations, especially when open slots are involved. For the sake of simplicity, only the last word of the last stable utterance is repeated in this thesis but one can go further and try to repeat any part of the user's current partial request. It is used in (Skantze and Schlangen, 2009; Khouzaimi et al., 2014b) in the case of numbers dictation (see Chapter 4 for an example).

However, implementing this TTP can be very challenging as the user must perceive that even though the system speaks out, it does not plan to take the floor. Moreover, negative reactions to a feedback like *No, I said 45 52* must be clearly separated from positive ones. For the reasons explained in Chapter 3, it will only be implemented in simulation in this thesis.

- **BARGE\_IN\_RESP (System):** When the user has provided all the necessary information for the system to provide an answer, the latter can decide to take the floor right away to improve the dialogue efficiency. The implementation of this behaviour is similar to INCOHERENCE\_INTERP. A list of dialogue acts has to be determined in advance and as soon as the service's response to the last stable utterance belongs to that list, the Scheduler decides to SPEAK.
- **BARGE\_IN\_RESP (User):** This TTP does not involve the Scheduler as the decision to speak is made by the user. However, it is important that the system does not allow user interruptions all the time, otherwise, none of the previous TTP implementations are possible. Suppose that the ASR is always listening and as soon as it detects a new input, if the system speaks, it releases the floor. Also, suppose that the user is speaking and for some reason (previous TTP), the system decides to interrupt him. During a little time window, both the user and the system are speaking and as the system is speaking it is interrupted right away. A simple so-

lution that is adopted in this thesis is to define a time interval during which the ASR is no longer listening after the system takes the floor. Of course, this is only the case when a SPEAK is performed (the REPEAT action is not treated as such).

To conclude, this chapter clarified the way concepts are manipulated in most current dialogue systems as well as a few DM strategies that will be used in the following. On top of that, a few turn-taking behaviours (based on the TTP taxonomy introduced in Chapter 3) and the way they can be implemented have been discussed. In the rest of this thesis, they will be instantiated in two different domains in order to form a new rule-based turn-taking strategy.



## Chapter 6

# Incremental dialogue simulation

Using dialogue simulation techniques is very common in the research community ([Eckert et al., 1997](#); [Pietquin and Dutoit, 2006](#)) for several reasons like: the ability to quickly generate dialogue corpora that can be used to develop machine learning techniques, an easy way to model different populations of users and the possibility to use the same user simulator to test and compare concurrent dialogue strategies (see Chapter 2 for more details). In this chapter, a new incremental dialogue simulation framework is introduced (published in [\(Khouzaimi et al., 2016a\)](#)). Its novelty resides in the fact that it is able to simulate the ASR instability phenomenon. First, it is presented in its most generic and abstract form that can be used by the reader to instantiate his/her own simulator that is adapted to any target domain. Then, these principles are applied in order to implement a showcase simulated environment where the service is a personal agenda manager.

Later on, this simulated environment is used for two main purposes. Firstly, the slot-filling and the incremental dialogue strategies described in Chapter 5 are implemented and compared. This somehow validates the preliminary efficiency analysis led in that chapter. It also provides new analysis elements to go further and prepare a basis for the experiments with real users. Secondly, it is a very useful tool for generating data to train machine learning algorithms. In Chapter 8, it is used to train a reinforcement learning algorithm which purpose is to optimise turn-taking decisions.

### 6.1 Overview

How to run dialogues with no users? The well-known answer is: by designing a User Simulator (US). Rigorously, in the case of SDSs, a US should be able to process an input audio signal and to output a new audio signal as well. Even though this method has its merits (noise and ASR imperfections are naturally taken into account), it goes against one of the main advantages of user simulation techniques which is the ability to quickly generate an important number of dialogues. Also, making an ASR module listen to a TTS and understand its message is not easy. Therefore, the user simulator

elaborated here inputs and outputs text. An ASR output simulator is in charge of replicating the ASR behaviour. Figure 6.1 gives an overview of how these parts fit together in the whole architecture as well as the composition of the US. The latter is composed of five modules: The Intent Manager, the NLU, the Verbosity Manager, the NLG and the Patience Manager.

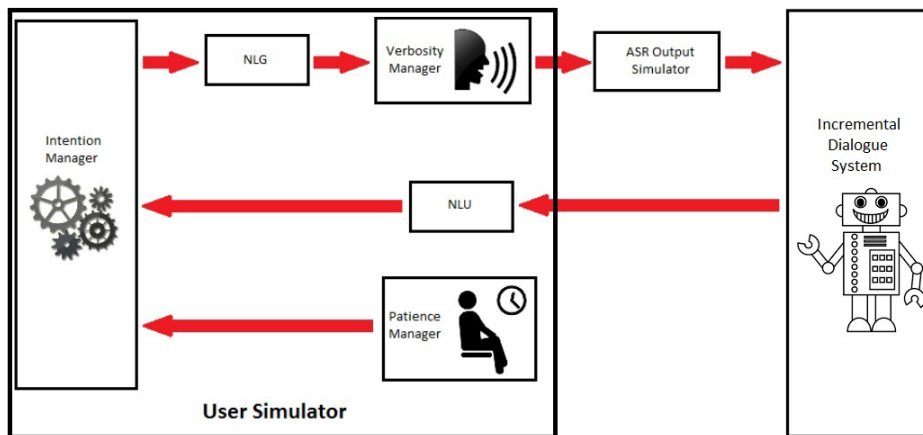


Figure 6.1: Simulated environment architecture

## 6.2 Incremental dialogue simulation

In a nutshell, at the  $p^{th}$  micro-turn of the  $k^{th}$  user turn  $\mu T_p^{k,U}$ , the US generates a partial utterance  $Req_p^k$  that is transformed into an N-Best  $(score_{p,1}^k, hyp_{p,1}^k), \dots, (score_{p,N}^k, hyp_{p,N}^k)$ , which corresponds to the N recognition hypotheses that have the best confidence scores. It corresponds to the whole utterance pronounced during the partial turn  $T_p^{k,U}$  (*restart incremental mode* (Schlangen and Skantze, 2011)). On the other hand, either the US receives an answer from the dialogue system at a certain micro-turn and it stops speaking<sup>1</sup>, either it does not and it continues speaking if it has additional things to say (releasing the floor otherwise). When the dialogue lasts for too long without achieving the task at hand, the US can end the dialogue.

In the following, the role and the functioning of the US and the ASR output simulator is described in an abstract fashion before being instantiated later on to give birth to a personal agenda management simulated environment.

<sup>1</sup>This is of course an approximation of real barge-in cases since the overlap is neglected.

### 6.2.1 User Simulator

#### Intent Manager

The Intent Manager is in charge of computing the dialogue acts that the US performs. It maintains an internal dialogue context and takes the dialogue acts coming from the dialogue system as inputs. Thus, it can be viewed as a dialogue manager in itself but with the difference that it is aimed to generate requests and to lead the dialogue instead of serving a user (at least in task-oriented situations). Therefore, it is given a task or a list of tasks to accomplish before it starts interacting with the dialogue system.

A common approach to design such a module is the agenda-based method (Wei and Rudnicky, 1999; Schatzmann et al., 2007). Inspired by the latter, the approach adopted in this thesis suggests that the tasks the Intent Manager should accomplish are given in the form of a stack (LIFO structure): the action stack (AS). They are removed and executed one by one and during each step, new actions could be added. The Algorithm 1 describes a function called *run* with AS as an argument and that is in charge of unstacking all the corresponding actions and executing them by using the method *perform()*. The latter tries to execute the top element of the action stack which might lead to the removal of the top element of the stack or the creation of new actions that are added on top of AS (which justifies the fact that the whole stack is passed as an argument and not the top element only). A loop is run over the elements of AS until it is empty.

```

Algorithm run (AS)
  while AS.size > 0 do
    | perform(AS);
  end

```

**Algorithm 1:** Intent Manager abstract algorithm

#### NLG and Verbosity Manager

The NLG module of the simulator transforms the Intent Manager’s output into a simple and straightforward utterance. For example:

- Book a room for tomorrow.
- Record channel 2 from 6pm until 8pm.
- Delete the event football game from the agenda.

Compared to human/human conversations, limiting interactions to this kind of simple utterances is not realistic. Therefore, they are enhanced in the Verbosity Manager with prefixes like *I would like to*, *Is it possible to...* and suffixes like *if possible, please...* In (Ghigi et al., 2014), a corpus study showed that users tend to go off-domain and to repeat the same information several times in the same sentence. These behaviours are also replicated in the Verbosity Manager: with a probability  $p_{od}$  the NLU output is replaced with an off-domain sentence randomly picked in a predefined list, moreover,

with a probability  $p_{rep}$  and given that the system just reported a misunderstanding, the utterance is repeated twice (for example, *Check my account, I repeat, check my account*).

### Timing and patience manager

When it comes to incremental processing, timing is key. However, the main objective of simulation is to generate dialogues as fast as possible, hence, real time stamps cannot be used. In order to approximate durations, the user's and the system's speech rates are considered to be constant with value  $SR$ .

Users tend to get impatient, at various degrees, when dialogue systems take too long to accomplish the task they are asked for. To simulate this behaviour, a duration threshold is chosen at each new dialogue that will cause the user to hangup as soon as it is reached. It is computed as follows

$$d_{pat} = 2\mu_{pat}.sigmoid(X) \quad (6.1)$$

where  $X$  follows a Gaussian distribution of mean 0 and variance 1 and  $\mu_{pat}$  is the mean duration since

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (6.2)$$

### 6.2.2 ASR output simulator

The ASR output simulator generates an N-Best that is updated at each new micro-turn. For instance, if at a certain point, the US uttered *I would like to add the event birthday party on...*, a possible N-Best could be (the numbers between brackets represent ASR scores):

- (0.82) I would like to add the event birthday party on
- (0.65) I like to add the event birthday party on
- (0.43) I have had the event birthday party
- (0.33) I would like to add the holiday party
- (0.31) I like to add the holiday party on

More formally, at the  $p^{th}$  micro-turn of the  $k^{th}$  user turn  $\mu T_p^{k,U}$ , the N-Best is an N-uplet  $(score_{p,1}^k, hyp_{p,1}^k), \dots, (score_{p,N}^k, hyp_{p,N}^k)$ . At time  $t+1$ , a new word  $w_{t+1}$  is sent to the ASR output simulator and the latter calculates a new associated N-Best. Therefore, at this stage, the system has two N-Bests:



Figure 6.2: ASR score sampling distribution ( $\sigma_{conf} = 1$ )

- **The word N-Best:** It corresponds to the different hypotheses related to the last word pronounced. In Figure 6.3, the top right box represents the word N-Best associated with the word 6<sup>th</sup>.
- **The utterance N-Best:** It designates the N-Best associated with the whole partial utterance pronounced so far. In Figure 6.3, the top left box is an example of such N-Best associated with the partial utterance *I want to add the event birthday party on January*.

Both are combined to form the new utterance N-Best. In the following, the way the word N-Best is calculated and the way it is incorporated into the partial utterance N-Best are described.

In order to simulate noise and ASR imperfections, the ASR output simulator uses a module called the Scrambler. It receives a word as input and performs one of the three following operations in order to compute the output<sup>2</sup>:

- Replace the word with a different word taken randomly from a dictionary (probability:  $p_{repl}$ )
- Add a new word (probability :  $p_{add}$ )
- Delete the word (probability :  $p_{del}$ )

A Word Error Rate (WER) is given as a parameter to the ASR output simulator. It controls the noise level that one wants to simulate. The algorithm used to generate the N-Best associated with a single word is described below:

1. Determine whether  $w_{t+1}$  is among the N-Best or not with a probability that is computed as follows:  $(1 - \text{WER}) + \text{INBF} \cdot \text{WER}$ , where INBF (In N-Best Factor) is a parameter between 0 and 1. If  $w_{t+1}$  is not in the N-Best, then the latter contains only scrambled versions of this word and jump to step 4.

<sup>2</sup>The Scrambler always performs one these three operations. In other words  $p_{repl} + p_{add} + p_{del} = 1$



2. The first hypothesis is set to be  $w_{t+1}$  with a probability of (1-WER), otherwise, it is a scrambled version of it.
3. If the the first hypothesis is not  $w_{t+1}$ , then this word's position is randomly chosen between 2 and N. Moreover, the other hypotheses are scrambled versions of it.
4. The confidence score associated with the best hypothesis ( $score_0$ ) is sampled as  $\text{sigmoid}(X)$  where  $X$  follows a Gaussian distribution. More precisely,  $X \sim \mathcal{N}(c_{err}, \sigma_{conf})$  if the first hypothesis is wrong and  $X \sim \mathcal{N}(c_{right}, \sigma_{conf})$  when it is right (with  $c_{err} < 0 < c_{right}$ ). By taking the sigmoid, this leads to two distributions<sup>3</sup> (depicted in Figure 6.2 for  $\sigma_{conf} = 1$ ) with a mean on both sides of 0.5 and the same standard deviation for both (which is a growing function of  $\sigma_{conf}$  and which can be changed to simulate different levels of accuracy of the confidence score model). Big  $\sigma_{conf}$  values lead to spread recognition scores and small differences between  $c_{err}$  and  $c_{right}$  engender close scores for both cases: right and wrong first hypothesis. Therefore, discriminative models are obtained for small values of  $\sigma_{conf}$  and high difference  $c_{right} - c_{err}$ .
5. The scores for the other hypotheses are computed in an iterative way. For  $i$  between 2 and N,  $score_i$  is uniformly sampled in  $[0, score_{i-1}]$ .

As already mentioned in this thesis, early partial utterances are not necessarily prefixes of later ones with a true ASR system (ASR instability phenomenon). To replicate this behaviour, a language model is needed to compute the scores corresponding to the different hypotheses in the N-Best. Therefore, sentences that are more in alignment with the model have higher scores thus being pushed to the top of this N-Best. Here, the NLU knowledge is used as a proxy for the language model by making the following assumption: *the more an utterance generates key concepts once fed to the NLU, the more it is likely to be the correct one*. Therefore, as soon as a new concept is detected in  $hyp_{p,i}^k$ ,  $score_{p,i}^k$  is boosted as follows:

$$score_{p,i}^k \leftarrow score_{p,i}^k + BF \cdot (1 - score_{p,i}^k)$$

where BF is the Boost Factor parameter. An illustration of this mechanism with BF=0.2 is provided in Fig. 6.3.

## 6.3 Personal Agenda management simulated environment

So far, an abstract framework for simulating incremental dialogues has been described. In this section, it is instantiated to be able to generate dialogues with an incremental dialogue system which purpose is to help users manage their personal agenda. This task will be used in Chapter 7 in order to run a few experiments regarding the dialogue

<sup>3</sup>Confidence score estimation is a complex problem and it is still a research topic (Jiang, 2005; Seigel and Woodland, 2011). The simple model introduced here is inspired by (Pietquin and Beaufort, 2005). Also, notice that the scores are between 0 and 1 but they do not sum up to 1 since they are not probabilities.

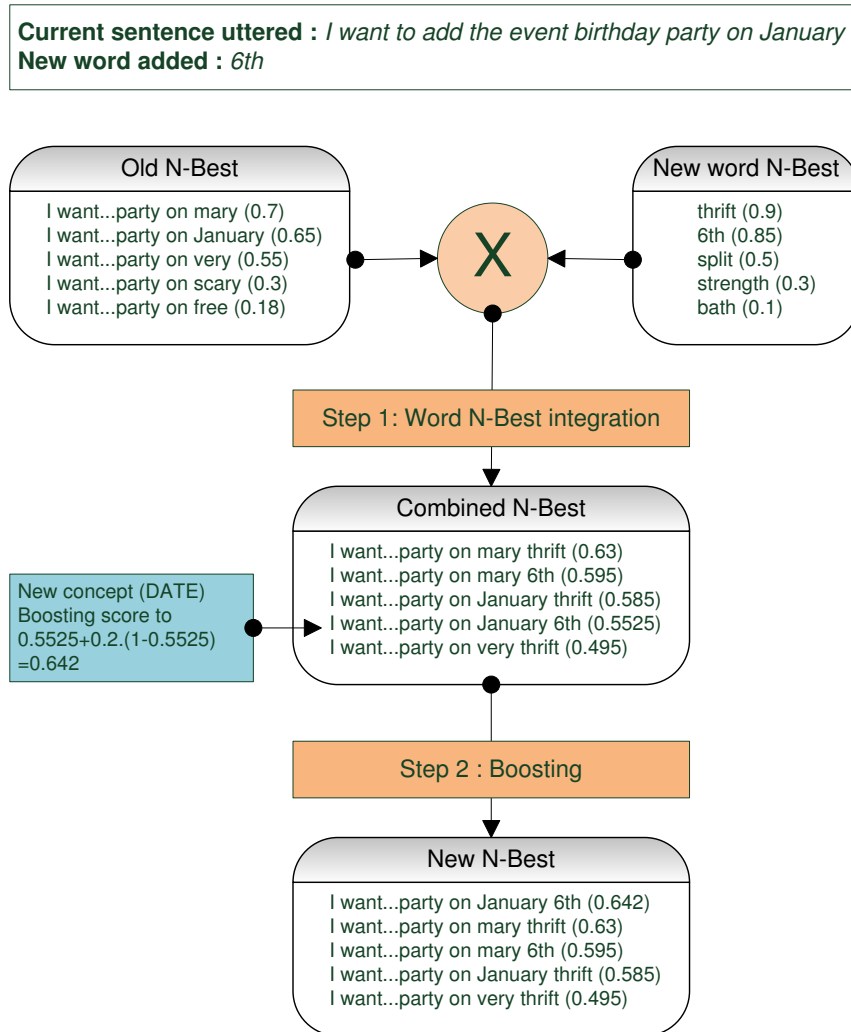


Figure 6.3: An illustration of the incremental ASR output N-Best update (BF=0.2)

strategies introduced in Chapter 5 and in Chapter 8 to test a new reinforcement learning approach. First the Service (system side, Section 6.3.1) and its dialogue task are presented, then the way the US (user side, Section 6.3.2) and the ASR output simulator are instantiated is described.

### 6.3.1 The Service: personal agenda assistant

#### Dialogue Manager

A personal agenda assistant has been implemented as a new task for the experiments. The user can add, move or delete events in his agenda. For instance, a request could be:

*I would like to add the event football game on March 3<sup>rd</sup> from 9 to 10 pm*<sup>4</sup>. This is a slot-filling task with four slots:

- **ACTION:** The type of action the user wants to perform. Can take three different values: ADD, MODIFY or DELETE.
- **DESCRIPTION:** The description of the event.
- **DATE:** The date of the event.
- **WINDOW:** The time window of the event.

However, no overlap is tolerated between events in the agenda.

Each interaction between the US and the system is defined by a scenario. Here, the US is given two lists of events: *InitList* and *ToAddList*. The first one contains the events that already exist in the agenda before the dialogue and the second one contains the ones that the US is supposed to add during the dialogue. Each event is associated with a priority value and the US must prefer adding the ones with high priority first<sup>5</sup>. Its aim is to make as many events with the highest priority values as possible fit in the agenda.

### Natural Language Understanding

A rule-based algorithm transforms the user's utterance hypothesis into concepts. To do that, a set of rules have been defined. Each rule transforms a word, a concept or any combination of the two into a new concept. Three types of rules are used; they are depicted in Table 6.1.

For instance, parsing the sentence *I want to add the event birthday party on January 6<sup>th</sup> from 9pm to 11pm* is performed following these steps:

1. **I want to ADD the TAG\_EVENT birthday party on MONTH(January) NUMBER(6) from TIME(9,0) to TIME(11,0)**
  - add : [ADD]
  - event : [TAG\_EVENT]
  - Regex(janvier | ... | decembre) : MONTH(\$word)
  - Regex([0-9]+) : NUMBER(\$word)
  - Regex((( [0-1]?[0-9] ) | (2[0-3]))h([0-5][0-9])?) : TIME(\$word)<sup>6</sup>
2. **I want to ADD the TAG\_EVENT birthday party on DATE(6,1) WINDOW(TIME(21,0),TIME(23,0))**

---

<sup>4</sup>The dialogues are actually in French but they are translated in English to ensure language coherence in this thesis.

<sup>5</sup>To make the algorithms easier to understand, the larger priority value, the more important the event, unlike common usage.

<sup>6</sup>Adapted to the french way of uttering time values.

Rule type	Description	Example
Tag	Words are associated with labels	remove : [DELETE]
Regular expressions	Words matching a regular expression are transformed into concepts	Regex([0-9]+) : NUMBER(\$word)
Combine	Words and concepts are mapped into a new concept	Combine(NUMBER,MONTH) : DATE

**Table 6.1:** NLU rules types

- Combine(NUMBER,MONTH) : DATE(NUMBER,MONTH)
  - Combine(from,TIME\_1,to,TIME\_2) : WINDOW(TIME\_1,TIME\_2)
3. I want to ADD EVENT(birthday party, DATE(6,1), WINDOW(TIME(21,0),TIME(23,0)))
- Combine(TAG\_EVENT,\$x,on,DATE,WINDOW) : EVENT(\$x,DATE,WINDOW)
4. I want ACTION(ADD, EVENT(birthday party, DATE(6,1), WINDOW(TIME(21,0),TIME(23,0))))
- Combine(ADD,EVENT) : ACTION(ADD,EVENT)

### 6.3.2 Simulator instantiation

The aim of the US is to make the biggest number possible of events with the highest priority values taken from *InitList* and *ToAddList* fit in the agenda. To do so, the Intent Manager executes the `run()` function (taking the action stack (AS) as an argument) depicted in Algorithm 2 (an instance of the abstract function depicted in Algorithm 1) where

- The manipulated structures are stacks and queues.
- If `st` is a stack, the top element is called `st.top` and `st.removeTop()` removes it from `st`.
- If `q` is a queue, the next element is called `q.next` and `q.removeNext()` removes it from `q`.
- The size of the structure `x` is called `x.size`.
- If `m` is a map, `m.add(x, y)` adds a new entry to `m` such that `m(x) = y`.
- AS is a stack of actions and each action `a` is a couple (`a.action, a.event`) where `a.action` is an action type (ADD, MODIFY or DELETE) and `a.event` is an event characterised by a description, a date, a time window and a priority value.
- For each event `e`, `e.alt` refers to all its alternatives including itself<sup>7</sup>. Its priority is called `e.prio`.

<sup>7</sup>The elements of `e.alt` share the same description and priority value but they have different dates and time windows.

```
Algorithm run (AS)
  while AS.size > 0 do
    altQ  $\leftarrow$  AS.top.event.alt
    c  $\leftarrow$  empty map
    M  $\leftarrow$  empty map
    do
      conflQ  $\leftarrow$  execute((AS.top.action, altQ.next))
      if conflQ.size > 0 then
        c.add(altQ.next, conflQ)
        M.add(altQ.next,  $\max_{e \in \text{conflQ}} e.\text{prio}$ )
        altQ.removeNext()
      end
    while altQ.size > 0 AND conflQ.size > 0
    if conflQ.size > 0 then
       $e_{\text{best}} \leftarrow \text{argmin}(M)$ 
      if M( $e_{\text{best}}$ ) < AS.top.event.prio then
        for  $e \in c(e_{\text{best}})$  do
          if e.alt.size > 1 then
            for  $e' \in e.\text{alt} - \{e\}$  do
              AS.add((MOVE,  $e'$ ))
            end
          else
            AS.add((DELETE, e))
          end
        end
      else
        AS.removeTop()
        if AS.top.action == MOVE then
          AS.add((DELETE, AS.top.event))
        end
      end
    else
      AS.removeTop()
    end
  end
```

**Algorithm 2:** Intent Manager algorithm

Parameters	Value	Justification
$p_{od}$	0.1	Based on the corpus study led in (Ghigi et al., 2014)
$p_{rep}$	0.3	Id.
SR	200 words per minute	See (Yuan et al., 2006)
$\mu_{pat}$	3 min	Empirical given the task
$p_{repl}$	0.7	Empirical (variable and difficult to estimate)
$p_{add}$	0.15	Id.
$p_{del}$	0.15	Id.
N	5	Empirical (can be anything depending on the ASR configuration)
INBF	0.7	Tuned for a reasonable boosting effect with N=5
BF	0.2	Id.
$c_{err}$	-1	Empirical (variable and difficult to estimate)
$c_{right}$	1	Empirical (variable and difficult to estimate)
$\sigma_{conf}$	1	Id.
SM	2	words that lasted for more than 0.6 seconds have 90% chance of staying unchanged (McGraw and Gruenstein, 2012) and SR = 200 wpm
WER	Variable during the experiments	

Table 6.2: User simulator and ASR output simulator values

- The function `execute()` communicates an intent to the NLG and gets the corresponding response through the NLU. If this response is a misunderstanding declaration, then the same intent is sent again. If it is a declaration of conflict with a set of other events, then a queue with all these events is returned. Finally, if no conflict was detected, then this function returns an empty queue. This is how the US updates AS.

The other parameters of the US and the ASR output simulator are set as shown in Table 6.2.

## 6.4 Functionning illustration

The following illustration provides the reader with a global view of all the notions introduced in this chapter through an application to a concrete interaction in the agenda management domain. Consider the following scenario:

- **InitList:** {**title:** *house cleaning*, **date:** *January 6<sup>th</sup>*, **window:** *from 18 to 20*, **priority:** 3, **alternative 1:** *January 7<sup>th</sup>, from 6pm until 8pm*, **alternative 2:** *January 9<sup>th</sup>, from 10am until 12am*}
- **ToAddList:** {**title:** *birthday party*, **date:** *January 6<sup>th</sup>*, **window:** *from 6pm until 11pm*, **priority:** 5}

AS initially contains the ADD action associated with the event *birthday party* since it is the only event in *ToAddList*. Let  $e$  be this event. The function `run()` in Algorithm 2 is then run over this stack. Thus the first result generated by the Intent Manager is the ADD action corresponding to the event *birthday party*. Once communicated to the NLG, the latter outputs the sentence *Add the event birthday party on January 6<sup>th</sup> from 6pm until 8pm* which is in turn transfered to the Verbosity Manager.  $p_{od} = 0.1$  so there

is a 10% chance that the result of the NLG is ignored and replaced by an off-domain sentence. Suppose it is the case for this first trial then the ASR output simulator's result is also off-domain. As a consequence, the service responds by saying that the user's requests has not been understood which is recognised as a dialogue act by the NLU and communicated to the Intent Manager. The function `execute()` then sends the same intent once again to the NLG which generates the same utterance as it did the first time. As the last system's response is a misunderstanding declaration, the Verbosity Manager has  $p_{rep} = 0.3$  chance of repeating the same sentence twice. Suppose it is the case, then the Verbosity Manager's output is the following (a prefix and a suffix are also randomly added by this module): *Can you please add the event birthday party on January 6<sup>th</sup> from 6pm until 8pm, I repeat, add the event birthday party on January 6<sup>th</sup> from 6pm until 8pm if it is possible.* The ASR output simulator generates an N-Best as it is described in Figure 6.3.

Suppose that *CommitCond* in the Scheduler is true if and only if the response to the last stable utterance is either a confirmation question or a conflict declaration (in other words, all the slots have been successfully communicated to the system). Also, suppose that this time, the best ASR hypothesis is 100% correct. Since  $SM=2$ , the Scheduler decides to take the floor as soon as the users decides to utter *Can you please add the event birthday party on January 6<sup>th</sup> from 6pm until 8pm, I repeat.*

Since the time window required to organise the birthday party is already taken by the house cleaning event `execute()` returns a queue (`conflQ`) containing the latter only, and the entry  $(e, conflQ)$  is added to  $c$ . The priority of the house cleaning event is 3, therefore, the entry  $(e, 3)$  is added to  $M$ . Therefore,  $e_{best} = e$  since it is the only element in  $M$ . Its priority is lower than  $AS.top.event.prio$ . As a consequence, the Intent Manager tries to move the house cleaning event (details about the sentence formulation and management are no longer given) and it ends up to be successful using the first alternative: *January 7<sup>th</sup>, from 6pm until 8pm.* By doing so, the event *birthday party* is still in the stack and it is added facing no problem this time since the corresponding time window has been freed.

To sum up, a new incremental dialogue simulation which supports dialogue instability has been introduced in detail in this chapter. Its different components have been thoroughly described as well as the way they interact with each other to incrementally send requests to the system and to process its responses. In the rest of this thesis, the strategies formerly described are implemented and evaluated in this simulated environment, then the latter is used to train a reinforcement learning turn-taking strategy.

## Chapter 7

# Handcrafted strategies for improving dialogue efficiency

Several strategies for handling slot-filling dialogue tasks have been discussed in Chapter 5. The hypothesis laid stipulates that they perform differently under different noise conditions. A rough study has been made leading to a first model that made it possible to derive the tendencies of these strategies in terms of dialogue duration as a function of noise. In this chapter, the personal agenda simulated environment introduced in Chapter 6 is used in order to validate these results.

On top of that, incremental processing offers a new degree of freedom that can be used in order to derive new dialogue strategies. Inspired by the TTP taxonomy introduced in Chapter 3, five turn-taking strategies have been selected for their a priori potential to improve task oriented dialogue. Here, they are also implemented in the simulated environment and mixed with the previous turn-taking strategies. The new formed strategies are also compared in terms of dialogue duration and task completion (published as (Khouzaimi et al., 2015a)).

The best strategy obtained in this chapter will serve as a baseline for the one elaborated later (automatically learnt from data).

## 7.1 User and system initiative

### 7.1.1 Strategies

In Chapter 5, three slot-filling strategies have been addressed: SysIni, UsrIni and MixIni. Here, they are implemented in the personal agenda management simulated environment and compared. In the following, the way they are instantiated as well as dialogue examples are provided.

- **System Initiative (SysIni):** Recall that while using this strategy, the system asks



the user for the different slot values one by one. The first slot the system asks for is the action type: whether it is an ADD, a MODIFY or a DELETE action. To improve the system's performance, the order the remaining slots are asked for is made dependent on this type of action.

In the case of an ADD action, the date and the slot are asked first. Therefore, in the case of a conflict, there is no need to ask for the description. If no conflict is detected, the description is asked for and as it is an open slot (anything could be taken as a valid description), the user is asked for a confirmation of that particular slot before moving to the general confirmation (related to the complete request with all the associated slots).

As far as MODIFY actions are concerned, the description slot can also be an invalid utterance if the user provides a non existing event<sup>1</sup>. Therefore, as it involves only one dialogue turn (instead of the two turns required to determine the date and the time window) and as it is more natural to perform a MODIFY action by asking which event is to be moved, the description slot is asked for first.

Finally, DELETE actions are the most simple as they require only one additional slot: the description. The following dialogue illustrates the way these three types of actions are handled in the context of a noisy environment where a few requests fail to be communicated to the system.

SYSTEM: What kind of action do you want to perform?

USER: Add.

SYSTEM: Please specify a date.

USER: March <noise>.

SYSTEM: Sorry I don't understand.

USER: March 6<sup>th</sup>.

SYSTEM: Please specify a time window.

USER: From 10am until 11am.

SYSTEM: Sorry. The time window on March 6<sup>th</sup> from 10am until 11am is already taken by the event football game that is planned on March 6<sup>th</sup> from 10am until 12am. Do you want to specify another date and time window?

USER: Yes.

SYSTEM: Please specify a date.

USER: March 7<sup>th</sup>.

---

<sup>1</sup>In order to improve the dialogue efficiency, in the case of MODIFY and DELETE requests, only a prefix of the description is necessary to identify an event. For instance, if the event *Guitar lesson* is planned in the agenda, then the user saying *Guitar* is enough for the system to understand that she is referring to that event. In the case it is a prefix of more than one event, the first one in the system's internal list is taken.

SYSTEM: Please specify a time window.

USER: From 10am until 11am.

SYSTEM: Please specify a description.

USER: Long list. (*Recognition error: the user actually said Dentist*).

SYSTEM: You said long list right?

USER: No.

SYSTEM: Please specify a description.

USER: Dentist.

SYSTEM: You said dentist right?

USER: Yes.

SYSTEM: Ok. So you want to add the event dentist on March 7<sup>th</sup> from 10am until 11am. Is that right?

USER: Yes.

SYSTEM: Done. Would you like to perform other actions?

USER: Yes.

SYSTEM: What kind of action do you want to perform?

USER: Move<sup>2</sup>.

SYSTEM: Please specify a description.

USER: Guitar <noise>

SYSTEM: Please specify a new date.

USER: March 10<sup>th</sup>.

SYSTEM: Please specify a new time window.

USER: From 2pm until 3pm.

SYSTEM: Ok. So you want to move the event guitar lesson to March 10<sup>th</sup> from 2pm until 3pm. Is that right?

USER: Yes.

SYSTEM: Done. Would you like to perform other actions?

USER: Yes.

SYSTEM: What kind of action do you want to perform?

USER: Delete.

---

<sup>2</sup>Move and modify actions are equivalent

SYSTEM: Please specify a description.

USER: Visit George.

SYSTEM: Ok. So you want to delete the event visit George that is planned for March 9<sup>th</sup> from 6:30pm until 7:30pm. Is that right?

USER: Yes.

SYSTEM: Done. Would you like to perform other actions?

USER: No.

SYSTEM: Thank you. Bye.

As discussed in Chapter 5, this strategy is very tedious, however, it has the advantage of baby-stepping the dialogue and recovering from errors as soon as they occur.

- **User Initiative (UsrIni):** In the case of ADD or MODIFY actions, the event description, its date and its time window should be provided in the user's request, otherwise, an error is reported. For DELETE actions, only the description is required. For example:

SYSTEM: What can I do for you?

USER: I would like to add the event dentist on March <noise> from 10am until 11am.

SYSTEM: Sorry I don't understand.

USER: I said, I want to add the event dentist on March 6<sup>th</sup> from <noise> until 11am.

SYSTEM: Sorry I don't understand.

USER: Add the event dentist on March 6<sup>th</sup> from 10am until 11am.

SYSTEM: Sorry. The time window on March 6<sup>th</sup> from 10am until 11am is already taken by the event football game that is planned on March 6<sup>th</sup> from 10am until 12am. What can I do for you?

USER: I would like to add the event dentist on March 7<sup>th</sup> from 10am until 11am.

SYSTEM: Ok. So you want to add the event dentist on March 7<sup>th</sup> from 10am until 11am. Is that right?

USER: Yes.

SYSTEM: Done. Would you like to perform other actions?

USER: Yes.

SYSTEM: What can I do for you?

USER: I want to move the event guitar <noise> to March 10<sup>th</sup> from 2pm until 3pm please.

SYSTEM: Ok. So you want to move the event guitar lesson to March 10<sup>th</sup> from 2pm until 3pm. Is that right?

USER: Yes.

SYSTEM: Done. Would you like to perform other actions?

USER: Yes.

SYSTEM: What can I do for you?

USER: Delete the event visit Georges.

SYSTEM: Ok. So you want to delete the event visit George that is planned for March 9<sup>th</sup> from 6:30pm until 7:30pm. Is that right?

USER: Yes.

SYSTEM: Thank you. Bye.

Again, as pointed out in Chapter 5, with this strategy, the dialogue can progress much faster if no errors occur. Nevertheless, unlike SysIni, when there is still missing information (even one slot), the user has to repeat his request from the start.

- **Mixed Initiative (MixIni):** As previously discussed, this strategy starts in a similar way as UsrIni. The action type slot is important to determine which other slots should be gathered and how they should be processed. Therefore, if it is not detected in the user's utterance, the other slots are ignored and a misunderstanding is reported. On the contrary, if it has been detected, the system spots all the other slots present in the sentence and tries to gather the missing ones by switching to the SysIni strategy. For example:

SYSTEM: What can I do for you?

USER: I would like to add the event dentist on March <noise> from 10am until 11am.

SYSTEM: Please specify a date.

USER: <noise>.

SYSTEM: Sorry I don't understand.

USER: March 7<sup>th</sup>.

SYSTEM: Ok. So you want to add the event dentist on March 7<sup>th</sup> from 10am until 11am. Is that right?

USER: Yes.

By elaborating such a hybrid strategy, an a priori approach introduced in Chapter 5 showed that it allies both advantages of SysIni and UsrIni. In the following, the simulated environment is used to validate this affirmation.

### 7.1.2 Experiments

For the experiments, three dialogue scenarios have been used. As described in 6.2.1, a scenario is specified by two lists of events: *InitList* and *ToAddList*. The lists corresponding to these three scenarios are given in Table 7.1.

Scenario	Title (Priority)	Date	Window	DateAlt1	WindowAlt1	DateAlt2	WindowAlt2
1-Init	Guitar lesson(4)	November 17 <sup>th</sup>	14:00-15:30	November 15 <sup>th</sup>	9:45-11:15		
1-ToAdd	Book reading(8)	November 19 <sup>th</sup>	10:30-12:30	November 14 <sup>th</sup>	9:30-11:30	November 18 <sup>th</sup>	16:30-18:30
	Watch the lord of the rings(12)	November 13 <sup>th</sup>	9:30-12:30	November 15 <sup>th</sup>	11:15-14:15		
2-Init	Guitar lesson(4)	November 17 <sup>th</sup>	14:00-15:30	November 15 <sup>th</sup>	9:45-11:15		
2-ToAdd	Tennis(5)	November 17 <sup>th</sup>	13:15-15:15	November 19 <sup>th</sup>	15:15-17:15		
	Gardening(9)	November 18 <sup>th</sup>	13:15-15:15	November 14 <sup>th</sup>	12:30-14:30		
3-Init	Guitar lesson(4)	November 17 <sup>th</sup>	14:00-15:30	November 15 <sup>th</sup>	9:45-11:15		
3-ToAdd	Holidays preparation(1)	November 16 <sup>th</sup>	12:30-14:30	November 17 <sup>th</sup>	12:15-14:15		
	House cleaning(6)	November 13 <sup>th</sup>	14:15-16:15	November 17 <sup>th</sup>	15:30-17:30		
	Give back book(7)	November 16 <sup>th</sup>	14:00-14:30	November 13 <sup>th</sup>	14:00-14:30		

Table 7.1: The three scenarios used in the simulation

The WER in real dialogue systems generally varies between 0.1 and 0.3. Here, it is varied between 0 and 0.3 with a step of 0.03 in order to analyse the impact of noise on the different strategies. For each noise level, the three scenarios are run 1000 times. The mean duration of the dialogues, their task completion as well as the corresponding 95% confidence intervals are depicted in Figure 7.1.

The results obtained here clearly validate the analysis led in Chapter 5, even though the simplifying assumptions made in the latter are relaxed here. The graph in Figure

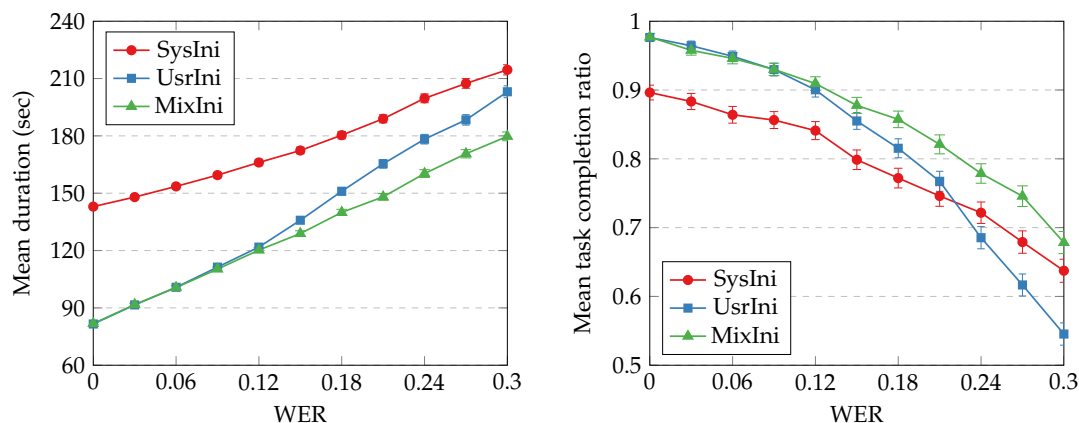


Figure 7.1: Simulated mean duration (left) and dialogue task completion (right) for different noise levels

5.1 is very similar to the ones depicted in Figure 7.1 which also confirms the well functioning of the simulated environment. In low noise setups, SysIni is clearly less efficient than UsrIni; the dialogues take twice more time to finish which in turn translates into a lower task completion ratio. Nevertheless when the noise level reaches about 0.2, SysIni offers better completion rates. The duration is still lower in spite of the correlation between the two metrics. This is due to the fact that the durations distribution for UsrIni is centered on short dialogues whereas the distribution for SysIni is centered on average ones. Finally, MixIni is the best strategy since it allies both the advantages of both UsrIni and SysIni.

## 7.2 Incremental strategies

Five TTP have been selected for implementation in Chapter 3: FAIL\_RAW, INCOHERENCE\_INTERP, FEEDBACK and BARGE\_IN\_RESP both from the system's and the user's point of view. Their generic implementation have been discussed in Chapter 5 and in the following, more specific implementation details in the personal agenda management domain are provided.

### 7.2.1 TTP implementation

First, recall that as announced in Chapter 6, SM=2 and therefore, the last stable utterance corresponds to the last user's request without the last two words.

**FAIL\_RAW:** Happens when the user speaks for too long with no key concept detected in her speech. It depends on the system's last question (the type of information it is waiting for). For open questions (all the slots at once), this key concept is the action type (ADD, MODIFY or DELETE). For the dates and the time windows, it corresponds to the slot value whereas in the case of yes/no questions, it corresponds to the concepts YES and NO. This TTP is not relevant when the user utters an event description since it is an open slot. Let *KeyConcept* be a boolean indicating whether the key concept the system is waiting for has been pronounced or not and let  $\Delta t$  be the duration of the user's utterance so far. Therefore:

$$CommitCond = \neg KeyConcept \wedge (\Delta t \geq \xi)$$

where  $\xi$  is set to 6 for open questions, 4 for date questions, 6 for time window questions and 3 for yes/no questions (values which showed empirical best values on the task).

**INCOHERENCE\_INTERP:** In this implementation, this event is triggered as soon as the last stable utterance generates an overlap with an existing event in the agenda, or it tries to move or delete a non existing event.

**FEEDBACK:** At each new micro-turn, given that no boosting is involved, the N-Best best score relative variation is equal to the best score of the most recent word N-Best. Therefore, as the word score is not accessible by the Scheduler, the ratio  $s_t/s_{t-1}$  can be used as a proxy for that value. Similarly, the score of the last word of the last stable utterance can be estimated by  $s_{t-SM}/s_{t-SM-1}$ . Let *persist* be a boolean that determines whether the last word of the last stable utterance has changed since it has first been pronounced, then (0.7 being an empirical value):

$$CommitCond = persist \wedge \left( \frac{s_{t-SM}}{s_{t-SM-1}} \leq 0.7 \right)$$

**BARGE\_IN\_RESP (System):** Depending on the last system's dialogue act (apart from dialogue acts reporting errors), the system can choose to barge-in once it has all the information needed to provide an answer. Again, it should also wait for the SM.

**BARGE\_IN\_RESP (User):** When the user gets familiar with the system, it tends to predict the system's dialogue act before the system finishes its sentence. Unlike the previous phenomena, this one is due the user's decision. Hence, it has been implemented in the US: a barge-in point has been manually determined in advance for each system's dialogue act template. The following excerpt dialogue illustrates this behaviour:

SYSTEM: Hello. Welcome to...

USER: Add

SYSTEM: Please specify a date.

USER: <noise>

SYSTEM: Sorry, I don't...

USER: July 7<sup>th</sup>

SYSTEM: Please specify a time...

USER: From 4pm until 6pm

SYSTEM: Sorry, this time window is already...

USER: Yes

SYSTEM: Please specify a date.

### 7.2.2 Experiments

In this experiment, the TTP described in Section 7.2.1 are used on the top of the slot-filling strategies introduced in Section 7.1.1. Incremental processing is more useful when the user makes long utterances. This is not the case in the SysIni strategy where her utterances are very short. Therefore, incremental behaviours have only been added to UsrIni and MixIni to form two new strategies: UsrIni+Incr and MixIni+Incr. The associated performances are depicted in Figure 7.2.

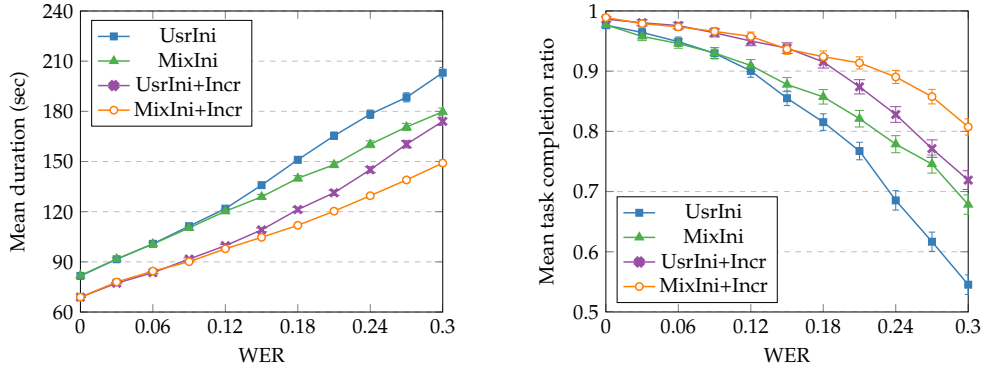


Figure 7.2: Mean dialogue duration and task completion for aggregated strategies.

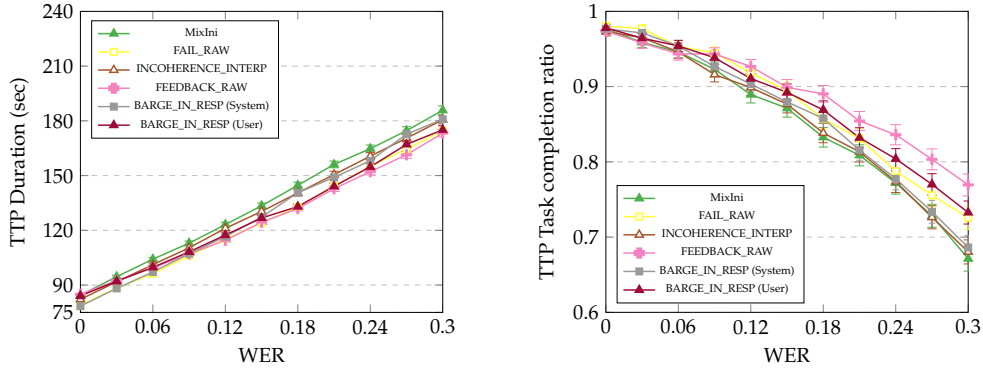


Figure 7.3: Mean dialogue duration and task completion for different turn-taking phenomena.

Adding mixed initiative behaviour or incrementality to UsrIni are both ways to improve its robustness to errors. Figure 7.2 shows that incrementality is more efficient. Most importantly, it is shown that MixIni and incremental behaviour can be combined to form the best strategy.

As already mentioned in Chapter 3, the main objective of the TTP taxonomy is to break human dialogue turn-taking into small pieces in order to get a better understanding of it. To illustrate this approach, a deeper look is taken at MixIni+Incr by isolating its different components<sup>3</sup>: FAIL\_RAW, INCOHERENCE\_INTERP, FEEDBACK,

<sup>3</sup>One of the advantages of the Scheduler approach is that the different TTP are implemented as inde-



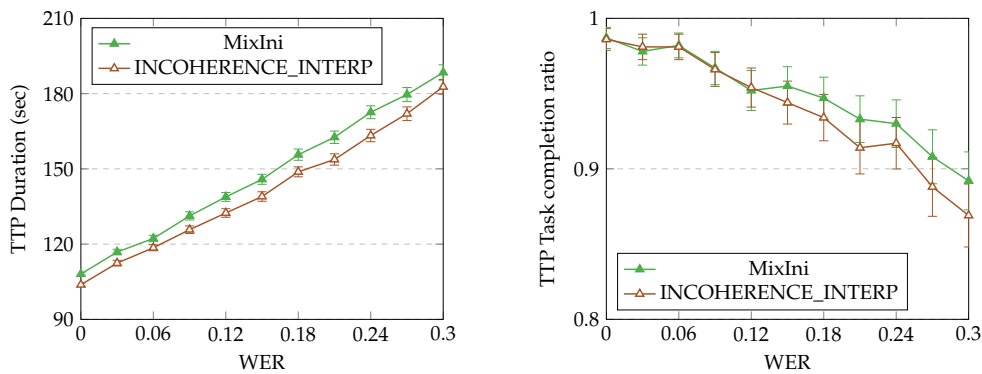


Figure 7.4: *INCOHERENCE\_INTERP* evaluated in a more adapted task

BARGE\_IN\_RESP (User) and BARGE\_IN\_RESP (System). The results reported in Figure 7.3 show that FEEDBACK contributes the most to improve the baseline followed by BARGE\_IN\_RESP (User) and FAIL\_RAW. *INCOHERENCE\_INTERP* and BARGE\_IN\_RESP (System) seem to have no effect. This is due to the fact that in general, to detect an incoherence, one must wait until the end of the utterance (same requirement for detecting all the information needed to barge-in and provide an answer). One might argue that in some cases, the US tries to refer to a non existing event (in the case of a MODIFY or DELETE action), therefore triggering an incoherence. However, as stated before, the service is able to recognise an existing event even if only a prefix of its description is recognised. As a consequence, *INCOHERENCE\_INTERP* is rarely triggered before the end of a request. Therefore, another scenario where that situation occurs more often has been tested: the US has to try several time window alternatives to add an event and only the fifth one is free. The results are shown in Figure 7.4: this time, *INCOHERENCE\_INTERP* has a visible added value.

This experiment raises an important point when it comes to studying efficiency in task oriented dialogues as far as turn-taking mechanisms are concerned. The nature of dialogue is very diverse, therefore, results and the following conclusions (whatever the chosen metric is) should not be given separately from the nature of the task at hand. In the previous example, *INCOHERENCE\_INTERP* is shown to have no impact in some kind of situations but it brings some improvement in other situations. Another example is the feedback which is more likely to improve the baseline when open slots are involved and especially long ones like this is the case in message dictation for example (*Send mom a message telling her that...*) whereas it may have very limited impact in a train booking dialogue for example, where only constrained slots come into play. It is one of the motivations behind adopting a data-driven approach, so that the Scheduler figures out by itself what are the appropriate behaviours given the situation at hand.

---

pendent decision makers that can easily be combined to form the aggregated strategies, but that can also be isolated and analysed separately.

So far, a handcrafted turn-taking strategy has been implemented in the simulated environment introduced in Chapter 6 along with several slot-filling strategies. These first results show that there is room for dialogue efficiency improvement by enhancing the system's turn-taking capabilities. In the following, this handcrafted strategy is improved using reinforcement learning and the results are validated in a live study with real users.



## Chapter 8

# Reinforcement learning for turn-taking optimisation

The experiments led in Chapter 7 showed that implementing incremental strategies in the Scheduler can improve dialogue efficiency. However, this approach requires the designer to handcraft the strategies most of the time in an empirical way. She has to come up with rules that are adapted to the type of the task at hand and to manually tune parameters. Moreover, the result is not guaranteed to be optimal.

In this chapter, a new approach is proposed where the Scheduler automatically learns optimal turn-taking behaviours through interactions (published as ([Khouzaimi et al., 2015b](#))). Reinforcement learning is applied in order to make decisions at the micro-turn level based on a new state representation model. A new simulation experiment shows that the resulting strategy outperforms the handcrafted one from Chapter 7.

Turn-taking is most of the time locally optimised using supervised learning ([Raux and Eskenazi, 2009](#); [Meena et al., 2013](#)). The different moments when a participant takes the floor (either interrupting the initial speaker or not) are treated independently. By using reinforcement learning, no such assumption is made and it is possible to learn different turn-taking behaviours according to the dialogue advancement. Moreover, the purpose of such studies is to minimise gaps and overlaps. However, even though this objective function is justified ([Sacks et al., 1974](#)), nothing proves that it leads to optimal dialogues in every situation. Again, using reinforcement learning makes it possible to avoid this second assumption too. Finally, it is also well adapted to learning from delayed rewards which is interesting here since the dialogue quality is only available at the end of the dialogue (or at the end of independent dialogue episodes like it is the case here).

## 8.1 Reinforcement learning Model

### 8.1.1 Background

This section recalls a few elements about reinforcement learning and MDPs (see Chapter 2 for more details). The Scheduler will be modeled as an MDP: a tuple  $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$  where

- $\mathcal{S}$  is the state space: all the states in which the agent could be.
- $\mathcal{A}$  is the action space: all the actions that it can perform.
- $\mathcal{T}$  is the transition model: the distributions over the state space where the random variable is the next state  $s' \in \mathcal{S}$  given that the agent is currently in state  $s \in \mathcal{S}$  and performs action  $a \in \mathcal{A}$ .
- $\mathcal{R}$  is the reward model: the distributions over  $\mathbb{R}$  where the random variable is the immediate reward  $r \in \mathbb{R}$  given that the agent is currently at state  $s \in \mathcal{S}$  and performs action  $a \in \mathcal{A}$ .
- $\gamma$  is a discount factor in  $[0, 1[$ . The more it is close to 1, the more the agent is farsighted (maximising long term returns).

The time step at which the Scheduler is run is the micro-turn. Therefore, at each new micro-turn, the Scheduler computes the current state and then makes a decision (which action to perform). As discussed in Chapter 2, the state space representation is a challenge in itself since it should embed all the necessary information while being tractable. In the next section, a new state representation adapted to incremental processing is introduced. It is general enough so that it can be easily implemented for any slot-filling task.

### 8.1.2 State representation

At each new micro-turn  $\mu T_i^{k,U}$ , the following features are used to describe the system state (an interaction example where the features' values are given is provided rightafter in Table 8.1):

- **SYSTEM\_REQ:** During the user dialogue turn  $T^{k,U}$ , the system is requiring a particular information. For instance, after an open question, it is waiting for all the slot values to be provided at once but it can also be waiting for a specific slot value or a response to a confirmation. This feature refers to the information that it is waiting for during  $T^{k,U}$ . It can take 6 different values<sup>1</sup> in the personal agenda management domain:

– General prompt: *What can I do for you?*

---

<sup>1</sup>The dialogue act *Done. Would you like to perform other actions?* is not a possible value since it is considered as a transition and not a part of the learning episode. For this specific dialogue turn, the Scheduler always choses to WAIT.

- Description question: *Please specify a description.*
- Date question: *Please specify a date.*
- Time window question: *Please specify a time window.*
- Description confirmation: *You said <description> right?*
- Confirmation: *Ok. So you want to add the event <event>. Is that right?*
- **LAST\_INCR\_RESP:** As described in Chapter 4, the Scheduler stores the last response it gets from the service at each micro-turn. It is used as a second feature which can take 11 different values, the 6 values possible for SYSTEM\_REQ and the five ones listed below:
  - Conflict: *Sorry. The time window <date> <time window> is already taken by the event <event>. What can I do for you?*
  - Misunderstanding: *Sorry I don't understand.*
  - Yes or no misunderstanding: *Sorry I don't understand. Please answer this question by saying yes or no.*
  - Not existing event: *Sorry, the event <event> does not exist. What can I do for you?*
  - Other operations: *Done. Would you like to perform other actions?*

The reason why this feature and SYSTEM\_REQ have different values possible even though they represent the same dialogue acts is that a single dialogue act can be viewed differently. For example, the system response *Sorry. The time window <date> <time window> is already taken by the event <event>. What can I do for you?* is a conflict declaration from the LAST\_INCR\_RESP point of view but it is an open question when viewed as a SYSTEM\_REQ. Moreover, when the system declares a misunderstanding, SYSTEM\_REQ does not change (the system is still waiting for the same information). LAST\_INCR\_RESP represents the reaction that the system would have if interrupted at each micro-turn.

- **NB\_USER\_WORDS:** This feature is a counter of the number of words since the last change of LAST\_INCR\_RESP (the number of words since the Service did not change its mind about the response to deliver). It is equal to zero at the exact micro-turn when the change happens and it is incremented until the next change.
- **NORMALISED\_SCORE:** At each micro-turn, the ASR score is updated: most of the time, it is multiplied by the ASR score corresponding to the new incoming word (see Figure 6.3). Except from the cases where a boost comes into play, the score keeps decreasing as the user speaks. To avoid penalising long sentences, the score is normalised by taking the geometric mean over the words (this induces a bias since the number of inputs that forms the current ASR hypothesis may not be exactly the number of words because of the Scrambler's additions and deletions). If  $s$  is the current score for  $n$  number of words,  $NORMALISED\_SCORE = s^{\frac{1}{n}}$ .

- **TIME:** Corresponds to the duration of the current episode in milliseconds.

It is noteworthy that the only domain-related features are SYSTEM\_REQ and LAST\_INCR\_RESP, therefore, in order to transpose this method to another domain, one should only provide the list of alternatives for both features. This is due to the fact that, instead of directly analysing the current partial utterance at each micro-turn, only the response it generates once fed to the Service is taken into account, and this is proven here to be a sufficient information.

A linear model is used to represent the Q-function (Sutton and Barto, 1998). First, it has been noticed that 21 combinations between SYSTEM\_REQ and LAST\_INCR\_RESP are frequently visited (the others barely happen or not at all). Therefore, 21 features are defined  $\delta_1, \dots, \delta_{21}$  where  $\delta_i = 1$  if and only if the current state corresponds to the  $i^{th}$  combination, and 0 otherwise. The rare combinations are not included in the model since they require maintaining heavier models with no real improvements over the simpler ones.

The Q-function should clearly not be monotonous with respect to NB\_USER\_WORDS since the user should not be interrupted too soon nor too late. It should be maximal around some value, therefore, the Q-function should not be linear with respect to that feature. Instead, NB\_USER\_WORDS is represented by three RBF functions  $\phi_1^{nw}$ ,  $\phi_2^{nw}$  and  $\phi_3^{nw}$  centered in 0, 5 and 10 with a standard deviation of 2, 3 and 3. In other words

$$\begin{aligned} \phi_i^{nw} &= \exp\left(\frac{(NB\_USER\_WORDS - \mu_i)^2}{2\sigma_i^2}\right) \\ \mu_1 &= 0, \mu_2 = 5, \mu_3 = 10 \\ \sigma_1 &= 2, \sigma_2 = 3, \sigma_3 = 3 \end{aligned} \tag{8.1}$$

These values are empirical but they give more flexibility for the model to approach the true Q-function. Similarly, NORMALISED\_SCORE is also represented using two RBF functions  $\phi_1^{ns}$  and  $\phi_2^{ns}$  centered in 0.25 and 0.75 and with a standard deviation of 0.3 for both.

Finally, TIME is normalised so that it is near zero at the beginning of the episode and around 1 when the duration reaches 6 minutes (the maximum duration due to patience limit):

$$T = \text{sigmoid}\left(\frac{TIME - 180}{60}\right) \tag{8.2}$$

There is no need to use RBFs for this last feature since the Q-function is supposed to be monotonous with respect to it. The longer the dialogue, the more likely the user is to hang up.

Therefore, the dialogue state is represented by the following vector

Turn or micro-turn	Utterance (with score for user)	SYSTEM_REQ	LAST_INCR_RESP	NB_USER_WORDS	NORMALISED_SCORE	TIME	Scheduler's action
$T^{1,U}$	What can I do for you?						
$\mu T_1^{1,U}$	[0.94] I	Open question	Misunderstanding	0	0.94	300	WAIT
$\mu T_2^{1,U}$	[0.92] I would	Open question	Misunderstanding	1	0.96	600	WAIT
$\mu T_3^{1,U}$	[0.81] I would like	Open question	Misunderstanding	2	0.93	900	WAIT
$\mu T_4^{1,U}$	[0.79] I would like to	Open question	Misunderstanding	3	0.94	1200	WAIT
$\mu T_5^{1,U}$	[0.65] I would like to add	Open question	Date question	0	0.92	1500	WAIT
$\mu T_6^{1,U}$	[0.61] I would like to add the	Open question	Date question	1	0.92	1800	WAIT
$\mu T_7^{1,U}$	[0.52] I would like to add the event	Open question	Date question	3	0.91	2100	WAIT
$\mu T_8^{1,U}$	[0.50] I would like to add the event birthday	Open question	Date question	4	0.92	2400	WAIT
$\mu T_9^{1,U}$	[0.48] I would like to add the event birthday party	Open question	Date question	5	0.92	2700	WAIT
$\mu T_{10}^{1,U}$	[0.47] I would like to add the event birthday party on	Open question	Date question	6	0.93	3000	WAIT
$\mu T_{11}^{1,U}$	[0.45] I would like to add the event birthday party on January	Open question	Date question	7	0.93	3300	WAIT
$\mu T_{12}^{1,U}$	[0.22] I would like to add the event birthday party on January 6 <sup>th</sup>	Open question	Time window question	0	0.88	3600	WAIT
$\mu T_{13}^{1,U}$	[0.17] I would like to add the event birthday party on January 6 <sup>th</sup> from	Open question	Time window question	1	0.87	3900	WAIT
$\mu T_{14}^{1,U}$	[0.15] I would like to add the event birthday party on January 6 <sup>th</sup> from 8pm	Open question	Time window question	2	0.87	4200	WAIT
$\mu T_{15}^{1,U}$	[0.15] I would like to add the event birthday party on January 6 <sup>th</sup> from 8pm until	Open question	Time window question	3	0.88	4500	WAIT
$\mu T_{16}^{1,U}$	[0.11] I would like to add the event birthday party on January 6 <sup>th</sup> from 8pm until 11pm	Open question	Conflict	0	0.87	4800	SPEAK
$T^{2,S}$	Sorry. The time window on January 6 <sup>th</sup> from 8pm until 11pm is already taken by the event writing essay that is planned on January 6 <sup>th</sup> from 8pm until 12am. What can I do for you?						

Table 8.1: Dialogue example with state representation features' values.



$$\Phi(s) = [1, \delta_1, \delta_2, \delta_3, \phi_1^{nw}, \phi_2^{nw}, \phi_3^{nw}, \phi_1^{ns}, \phi_2^{ns}, T] \quad (8.3)$$

### 8.1.3 Actions, rewards and episodes

The system can perform the action WAIT and the action SPEAK (see Chapter 7 for a description of these actions). The action REPEAT, which consists on repeating the last word of the last stable utterance (with no intention of interrupting the user), introduces more complexity to the system for the following reasons:

- When performing a feedback, the speaker is not interrupted (total overlap) and therefore, in the case of a right feedback (repeating the true word that the user uttered), there is no visible impact on the dialogue duration. As a consequence, in simulation mode, the optimal strategy would be to perform feedbacks all the time which is of course not acceptable in real dialogue.
- Determining whether the user actually agreed with the feedback content and continued his sentence or whether he spotted an error and tried to correct it is a complicated problem in itself. The FEEDBACK\_RAW strategy implementation in Chapter 7 used the simple assumption that if the user's next word after the feedback is not a *no*, it means that the feedback content is confirmed. This is of course an approximation that is aimed to make a first assessment of the potential of the corresponding TTP but it is not always the case in real dialogue.
- The way user's might react to a feedback is not certain. They might interpret such a behaviour as an interruption which is not the desired effect.

As a consequence, the action REPEAT is not addressed here.

At each micro-turn, the system receives a reward  $-\Delta t$  which corresponds to the opposite of the time elapsed since the micro-turn before. Moreover, there are two rewarding situations where the system gets a reward of 150 (the reward the Scheduler gets when an elementary task - see Chapter 5 - is completed):

- The system takes the floor to confirm that the task corresponding to the user's requests has been accomplished. Happens when the user says *yes* to a confirmation question like in the following dialogue:

SYSTEM: Ok. So you want to add the event birthday party on January 6<sup>th</sup> from 8pm until 11pm. Is that right?

USER: Yes.

SYSTEM: Ok, done. Is there anything else I can do for you?

- The system takes the floor to declare a conflict, for example: *Sorry. The time window on March 3<sup>rd</sup> from 2pm until 5pm is not available since the event dentist is scheduled on March 3<sup>rd</sup> from 2pm until 3pm.* Even though the task has not been accomplished,

the system has successfully done its job (all the information slots have been successfully gathered and a response has been computed, even though an incoherence is reported).

An episode is a portion of a dialogue that starts with an open question (where the user is supposed to utter a complete request with all the necessary slot values) and ends with either a new open question or a user hang up (open questions due to confirmation failures do not start a new episode).

#### 8.1.4 Fitted-Q Value Iteration

Fitted-Q iteration (Richard Bellman, 1959) has already been successfully applied to dialogue management in the traditional sense (Chandramohan et al., 2010). Here it is applied to the problem of turn-taking<sup>2</sup>. Recall that the Bellman optimality equation states that

$$Q^*(s, a) = \mathbb{E}[\mathcal{R}(s, a, s') + \gamma \max_{a'} Q^*(s, a') | s, a] \quad (8.4)$$

$$Q^* = T^* Q^* \quad (8.5)$$

The operator  $T^*$  is a contraction (Bellman, 1957). As a consequence, there is a way to estimate it in an iterative way: *Value Iteration* (Banach theorem). Each new iteration is linked to the previous one as follows:

$$Q_i = T^* Q_{i-1} \quad (8.6)$$

However, an exact representation of the Q-function is assumed which is not possible in the present case since the state space is infinite. Discretisation of continuous variables is generally a poor option since it implies arbitrary choices of granularity and phase. Also, it does not solve the curse of dimensionality (and may even make it worse). A classical solution consists in using a linear representation of the Q-function:

$$\hat{Q}(s, a) = \theta(a)^T \Phi(s) \quad (8.7)$$

where  $\theta(a)$  is a parameter vector associated with action  $a$ . The aim of Fitted-Q algorithm is to estimate the parameters that approximate the Q-function  $Q^*$  best.  $\hat{Q}$  is the projection of  $Q$  on the space of the functions that can be written as a linear combination of the state vector's components. Let  $\Pi$  be the corresponding projection operator, then

---

<sup>2</sup>Q-Learning has been tried at first but it learnt poorly (it needs very important amount of data to converge). This is coherent with previous literature (Lemon et al., 2006). In (Daubigney et al., 2013), Fitted-Q has been shown to perform better.

it can be shown that  $\Pi T^*$  is still a contraction and admits a unique fixed point that can be iteratively computed as follows:  $\hat{Q}_{\theta_i(a)} = \Pi T^* \hat{Q}_{\theta_{i-1}(a)}$  for each action  $a$ . Since the transition probabilities of the MDP and the reward function are not known, a sampled operator  $\hat{T}$  is used instead of  $T$ . For a transition  $(s_j, a_j, r_j, s'_j)$ , it is defined as

$$\hat{T}Q(s_j, a_j) = r_j + \gamma \max_{a'} Q(s'_j, a') \quad (8.8)$$

The Fitted-Q algorithm therefore estimates the  $\theta$  vector using the iteration rule:  $\hat{Q}_{\theta_i(a)} = \hat{T} \hat{Q}_{\theta_{i-1}(a)}$  for each action  $a$ . To compute the projection, the least square estimator is used:

$$\theta_i(a) = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{j=1}^N \left( r_j + \gamma \max_a [\theta_{i-1}(a)^T \phi(s'_j)] - \theta(a_j)^T \phi(s_j) \right)^2 \quad (8.9)$$

where  $N$  is the number of transitions in the data batch. This is a classic least square optimisation and  $\theta_i$  can be computed as follows (since the matrix inversion does not depend on  $i$ , it has to be performed only once):

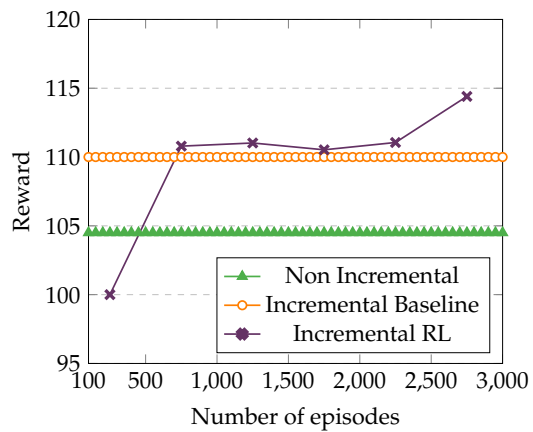
$$\theta_i(a) = \left( \sum_{j=1}^N \phi(s_j)^T \phi(s_j) \right)^{-1} \sum_{j=1}^N \phi(s_j) \left( r_j + \gamma \max_a \theta_{i-1}(a)^T \phi(s'_j) \right) \quad (8.10)$$

## 8.2 Experiment

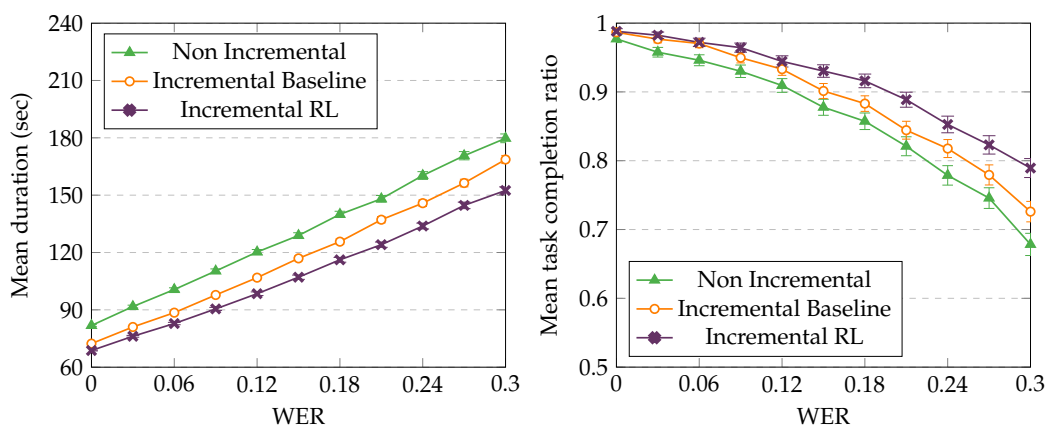
### 8.2.1 Setup

The dialogue scenarios described in Chapter 7 are used here. During learning, the noise level is fixed at 0.15. 50 parallel runs have been produced with 3000 episodes each and the average curve is depicted in Figure 8.1. The  $\theta$  parameters in the Q-function model are initiated to zeros and updated every 500 episodes, therefore, the learnt strategy reward evolution is representation by one average point in the middle of each 500 episodes interval. There are three phases to distinguish:

1. **Pure exploration (Episodes 0-500):** The actions are taken randomly with a probability of 0.9 for choosing WAIT and 0.1 for SPEAK. Picking equiprobable actions results in the user being interrupted so often that the interesting scenarios are very rarely explored.
2. **Exploration/exploitation (Episodes 500-2500):** An  $\epsilon$ -greedy policy is used with respect to the current Q-function, with  $\epsilon = 0.1$  (unlike the previous phase, when a random action is picked, the actions WAIT and SPEAK are equiprobable).
3. **Pure exploitation (Episodes 2500-3000):** A 100% greedy policy is used.



**Figure 8.1:** Learning curve (0-500: pure exploration, 500-2500: exploration/exploitation, 2500-3000: pure exploitation) with WER= 0.15.



**Figure 8.2:** Mean dialogue duration and task for the non-incremental, the baseline incremental and the RL incremental (after convergence) strategies under different noise conditions.

### 8.2.2 Results

Three different strategies are compared:

- **Non-incremental baseline:** It corresponds to the MixIni strategy defined in Chapter 7. The user is asked to provide all the information necessary to execute her request and when there are still missing slots, the corresponding values are asked for one after another.
- **Incremental baseline:** MixIni+Incr from Chapter 7 is selected as an incremental baseline. It is identical to the non-incremental baseline with the difference that it is enhanced with handcrafted turn-taking rules defined in Chapter 7.
- **Incremental RL:** It corresponds to the turn-taking strategy learned with reinforcement learning on top of the MixIni strategy for dialogue management.

Like in Chapter 7, these strategies are compared under different levels of noise. The non-incremental and the incremental baselines have already been compared in Chapter 7. In Figure 8.2, they are also compared to the new automatically learnt strategy. The differences becomes clearer as the WER increases. For WER=0.3, the non-incremental baseline reaches 3 minutes, the incremental baseline goes 10 seconds faster and the incremental RL still improves it by an additional 20 seconds (17% gain in total). In terms of task completion, the non-incremental baseline drops under 70%, the incremental baseline shows a performance of 73% whereas the incremental RL keeps the ratio at a level of 80%.

As a consequence, the reinforcement learning based strategy has been proven to improve the dialogue efficiency, even better than the handcrafted baseline. The proposed model was able to automatically learn optimal turn-taking decisions directly from interactions. In the next chapter, an experiment involving real users is run in order to validate these results.

## Chapter 9

# Experiment with real users

In spite of the efforts that have been made in order to approximate human-machine dialogue through simulation, there are still many complex behaviours and subtleties that could not be replicated. In this chapter the simulation results are tested with real users in a real dialogue setup as a validation ([Khouzaimi et al., 2016b](#)). In the following, the interaction domain is presented then the experimental protocol is described. Finally, the result experiments are depicted and discussed.

### 9.1 The Majordomo domain

The prototype used for the experiment (called Majordomo) has been implemented as a part of the VoiceHome<sup>1</sup> project at Orange Labs. This project is aimed at exploring the new opportunities that the use of the vocal modality to communicate with a smart home can bring. For that reason and in order to show that the methodology presented in this thesis is not domain dependent, this new task has been used for experiments instead of the agenda management task (event though they are somehow quite similar).

Majordomo is able to schedule a set of tasks during a specific time window. These tasks are depicted in Table 9.1.

However some tasks cannot be run simultaneously (in Table 9.1, the second column shows which tasks cannot be scheduled at the same time as the one in the first column). For example:

- The lawn cannot be mowed while watered.
- The heating cannot be activated when the windows are open.
- The swimming pool cannot be warmed and cleaned at the same time.
- When the calm mode is activated, no hoovering nor laundry are allowed.

---

<sup>1</sup>FUI project.

Task	Conflicting tasks
Alarm	Laundry, Hoover
Heating	Open windows, Air conditioning
Open windows	Heating, Air conditioning
Absence mode	Open windows
Laundry	Alarm, Calm mode
Air conditioning	Heating, Open windows
Swimming pool warming	Swimming pool cleaning
Swimming pool cleaning	Swimming pool warming
Calm mode	Laundry, Hoover, Mow lawn
Mow lawn	Water lawn
Water lawn	Mow lawn
Record channel 1	Record channel 2, Record channel 3
Record channel 2	Record channel 1, Record channel 3
Record channel 3	Record channel 1, Record channel 2
Hoover	Alarm, Calm mode
Run bath	

Table 9.1: Majordomo tasks

This is a slot-filling task where the user is supposed to provide the following information: the action type (ADD, MODIFY or DELETE), the task (see the list above), the date and the time window.

## 9.2 Experimental protocol

### 9.2.1 Implementation

The Client (interface to the users) has been developed in the form of a website where the users are supposed to read a few instructions before starting to interact with the system. Google ASR has been chosen since it is a powerful off-the-shelf solution that does not require to develop any acoustic nor language model, which is costly and not in the focus of this work. It also has a powerful named entity recognition engine (NER, see (Ben Jannet et al., 2015)) and it is able to provide incremental partial results. However, it is still not able to provide the partial confidence scores (only the confidence score at the end of the utterance is computed).

The implementation of the Service is similar to the personal agenda management case described in the previous chapters. The only difference is that home tasks are manipulated instead of events. Therefore, an open slot has been replaced with a slot where only a few alternatives are possible. Moreover, in the personal agenda management domain, no events can overlap whereas in the Majordomo domain some tasks can be run at the same time as others and others cannot. This is encoded in a compatibility matrix provided to the Service.

As far as the Scheduler is concerned, similarly to the strategies developed in the simulation case, a handcrafted (version presented in Chapter 7, with no FEEDBACK\_RAW) as well as a reinforcement learning strategy have been implemented. The reinforcement

learning strategy has been learnt in simulation (the US has also been ported to the new domain) and tested directly with real users. Since Google ASR does not provide the confidence scores incrementally, this feature has been removed from the model. Finally, when transitioning from simulation to the real world, there is no need to estimate timing from the number of words, therefore, real timestamps have been taken into account.

### 9.2.2 Conduct of the dialogue

Once the user decides to start a dialogue, the system displays the interface depicted in Figure 9.1. A small briefing paragraph explains the task to accomplish which is also synthesised in the form of a table. Ten different scenarios were designed and one of them was picked randomly at each new interaction. Similarly, the dialogue strategy is also picked randomly: the user can interact with the non-incremental strategy, with the handcrafted incremental or the reinforcement learning incremental strategy, but she is not aware which one is used for each dialogue.

When ready, the user clicks on the *Start* button. During the interaction, the ASR is always on so the Client is always listening except in the system barges-in, in which case it is disabled for two seconds. This is necessary in order to make the system take the floor, otherwise, it will be immediately interrupted before the user even realises that there is an intervention.

After the interaction, either the user ends the dialogue normally by saying *Goodbye* or hangs up by clicking on the *Hang up* button.

### 9.2.3 Key Performance Indicators

In order to evaluate the three turn-taking strategies, dialogue duration and task completion have been computed (the tasks scheduled by the Majordomo are logged at the end of the dialogue). In addition, the users filled a survey at the end of each dialogue where they provided the following subjective Key Performance Indicators (KPIs) on Lickert scales:

- **Reactivity:** The users are asked whether they found the system reactive or not. There are 6 possible answers going from 1 (*not reactive at all, very slow*) to 6 (*very reactive*).
- **Reactivity impact:** The system's reactivity does not necessarily improve the dialogue quality since it can be perceived as too intrusive. The objective of this metric is to assess the impact associated with the reactivity through 6 possible values going from 1 (*very negative, hurts the dialogue quality*) to 6 (*very positive, significant improvement of the dialogue quality*).
- **Realism:** The users are asked whether the system acts like a human operator. There are 6 possible answers going from 1 (*no, not at all*) to 6 (*yes, clearly*).



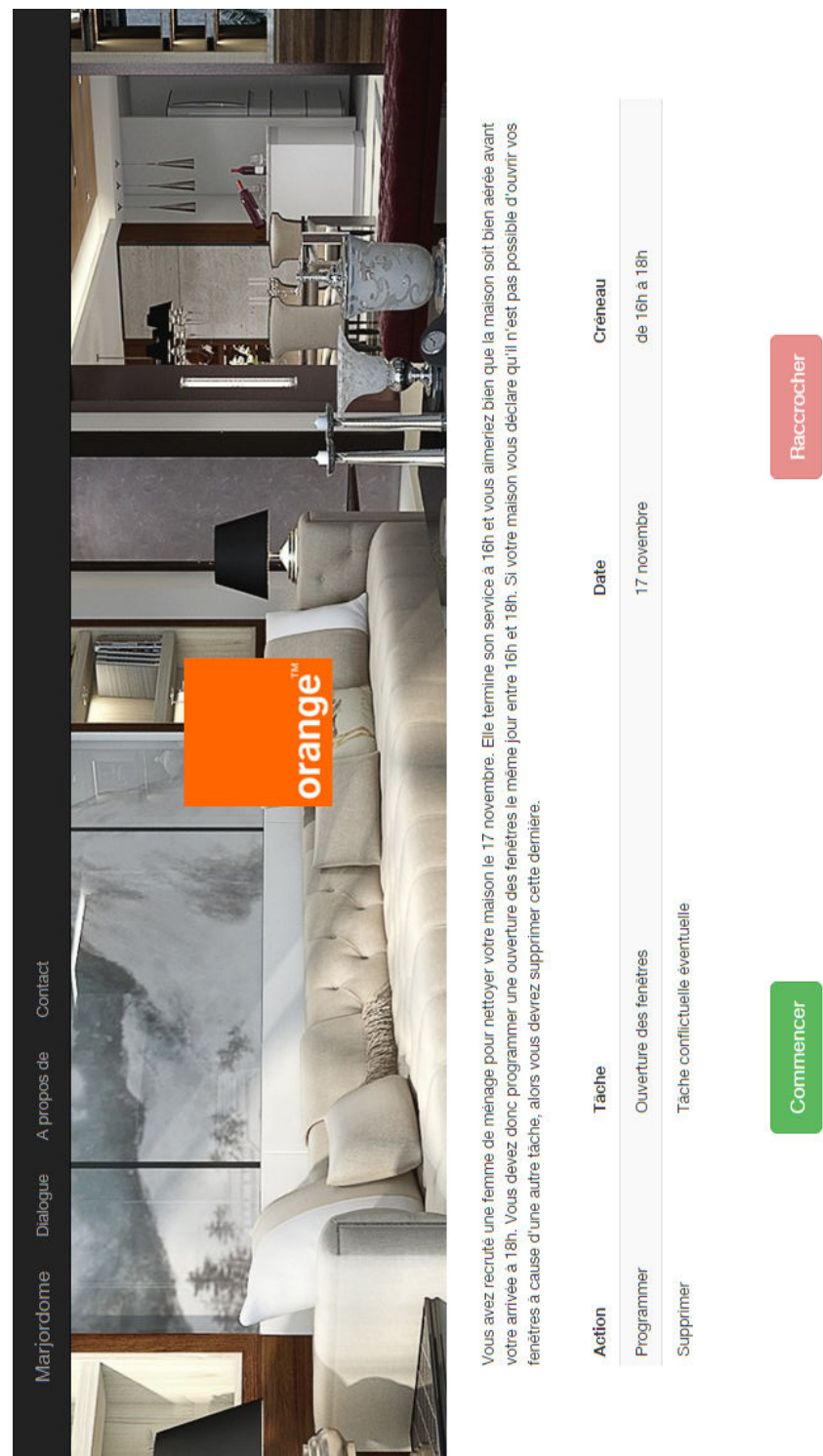


Figure 9.1: The Majordomo interface

- **Efficiency:** The users are asked to assess the dialogue efficiency by selecting one of the 6 possible answers, going from 1 (*very bad*) to 6 (*very good*).
- **Global quality:** The users are asked how they globally appreciated the dialogue on a scale from 1 (*Very unpleasant experience*) to 6 (*Very enjoyable experience*).
- **Potential user:** Finally, the users are also asked whether they would use the Majordomo at home (if it was a real commercialised product) on a scale from 1 (*clearly not*) to 4 (*absolutely*).

### 9.3 Results and discussion

206 dialogues have been collected with the collaboration of 47 volunteer users (from Orange Labs, LIA and personal network). 65 dialogues were run using the non-incremental strategy (*None*), 65 using the handcrafted strategy (*Handcrafted*) and 76 using the reinforcement learning one (*RL*). The experiment results are depicted in Table 9.2.

Category	KPI	<i>None</i>	<i>Handcrafted</i>	<i>RL</i>
Objective	Duration (sec)	94.7	<b>89.6</b>	90.6
	Task completion	0.60	0.63	<b>0.75</b>
Subjective	Reactivity	4.31	4.57	<b>4.62</b>
	Reactivity quality	<b>4.38</b>	4.25	4.36
	Human-likeness	3.63	3.66	<b>3.74</b>
	Efficiency	4.22	4.20	<b>4.36</b>
	Global quality	4.06	4.18	<b>4.20</b>
	Potential use	2.66	2.68	<b>2.82</b>

*Table 9.2: Global dialogue evaluation metrics*

As far as the mean dialogue duration is concerned, the incremental strategies slightly improve the dialogue duration (even though this is not statistically significant). At a first glance, this improvement can be viewed as an obvious result since by construction, incremental strategies are more reactive. Nevertheless, the user can also be interrupted before she has provided all the information that she wanted which complicates the dialogue and makes it last longer like it is the case in the experiment led in (Ghigi et al., 2014). Therefore, this result shows that the Majordomo successfully interrupted the user on average. However, there is no visible change when comparing the *Handcrafted* and the *RL* strategy.

The task completion ratio, on the other hand, has been significantly<sup>2</sup> improved by the *RL* strategy compared to *None* (by 15% with  $p = 0.030$ ). Moreover, an important difference has been reported between *RL* and *Handcrafted* (12%) even though it is not

<sup>2</sup>All the p-values are computed according to the Welch t-test since the number of samples is important enough for the means to be considered as following a normal distribution (and since it is more powerful than non-parametric tests). A binomial proportions test has also been run for task completion, leading to very similar p-values.

exactly statistically significant ( $p = 0.065$ ). Finally, *Handcrafted* shows a minor improvement over *None* (3%) with no statistical significance ( $p = 0.36$ ). The Majordomo task requires a certain level of engagement and focus in order to keep track of all what has been accomplished so far, while keeping the final objective in mind. When interacting with a reactive system that takes the floor in an intelligent way (to correct errors hence fixing desynchronisations, to deliver a response when all the information has been provided...) without overwhelming the user, the latter feels more engaged in the conversation thus accomplishing the task more efficiently, even when a certain cognitive load is involved. Another impact of such strategy as reported in (Ghigi et al., 2014) is that when the users realise that they can be interrupted in case of a problem, they tend to provide more concise and focused answers, which reduces the risk of a misunderstanding.

The subjective metrics are the noisier ones. Therefore, except from the *Reactivity* KPI where *RL* significantly improves it in comparison with *None* ( $p = 0.048$ ), the other *p*-values are above 0.05. Nevertheless, generally speaking, the metrics tend to favour the *RL* strategy.

KPI	<i>None</i>	<i>Handcrafted</i>	<i>RL</i>
Latency (ms)	1545 $\pm$ 61	1303 $\pm$ 78	<b>588 <math>\pm</math> 59</b>
FC ratio	No barge-in	0.31 $\pm$ 0.091	<b>0.068 <math>\pm</math> 0.023</b>

**Table 9.3:** Local dialogue evaluation metrics

To complete this study, more local metrics have been investigated: the latency and the false cut-in ratio. The latency is the mean delay involved in each human to machine floor transition<sup>3</sup> (549 transitions for *None*, 542 for *Handcrafted* and 727 for *RL*) whereas the false cut-in ratio refers to the proportion over all the SPEAK decisions (*Handcrafted* : 99, *RL* : 456) of the ones where the system should have waited longer before taking the floor (manually annotated). The results are reported in Table 9.3 (all the differences are statistically significant with  $p < 0.000001$ ). The *Handcrafted* strategy reduces the latency by 200ms compared to *None*, however, it interrupts the user too soon one third of the time thus being too aggressive. On the other hand, *RL* reduces the latency by 1 second while maintaining a more reasonable false cut-in ratio. As a consequence, *RL* takes more risk since it chooses to SPEAK more often and when it does, it is better managed (significantly less frequent false cut-ins).

## 9.4 Discussion

In comparison with previous work (see Chapters 1 and 2), this is the first direct application of reinforcement learning to turn-taking management in an incremental dialogue system that is evaluated in real conditions. In many previous studies, an indirect way

<sup>3</sup>Estimated as the delay between the moment when the Scheduler delivers the last message and the moment when it received the last ASR output. Computing the real latency requires a Voice Activity Detection (VAD) module in the Client which could slow it down.

of testing dialogue strategies is used: controlled dialogue acts (Aist et al., 2007), a posteriori evaluation using recordings of interactions (Meena et al., 2013), a posteriori comparison with human decision corpus (Jonsdottir et al., 2008; Dethlefs et al., 2012), etc... As it is said in (Aist et al., 2007), the objective is *to minimize variance due to extraneous factors such as interspeaker variability, acoustic noise, and so forth and concentrate specifically on the difference between incremental processing and its nonincremental counterpart*. However, the price to pay to reduce variance is a certain bias due to the fact that the experiment is not run in real conditions.

Some papers use handcrafted strategies (Raux and Eskenazi, 2009; Ghigi et al., 2014), some collect annotated corpora on which they run supervised learning algorithms (Meena et al., 2013) and others propose reinforcement learning based strategies (Jonsdottir et al., 2008; Selfridge and Heeman, 2010; Dethlefs et al., 2012). However, to our knowledge, live studies only fit in the first two categories and no purely autonomous system using reinforcement learning has been tested with real users and directly evaluated by them, in real dialogue conditions. More generally, previous work related to incremental dialogue processing and turn-taking optimisation can be split into two categories given the metrics that are involved:

- **Local metrics:** These studies are based on the principle introduced in (Sacks et al., 1974) and saying that gaps and overlaps should be minimised in order to achieve smooth turn-taking. As a consequence, local metrics where only floor transitions are considered are used, mainly the latency and the false cut-in ratio (Jonsdottir et al., 2008; Raux and Eskenazi, 2012).
- **Global metrics:** Considering the overall dialogue quality can also be a way of evaluating turn-taking strategies (Selfridge and Heeman, 2010; Ghigi et al., 2014). Such an approach has the advantage of not having to make any assumption about what would make the dialogue more appealing for the user. However, the metrics involved are more difficult to measure since they are noisier.

In this thesis, global metrics were used for training then both global and local metrics were evaluated. Interestingly, it is shown that by optimising global KPIs, local ones turn out to be improved as well. Finally, even though it has been shown that interrupting the user can hurt its opinion on the system in some cases (Hirasawa et al., 1999), the results described above show that when it is done at the right moment, the system is significantly more efficient while being judged slightly better from a subjective point of view.



## Conclusion and future work

Viewed as a whole, the several contributions made in this thesis constitute a thorough methodology to enhance turn-taking capabilities of spoken dialogue systems. First, turn-taking mechanisms involved in human conversation are analysed which led to the establishment of a new turn-taking phenomena taxonomy. Compared to existing classifications in the literature, new analysis dimensions where the meaning behind the dialogue participants' behaviours, as well as the motivations behind them, are considered. This leads to a more fine-grained taxonomy which is relevant from the human-machine dialogue point of view. In addition, this constitutes the starting point from which the turn-taking phenomena that are the more likely to improve the dialogue efficiency are chosen.

An incremental dialogue system can be built from scratch using an incremental version of each component in the dialogue chain. In this thesis, an alternative approach is proposed: a dialogue system is split into a Client and a Service part, then a new module, called the Scheduler, is inserted between the two. This new interface plays the role of a turn-taking manager and makes the set {Scheduler+Service} behaves like an incremental system from the Client's point of view. Two advantages are associated with this new approach: first, it makes it possible to transform an existing traditional dialogue system into an incremental one at a low cost and second, the traditional dialogue management part is clearly separated from the turn-taking management one and can be kept almost unchanged.

Based on this new architecture, an incremental dialogue simulator, unique of its kind, has been implemented. It is able to generate dialogues in a personal agenda management domain. A User Simulator, coupled with an ASR Output Simulator that replicates ASR imperfections and instability, sends incremental requests to the Scheduler which decides when to take the floor to provide a response. A first simulation study where several slot-filling strategies along with two turn-taking strategies (non-incremental and handcrafted incremental) showed that the mixed-initiative one along with incremental processing achieves the best performance in terms of dialogue duration and task completion.

Since handcrafting turn-taking strategies requires the designers to empirically set all the parameters and doing so for each new condition (task, language, etc.), this approach is not guaranteed to be optimal while requiring important labour and time resources. On the other hand, data-driven techniques make it possible to build optimal strate-

gies at lower costs. In the field of human-machine dialogue, reinforcement learning has been proven to be particularly useful since no annotation effort is required (unlike supervised techniques) and it is able to learn from delayed rewards (thus making it possible to take the whole dialogue quality as a reward function). In this thesis, a new reinforcement learning turn-taking strategy is proposed and trained using the previous dialogue simulator. In simulation, it has been shown to reduce dialogue duration while improving the task completion ratio when compared to the non-incremental and the handcrafted incremental baselines.

Finally, the previous strategies have been transposed to a new domain, the Marjoromo, then they have been tested through real users interactions. The handcrafted and the reinforcement learning strategies slightly reduce the dialogue duration, but the latter significantly improves the task completion ratio (by 15% compared to the non-incremental strategy). Also, compared with the handcrafted strategy, the data-driven one takes more risk by deciding to speak before a user silence (hence significantly reducing the response latency) while maintaining a low false cut-in rate of 6.8% instead of 31%.

To conclude, from a general point of view, this thesis provides new evidence showing the potential of incremental dialogue processing. More particularly, it goes a step further by showing that optimal turn-taking can be learnt automatically during the interactions. Also, the proposed architecture and the generality of the feature used for learning makes it possible to easily transfer this work to any domain and to bootstrap from existing dialogue systems.

During the course of this thesis, two questions were also tackled but due to time constraints, they are still under investigation and there is still work to be done in order to hopefully come up with interesting results. These questions are:

- How to adapt the exploration/exploitation mechanism in reinforcement learning to the case of incremental processing?
- How to learn both optimal dialogue management and optimal turn-taking decisions at the same time?

**Exploration/exploitation:** The exploration/exploitation dilemma in reinforcement learning is a research problem in itself. Inspired by the multi-armed bandit literature, the most simple and naive approaches are  $\epsilon$ -greedy and Boltzmann exploration (Softmax) (Sutton and Barto, 1998) (some other simple approaches also exist in the literature but they are less used and they are based on similar ideas). A new approach (initially proposed to solve the multi-armed bandit problem) called Upper Confidence Bound (UCB) (Auer et al., 2002) is also frequently used since it improves the convergence rate. This algorithm has also inspired other algorithms which are more adapted to reinforcement learning like UCRL (Auer and Ortner, 2006). Nevertheless, incremental processing raises a new challenge which is not tackled in existing approaches: in this case, the optimal policy is not balanced when it comes to the number of times each action has to be chosen. The action which consists in doing nothing (called WAIT in this thesis)



and wait for further information coming from the ASR should be picked the majority of the time. Therefore, incremental processing involves long episodes where the system should perform one action most of the time and rarely pick an alternative action. As a consequence, during the exploration process using existing methods, the agent rarely performs the right decisions at the right time. Therefore, there is a need to come up with a more adapted and more efficient way to deal with the exploration/exploitation dilemma in this case. In Chapter 8, an  $\epsilon$ -greedy policy with a bias in favour of the WAIT action has been proven to be successful in this case. However, there is no guarantee that this is the optimal way to proceed and more importantly, it is likely that this method would perform poorly if the micro-turn duration is increasing (imagine a setup with a new micro-turn every millisecond). To investigate this problem, two ideas have been proposed during this thesis:

- **SMDP-based approach:** Semi-Markov Decision Processes (SMDPs) (Bradtke and Duff, 1994) are different from conventional MDPs in the sense that time is continuous and the time interval between two decisions is not constant (the instants in which the agent is asked to make decisions is defined by a separate process). This is an interesting framework when it comes to incremental dialogue processing and the idea is to cast the Scheduler as an SMDP that is asked to make decisions at specific instants (the rest of the time, it is always picking the WAIT action). Most of the time, when the WAIT action is picked, the state associated with the next micro-turn is very similar to the state associated with the current state. Therefore, the agent will be asked to make decisions only when the state varies in a significant way or in other words, when the state variation given a specific metric reaches a certain threshold. The investigated method tries to learn an optimal value for this threshold in an online fashion.
- **DPS-based approach:** Direct Policy Search (DPS) consists in using classical optimisation methods directly over the policy space unlike conventional reinforcement learning algorithms where a Q-function is computed and then the policy is derived from it. The advantage of such methods is that they naturally embed the mechanism described in the previous point: when two states are similar, a policy is most likely to associate the same action to both of them.

**DM-Scheduler co-learning:** The approach depicted in Chapter 8 assumes that the dialogue manager (in the Service) has a constant behaviour. However, a very active research thread proposed many methods where this module is constantly improving its decisions mostly using reinforcement learning. Considering that, a natural and legitimate question comes into play: is it possible for both the turn-taking manager and the conventional dialogue manager to simultaneously learn optimal behaviours from interactions. Such a study can benefit from the large existing literature about multi-agent collaborative learning (Claus and Boutilier, 1998; Panait and Luke, 2005; Vogel et al., 2013).





# List of Figures

1.1	The dialogue chain . . . . .	30
1.2	A 5-Best example corresponding to the sentence “I would like to book a flight from New-York to Los Angeles”. . . . .	31
2.1	The interaction cycle between the agent and the environment in reinforcement learning . . . . .	43
4.1	The Scheduler: an interface between the client and the service . . . . .	76
4.2	Time sharing in traditional settings . . . . .	77
4.3	Time sharing in incremental settings . . . . .	77
4.4	Incremental request processing with the Scheduler: the conditions on the left side trigger the Client to send partial requests . . . . .	78
4.5	The incremental version of the CFAsT project. The traditional view is represented on the left and the new incremental one is depicted on the left. . . . .	80
5.1	Slot-filling strategies efficiency comparison ( $n_s = 5$ ) . . . . .	94
6.1	Simulated environment architecture . . . . .	100
6.2	ASR score sampling distribution ( $\sigma_{conf} = 1$ ) . . . . .	103
6.3	An illustration of the incremental ASR output N-Best update (BF=0.2) . . . . .	105
7.1	Simulated mean duration (left) and dialogue task completion (right) for different noise levels . . . . .	116
7.2	Mean dialogue duration and task completion for aggregated strategies. . . . .	119
7.3	Mean dialogue duration and task completion for different turn-taking phenomena. . . . .	119
7.4	INCOHERENCE_INTERP evaluated in a more adapted task . . . . .	120
8.1	Learning curve (0-500: pure exploration, 500-2500: exploration/exploitation, 2500-3000: pure exploitation) with WER= 0.15. . . . .	131
8.2	Mean dialogue duration and task for the non-incremental, the baseline incremental and the RL incremental (after convergence) strategies under different noise conditions. . . . .	131
9.1	The Majordomo interface . . . . .	136



# List of Tables

1.1	Example of incremental NLU processing . . . . .	35
3.1	Taxonomy labels . . . . .	59
3.2	Turn-taking phenomena taxonomy. The rows/columns correspond to the levels of information added by the floor holder/taker. . . . .	60
3.3	Taxonomy labels (1/3) . . . . .	68
3.4	Taxonomy labels (2/3) . . . . .	69
3.5	Taxonomy labels (3/3) . . . . .	70
4.1	A double context: the real context and the simulation context. . . . .	78
4.2	Available features for dialogue systems given the way they integrate incrementality . . . . .	85
6.1	NLU rules types . . . . .	107
6.2	User simulator and ASR output simulator values . . . . .	109
7.1	The three scenarios used in the simulation . . . . .	116
8.1	Dialogue example with state representation features' values. . . . .	127
9.1	Majordomo tasks . . . . .	134
9.2	Global dialogue evaluation metrics . . . . .	137
9.3	Local dialogue evaluation metrics . . . . .	138



# Bibliography

- (Aist et al., 2007) G. Aist, J. Allen, E. Campana, C. G. Gallo, S. Stoness, M. Swift, and M. K. Tanenhaus, 2007. Incremental understanding in human-computer dialogue and experimental evidence for advantages over nonincremental methods. In Proceedings of *the 11th Workshop on the Semantics and Pragmatics of Dialogue*.
- (Anand and Kahn, 1992) T. Anand and G. S. Kahn, 1992. Making sense of gigabytes: A system for knowledge-based market analysis. In Proceedings of *Innovative Applications of Artificial Intelligence (IAAI)*.
- (Arnold et al., 2000) J. E. Arnold, J. G. Eisenband, S. Brown-Schmidt, and J. C. Trueswell, 2000. The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition* 76(1), B13–B26.
- (Asri et al., 2016) L. E. Asri, B. PIOT, M. Geist, R. Laroche, and O. Pietquin, 2016. Score-based inverse reinforcement learning. In Proceedings of *International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)*.
- (Auer et al., 2002) P. Auer, N. Cesa-Bianchi, and P. Fischer, 2002. Finite-time analysis of the multiarmed bandit problem. *Machine Learning Journal* 47(2-3), 235–256.
- (Auer and Ortner, 2005) P. Auer and R. Ortner, 2005. Online-regret bounds for a new reinforcement-learning algorithm. In Proceedings of *1st Austrian Cognitive Vision Workshop*.
- (Auer and Ortner, 2006) P. Auer and R. Ortner, 2006. Logarithmic online regret bounds for undiscounted reinforcement learning. In Proceedings of *Advances in Neural Information Processing Systems (NIPS)*.
- (Austin, 1962) J. Austin, 1962. *How to Do Things with Words*. Oxford University Press.
- (Barlier et al., 2015) M. Barlier, R. Laroche, and O. Pietquin, 2015. Human-machine dialogue as a stochastic game. In Proceedings of *the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Barto and Mahadevan, 2003) A. G. Barto and S. Mahadevan, 2003. Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems* 13, 41–77.

- (Baumann, 2008) T. Baumann, 2008. Simulating spoken dialogue with a focus on realistic turn-taking. In Proceedings of *the 5th International Workshop on Constraints and Language Processing (CSLP)*.
- (Baumann, 2014) T. Baumann, 2014. Partial representations improve the prosody of incremental speech synthesis. In Proceedings of *Interspeech*.
- (Baumann and Schlangen, 2011) T. Baumann and D. Schlangen, 2011. Evaluation and optimisation of incremental processors. *Dialogue and Discourse* 2, 113–141.
- (Baumann and Schlangen, 2013) T. Baumann and D. Schlangen, 2013. Open-ended, extensible system utterances are preferred, even if they require filled pauses. In Proceedings of *the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Beattie, 1982) G. Beattie, 1982. Turn-taking and interruption in political interviews: Margaret thatcher and jim callaghan compared and contrasted. *Semiotica* 39, 93–114.
- (Bellman, 1957) R. Bellman, 1957. *Dynamic Programming*. Princeton University Press.
- (Ben Jannet et al., 2015) M. A. Ben Jannet, O. Galibert, M. Adda-Decker, and S. Rosset, 2015. How to evaluate asr errors impact on ner? In Proceedings of *Errors by Humans and Machines in Multimedia, Multimodal, Multilingual Data Processing (ERRARE)*.
- (Bengio et al., 2003) Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, 2003. A neural probabilistic language model. *The Journal of Machine Learning Research* 3, 1137–1155.
- (Berry and Fristedt, 1985) D. A. Berry and B. Fristedt, 1985. *Bandit problems : sequential allocation of experiments*. Chapman and Hall London.
- (Bradtke and Duff, 1994) S. J. Bradtke and M. O. Duff, 1994. Reinforcement learning methods for continuous-time markov decision problems. In Proceedings of *Advances in Neural Information Processing Systems (NIPS)*.
- (Bratman, 1987) M. Bratman, 1987. *Intention, plans, and practical reason*. Harvard University Press.
- (Bratman et al., 1988) M. E. Bratman, D. J. Israel, and M. E. Pollack, 1988. Plans and resource-bounded practical reasoning. *Computational Intelligence* 4, 349–355.
- (Breslin et al., 2013) C. Breslin, M. Gasic, M. Henderson, D. Kim, M. Szummer, B. Thomson, P. Tsiakoulis, and S. Young, 2013. Continuous asr for flexible incremental dialogue. In Proceedings of *ICASSP*, 8362–8366.
- (Bretier et al., 2010) P. Bretier, R. Laroche, and G. Putois, 2010. D5.3.4: Industrial self-help system ("system 3") adapted to final architecture. Prototype D5.3.4, CLASSIC Project.
- (Bubeck and Cesa-Bianchi, 2012) S. Bubeck and N. Cesa-Bianchi, 2012. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning* 5, 1–122.

- (Chandramohan et al., 2011) S. Chandramohan, M. Geist, F. Lefèvre, and O. Pietquin, 2011. User simulation in dialogue systems using inverse reinforcement learning. In Proceedings of *INTERSPEECH*.
- (Chandramohan et al., 2012a) S. Chandramohan, M. Geist, F. Lefèvre, and O. Pietquin, 2012a. Behavior specific user simulation in spoken dialogue systems. In Proceedings of *Speech Communication; 10th ITG Symposium*, 1–4. VDE.
- (Chandramohan et al., 2012b) S. Chandramohan, M. Geist, F. Lefèvre, and O. Pietquin, 2012b. Clustering behaviors of spoken dialogue systems users. In Proceedings of *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 4981–4984. IEEE.
- (Chandramohan et al., 2010) S. Chandramohan, M. Geist, and O. Pietquin, 2010. Optimizing spoken dialogue management with fitted value iteration. In Proceedings of *INTERSPEECH 11th Annual Conference of the International Speech*.
- (Chao and Thomaz, 2012) C. Chao and A. L. Thomaz, 2012. Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets. *Journal of Human-Robot Interaction*.
- (Clark, 1996) H. H. Clark, 1996. *Using Language*. Cambridge University Press.
- (Claus and Boutilier, 1998) C. Claus and C. Boutilier, 1998. The dynamics of reinforcement learning in cooperative multiagent systems. In Proceedings of *the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence, AAAI '98/IAAI '98*.
- (Cohen and Levesque, 1990) P. R. Cohen and H. J. Levesque, 1990. Persistence, intention, and commitment. In Proceedings of *Intentions in Communication*.
- (Collobert et al., 2011) R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. P. Kuksa, 2011. Natural language processing (almost) from scratch. *CoRR abs/1103.0398*.
- (Cuayáhuitl et al., 2015) H. Cuayáhuitl, S. Keizer, and O. Lemon, 2015. Strategic dialogue management via deep reinforcement learning. In Proceedings of *NIPS Workshop on Deep Reinforcement Learning*.
- (Cuayáhuitl et al., 2007) H. Cuayáhuitl, S. Renals, O. Lemon, and H. Shimodaira, 2007. Hierarchical dialogue optimization using semi-markov decision processes. In Proceedings of *Interspeech*.
- (Daubigney et al., 2013) L. Daubigney, M. Geist, and O. Pietquin, 2013. Model-free pomdp optimisation of tutoring systems with echo-state networks. In Proceedings of *the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIG-DIAL)*.



- (Deng et al., 2013) L. Deng, J. Li, J.-T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, Y. Gong, and A. Acero, 2013. Recent advances in deep learning for speech research at microsoft. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
- (Dethlefs et al., 2012) N. Dethlefs, H. W. Hastie, V. Rieser, and O. Lemon, 2012. Optimising incremental dialogue decisions using information density for interactive systems. In *Proceedings of EMNLP-CoNLL*.
- (Doya, 2007) K. Doya, 2007. Reinforcement learning: Computational theory and biological mechanisms. *HFSP journal* 1, 30–40.
- (Dudley, 1939) H. Dudley, 1939. The vocoder. *Bell Labs Record* 17, 122–126.
- (Duncan, 1972) S. Duncan, 1972. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology* 23, 283–292.
- (Dybkjaer et al., 2004) L. Dybkjaer, N. O. Bernsen, and W. Minker, 2004. Evaluation and usability of multimodal spoken language dialogue systems. In *Proceedings of Speech Communications*, 33–54.
- (Eberhard et al., 1995) K. Eberhard, M. Spivey-Knowlton, J. Sedivy, and M. Tanenhaus, 1995. Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research* 24(6), 409–436.
- (Eckert et al., 1997) W. Eckert, E. Levin, and R. Pieraccini, 1997. User modeling for spoken dialogue system evaluation. In *Proceedings of Automatic Speech Recognition and Understanding*.
- (Edlund et al., 2008) J. Edlund, J. Gustafson, M. Heldner, and A. Hjalmarsson, 2008. Towards human-like spoken dialogue systems. *Speech Communication* 50, 630–645.
- (Ekeinhor-Komi et al., 2014) T. Ekeinhor-Komi, H. Falih, C. Chardenon, R. Laroche, and F. Lefèvre, 2014. Un assistant vocal personnalisable. In *Proceedings of the 21st Conference sur le Traitement Automatique des Langues Naturelles (TALN)*.
- (El Asri et al., 2014) L. El Asri, H. Khouzaimi, R. Laroche, and O. Pietquin, 2014. Ordinal regression for interaction quality prediction. In *Proceedings of Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*.
- (El Asri and Laroche, 2013) L. El Asri and R. Laroche, 2013. Will my spoken dialogue system be a slow learner ? In *Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (El Asri et al., 2012) L. El Asri, R. Laroche, and O. Pietquin, 2012. Reward function learning for dialogue management. In *Proceedings of STAIRS*.
- (El Asri et al., 2014) L. El Asri, R. Laroche, and O. Pietquin, 2014. DINASTI: Dialogues with a Negotiating Appointment Setting Interface. In *Proceedings of Proc. of LREC*.

- (El Asri et al., 2016) L. El Asri, R. Laroche, and O. Pietquin, 2016. Compact and interpretable dialogue state representation with genetic sparse distributed memory. In Proceedings of *International Workshop Series on Spoken Dialogue Systems Technology (IWSDS)*.
- (El Asri et al., 2014) L. El Asri, R. Lemonnier, R. Laroche, O. Pietquin, and H. Khouzami, 2014. NASTIA: Negotiating Appointment Setting Interface. In Proceedings of *LREC*.
- (Evanini et al., 2008) K. Evanini, P. Hunter, J. Liscombe, D. Suendermann, K. Dayanidhi, and R. Pieraccini, 2008. Caller experience: A method for evaluating dialog systems and its automatic prediction. In Proceedings of *Spoken Language Technology Workshop, 2008. SLT 2008. IEEE*.
- (Ferguson and Allen, 2007) G. Ferguson and J. F. Allen, 2007. Mixed-initiative systems for collaborative problem solving. *AI Magazine* 28.
- (Ferreira et al., 2015a) E. Ferreira, B. Jabaian, and F. Lefèvre, 2015a. Online adaptive zero-shot learning spoken language understanding using word-embedding. In Proceedings of *ICASSP*.
- (Ferreira et al., 2015b) E. Ferreira, B. Jabaian, and F. Lefèvre, 2015b. Zero-shot semantic parser for spoken language understanding. In Proceedings of *INTERSPEECH*.
- (Ferreira and Lefèvre, 2013) E. Ferreira and F. Lefèvre, 2013. Social signal and user adaptation in reinforcement learning-based dialogue management. In Proceedings of *MLIS*.
- (Ferreira and Lefèvre, 2015) E. Ferreira and F. Lefèvre, 2015. Reinforcement-learning based dialogue system for human-robot interactions with socially-inspired rewards. *Computer Speech & Language* 34(1), 256–274.
- (Ferreira et al., 2015c) E. Ferreira, G. Milliez, F. Lefèvre, and R. Alami, 2015c. Users’ belief awareness in reinforcement learning-based situated human-robot dialogue management. In Proceedings of *IWSDS*.
- (Ferreira et al., 2016) E. Ferreira, A. Reiffers-Masson, B. Jabaian, and F. Lefèvre, 2016. Adversarial bandit for online interactive active learning of zero-shot spoken language understanding. In Proceedings of *IEEE International Conference on Acoustics Speech and Signal Processing*.
- (Fink et al., 1998) G. A. Fink, C. Schillo, F. Kummert, and G. Sagerer, 1998. Incremental speech recognition for multimodal interfaces. In Proceedings of *IECON*.
- (Gales and Young, 2007) M. Gales and S. Young, 2007. The application of hidden markov models in speech recognition. *Found. Trends Signal Process.* 1.
- (Galibert et al., 2005) O. Galibert, G. Illouz, and S. Rosset, 2005. An open-domain, human-computer dialog system. In Proceedings of *INTERSPEECH*.

- (Gasic et al., 2013) M. Gasic, C. Breslin, M. Henderson, D. Kim, M. Szummer, B. Thomson, P. Tsiakoulis, and S. Young, 2013. Pomdp-based dialogue manager adaptation to extended domains. In *Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Gelin et al., 2010) R. Gelin, C. d'Alessandro, Q. A. Le, O. Deroo, J.-C. Doukhan, Davidand Martin, C. Pelachaud, A. Rilliard, and S. Rosset, 2010. Towards a storytelling humanoid robot. In *Proceedings of Workshop "Dialog with Robots", AAAI Fall Symposium*.
- (Ghigi et al., 2014) F. Ghigi, M. Eskenazi, M. I. Torres, and S. Lee, 2014. Incremental dialog processing in a task-oriented dialog. In *Proceedings of Fifteenth Annual Conference of the International Speech Communication Association*.
- (Gravano and Hirschberg, 2011) A. Gravano and J. Hirschberg, 2011. Turn-taking cues in task-oriented dialogue. *Comput. Speech Lang.* 25(3), 601–634.
- (Hahn et al., 2010) S. Hahn, M. Dinarelli, C. Raymond, F. Lefèvre, P. Lehnen, R. De Mori, A. Moschitti, H. Ney, and G. Riccardi, 2010. Comparing stochastic approaches to spoken language understanding in multiple languages. *IEEE TASLP* 19(6), 1569–1583.
- (Hastie et al., 2013) H. Hastie, M.-A. Aufaure, P. Alexopoulos, H. Cuayáhuítl, N. Dethlefs, M. Gasic, J. Henderson, O. Lemon, X. Liu, P. Mika, N. Ben Mustapha, V. Rieser, B. Thomson, P. Tsiakoulis, and Y. Vanrompay, 2013. Demonstration of the parlance system: a data-driven incremental, spoken dialogue system for interactive search. In *Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Hirasawa et al., 1999) J. Hirasawa, M. Nakano, T. Kawabata, and K. Aikawa, 1999. Effects of system barge-in responses on user impressions. In *Proceedings of EUROSPEECH*.
- (Hone and Graham, 2000) K. S. Hone and R. Graham, 2000. Towards a tool for the subjective assessment of speech system interfaces (sassi). *Natural Language Engineering* 6, 287–303.
- (Huet et al., 2013) S. Huet, E. Manishina, and F. Lefèvre, 2013. Factored machine translation systems for russian-english. *the Eighth Workshop on Statistical Machine Translation*, 152–155.
- (Ilkin and Sturt, 2011) Z. Ilkin and P. Sturt, 2011. Active prediction of syntactic information during sentence processing. *Dialogue and Discourse* 2, 35–58.
- (Israel and Mitchell, 1978) J. E. Israel and J. G. Mitchell, 1978. Separating data from function in a distributed file system. In *Proceedings of Second International Symposium on Operating Systems Theory and Practice*.

- (Jabaian et al., 2013) B. Jabaian, L. Besacier, and F. Lefèvre, 2013. Comparison and combination of lightly supervised approaches for language portability of a spoken language understanding system. *Audio, Speech, and Language Processing, IEEE Transactions on* 21(3), 636–648.
- (Jabaian et al., 2016) B. Jabaian, F. Lefèvre, and L. Besacier, 2016. A unified framework for translation and understanding allowing discriminative joint decoding for multilingual speech semantic interpretation. *Computer Speech & Language* 35, 185–199.
- (Jaffe and Feldstein, 1970) J. Jaffe and S. Feldstein, 1970. *Rhythms of dialogue*. Academic Press New York.
- (Jiang, 2005) H. Jiang, 2005. Confidence measures for speech recognition: A survey. *Speech communication* 45, 455–470.
- (Johnston et al., 2014) M. Johnston, J. Chen, P. Ehlen, H. Jung, J. Lieske, A. Reddy, E. Selfridge, S. Stoyanchev, B. Vasilieff, and J. Wilpon, 2014. Mva: The multimodal virtual assistant. In Proceedings of *the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Jonsdottir et al., 2008) G. R. Jonsdottir, K. R. Thorisson, and E. Nivel, 2008. Learning smooth, human-like turntaking in realtime dialogue. In Proceedings of *Intelligent Virtual Agents (IVA 08)*, 162–175. Springer.
- (Jordan et al., 2015) P. Jordan, P. Albacete, and S. Katz, 2015. Exploring the effects of redundancy within a tutorial dialogue system: Restating students’ responses. In Proceedings of *the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Khouzaimi et al., 2014a) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2014a. Dictanum : système de dialogue incrémental pour la dictée de numéros. In Proceedings of *the TALN 2014 Conference*.
- (Khouzaimi et al., 2014b) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2014b. An easy method to make dialogue systems incremental. In Proceedings of *the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Khouzaimi et al., 2015a) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2015a. Dialogue efficiency evaluation of turn-taking phenomena in a multi-layer incremental simulated environment. In Proceedings of *the HCI International 2015 Conference*.
- (Khouzaimi et al., 2015b) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2015b. Optimising turn-taking strategies with reinforcement learning. In Proceedings of *the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Khouzaimi et al., 2015c) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2015c. Turn-taking phenomena in incremental dialogue systems. In Proceedings of *the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

- (Khouzaimi et al., 2016a) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2016a. Incremental human-machine dialogue simulation. In Proceedings of *the International Workshop on Spoken Dialogue Systems (IWSDS)*.
- (Khouzaimi et al., 2016b) H. Khouzaimi, R. Laroche, and F. Lefèvre, 2016b. Reinforcement learning for turn-taking management in incremental spoken dialogue systems. In Proceedings of *International Joint Conference on Artificial Intelligence (IJCAI)*.
- (Kilger and Finkler, 1995) A. Kilger and W. Finkler, 1995. Incremental generation for real-time applications. Rapport technique, German Research Center for Artificial Intelligence.
- (Kim and Banchs, 2014) S. Kim and R. E. Banchs, 2014. Sequential labeling for tracking dynamic dialog states. In Proceedings of *the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Konolige and Pollack, 1993) K. Konolige and M. E. Pollack, 1993. A representationalist theory of intention. In Proceedings of *the 13th International Joint Conference on Artificial Intelligence (IJCAI)*.
- (Lamel et al., 2000) L. Lamel, S. Rosset, J.-L. Gauvain, S. Bennacef, M. Garnier-Rizet, and B. Prouts, 2000. The LIMSI ARISE System. *Speech Communication* 31(4), 339–354.
- (Laroche, 2010) R. Laroche, 2010. *Raisonnement sur les incertitudes et apprentissage pour les systemes de dialogue conventionnels*. Phd Thesis, Paris VI University.
- (Laroche, 2014) R. Laroche, 2014. CFAsT : Content Finder AssistanT. In Proceedings of *the 21st Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*.
- (Laroche, 2015) R. Laroche, 2015. Content finder assistant. In Proceedings of *the 18th International Conference on Intelligence in Next Generation Networks (ICIN)*.
- (Laroche et al., 2008) R. Laroche, B. Bouchon-Meunier, and P. Bretier, 2008. Uncertainty management in dialogue systems. In Proceedings of *European Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*.
- (Laroche et al., 2011) R. Laroche, G. Putois, P. Bretier, M. Aranguren, J. Velkovska, H. Hastie, S. Keizer, K. Yu, F. Jurcicek, O. Lemon, and S. Young, 2011. D6.4: Final evaluation of classic towninfo and appointment scheduling systems. Report D6.4, CLASSIC Project.
- (Laroche et al., 2013) R. Laroche, L. Roussarie, P. Baczyk, and J. Dziekan, 2013. Cooking coach spoken/multimodal dialogue systems. In Proceedings of *the IJCAI Workshop on Cooking with Computers*.
- (Lefèvre et al., 2012) F. Lefèvre, D. Mostefa, L. Besacier, Y. Estève, M. Quignard, N. Camelin, B. Favre, B. Jabaian, and L. M. Rojas-Barahona, 2012. Leveraging study of robustness and portability of spoken language understanding systems across languages and domains: the portmedia corpora. In Proceedings of *the Eight International Conference on Language Resources and Evaluation (LREC'12)*.

- (Lemon et al., 2003) O. Lemon, L. Cavedon, and B. Kelly, 2003. Managing dialogue interaction: A multi-layered approach. In Proceedings of *the 4th SIGDIAL Workshop on Discourse and Dialogue*.
- (Lemon et al., 2006) O. Lemon, K. Georgila, J. Henderson, and M. Stuttle, 2006. An interactive dialogue system exhibiting reinforcement learning of dialogue policies: Generic slot-filling in the talk in-car system. In Proceedings of *the 11th Conference of the European Chapter of the Association for Computational Linguistics, EACL '06*, 119–122.
- (Lemon and Pietquin, 2007) O. Lemon and O. Pietquin, 2007. Machine learning for spoken dialogue systems. In Proceedings of *the European Conference on Speech Communication and Technologies (Interspeech'07)*.
- (Levelt, 1989) W. J. M. Levelt, 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- (Levin and Pieraccini, 1997) E. Levin and R. Pieraccini, 1997. A stochastic model of computer-human interaction for learning dialogue strategies. In Proceedings of *In EUROSPEECH 97*.
- (Levin et al., 1997) E. Levin, R. Pieraccini, and W. Eckert, 1997. Learning dialogue strategies within the markov decision process framework. In Proceedings of *Automatic Speech Recognition and Understanding, 1997. Proceedings., 1997 IEEE Workshop on*.
- (Lock, 1965) K. Lock, 1965. Structuring programs for multiprogram time-sharing on-line applications. In Proceedings of *AFIPS '65 (Fall, part I)*.
- (Louis, 2002) V. Louis, 2002. *Conception et mise en oeuvre de modèles formels du calcul de plans d'action complexes par un agent rationnel dialoguant*. Phd Thesis, Université de Caen/Basse-Normandie, France.
- (Lu et al., 2011) D. Lu, T. Nishimoto, and N. Minematsu, 2011. Decision of response timing for incremental speech recognition with reinforcement learning. In Proceedings of *ASRU*.
- (Macherey, 2009) K. Macherey, 2009. *Statistical methods in natural language understanding and spoken dialogue systems*. Phd Thesis.
- (Mairesse et al., 2010) F. Mairesse, M. Gašić, F. Jurčiček, S. Keizer, B. Thomson, K. Yu, and S. Young, 2010. Phrase-based statistical language generation using graphical models and active learning. In Proceedings of *the 48th Annual Meeting of the Association for Computational Linguistics*, 1552–1561. Association for Computational Linguistics.
- (Manishina et al., 2016) E. Manishina, B. Jabaian, S. Huet, and F. Lefèvre, 2016. Automatic corpus extension for data-driven natural language generation. In Proceedings of *the Language Resources and Evaluation Conference (LREC)*.

- (Matthias, 2009) G. M. Matthias, 2009. Incremental speech understanding in a multi-modal web-based spoken dialogue system. Mémoire de Master, Massachusetts Institute of Technology.
- (Mccarthy, 1979) J. Mccarthy, 1979. Ascribing mental qualities to machines. In Proceedings of *Philosophical Perspectives in Artificial Intelligence*, 161–195.
- (Mccarthy and Hayes, 1969) J. Mccarthy and P. J. Hayes, 1969. Some philosophical problems from the standpoint of artificial intelligence. In Proceedings of *Machine Intelligence Journal*, 463–502.
- (McGraw and Gruenstein, 2012) I. McGraw and A. Gruenstein, 2012. Estimating word-stability during incremental speech recognition. In Proceedings of *the INTER-SPEECH 2012 Conference*.
- (Meena et al., 2013) R. Meena, G. Skantze, and J. Gustafson, 2013. A data-driven model for timing feedback in a map task dialogue system. In Proceedings of *the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Meyer, 2006) G. Meyer, 2006. *Formalisation logique de préférences qualitatives pour la sélection de la réaction d'un agent rationnel dialoguant*. Phd Thesis, Université de Paris XI.
- (Mitchell et al., 2014) M. Mitchell, D. Bohus, and E. Kamar, 2014. Crowdsourcing language generation templates for dialogue systems. In Proceedings of *the INLG and SIGDIAL 2014 Joint Session*.
- (Mohamed et al., 2009) A. Mohamed, G. Dahl, and G. Hinton, 2009. Deep belief networks for phone recognition. *NIPS 22nd workshop on deep learning for speech recognition*.
- (Möller et al., 2007) S. Möller, P. Smeele, H. Boland, and J. Krebber, 2007. Evaluating spoken dialogue systems according to de-facto standards: A case study. *Computer Speech & Language* 21, 26–53.
- (Mori, 1970) M. Mori, 1970. The uncanny vally. In Proceedings of *Energy*.
- (Newell, 1980) A. Newell, 1980. The knowledge level. *AI Magazine* 2, 1–20.
- (Paek, 2007) T. Paek, 2007. Toward evaluation that leads to best practices: reconciling dialog evaluation in research and industry. In Proceedings of *the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies*, 40–47. Association for Computational Linguistics (ACL).
- (Paek and Pieraccini, 2008) T. Paek and R. Pieraccini, 2008. Automating spoken dialogue management design using machine learning: An industry perspective. *Speech Communication* 50, 716–729.
- (Pakucs, 2003) B. Pakucs, 2003. Towards dynamic multi-domain dialogue processing. In Proceedings of *8th European Conference on Speech Communication and Technology, EUROSPEECH - INTERSPEECH*.

- (Panait and Luke, 2005) L. Panait and S. Luke, 2005. Cooperative multi-agent learning: The state of the art. *Autonomous Agents and Multi-Agent Systems* 11, 387–434.
- (Pieraccini and Huerta, 2005) R. Pieraccini and J. Huerta, 2005. Where do we go from here? research and commercial spoken dialog systems. In *Proceedings of the 6th SIGDIAL Workshop on Discourse and Dialogue*.
- (Pietquin and Beaufort, 2005) O. Pietquin and R. Beaufort, 2005. Comparing asr modeling methods for spoken dialogue simulation and optimal strategy learning. In *Proceedings of 9th European Conference on Speech Communication and Technology (Eurospeech/Interspeech)*.
- (Pietquin and Dutoit, 2006) O. Pietquin and T. Dutoit, 2006. A probabilistic framework for dialog simulation and optimal strategy learning. *IEEE Transactions on Audio, Speech and Language Processing* 14(2), 589–599.
- (Pietquin and Hastie, 2013) O. Pietquin and H. Hastie, 2013. A survey on metrics for the evaluation of user simulations. *The Knowledge Engineering Review* 28.
- (Plátek and Jurčiček, 2014) O. Plátek and F. Jurčiček, 2014. Free on-line speech recogniser based on kaldi asr toolkit producing word posterior lattices. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Potet et al., 2011) M. Potet, R. Rubino, B. Lecouteux, S. Huet, H. Blanchon, L. Besacier, and F. Lefèvre, 2011. The liga (lig/lia) machine translation system for wmt 2011. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, 440–446. Association for Computational Linguistics.
- (Povey et al., 2011) D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, 2011. The kaldi speech recognition toolkit. In *Proceedings of IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*.
- (Putois et al., 2010) G. Putois, R. Laroche, and P. Bretier, 2010. Enhanced monitoring tools and online dialogue optimisation merged into a new spoken dialogue system design experience. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Raux and Eskenazi, 2007) A. Raux and M. Eskenazi, 2007. A multi-layer architecture for semi-synchronous event-driven dialogue management. In *Proceedings of ASRU*.
- (Raux and Eskenazi, 2008) A. Raux and M. Eskenazi, 2008. Optimizing endpointing thresholds using dialogue features in a spoken dialogue system. In *Proceedings of the 9th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Raux and Eskenazi, 2009) A. Raux and M. Eskenazi, 2009. A finite-state turn-taking model for spoken dialog systems. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL '09*, 629–637.



- (Raux and Eskenazi, 2012) A. Raux and M. Eskenazi, 2012. Optimizing the turn-taking behavior of task-oriented spoken dialog systems. *ACM Transactions on Speech and Language Processing* 9(1), 1:1–1:23.
- (Richard Bellman, 1959) S. D. Richard Bellman, 1959. Functional approximations and dynamic programming. *Mathematical Tables and Other Aids to Computation* 13, 247–251.
- (Rieser and Lemon, 2011) V. Rieser and O. Lemon, 2011. Learning and evaluation of dialogue strategies for new applications: Empirical methods for optimization from small data sets. *Computational Linguistics* 37, 153–196.
- (Rosenthal et al., 2013) S. Rosenthal, D. Bohus, E. Kamar, and E. Horvitz, 2013. Look versus leap: Computing value of information with high-dimensional streaming evidence. In Proceedings of *the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*.
- (Rosset et al., 2013) S. Rosset, D. Luzzati, C. Grouin, I. Vasilescu, M. Adda-Decker, E. Bilinski, N. Camelin, J. Kahn, C. Lailler, and L. Lamel, 2013. Human annotation of asr error regions: Is "gravity" a sharable concept for human annotators? In Proceedings of *Errors by Humans and Machines in Multimedia, Multimodal, Multilingual Data Processing (ERRARE)*.
- (Roy et al., 2000) N. Roy, J. Pineau, and S. Thrun, 2000. Spoken dialogue management using probabilistic reasoning. In Proceedings of *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL)*.
- (Rubino et al., 2012) R. Rubino, S. Huet, F. Avignon, F. Lefèvre, and G. Linares, 2012. Statistical post-editing of machine translation for domain adaptation.
- (Russell, 1998) S. Russell, 1998. Learning agents for uncertain environments (extended abstract). In Proceedings of *the Eleventh Annual Conference on Computational Learning Theory*.
- (Sacks et al., 1974) H. Sacks, E. A. Schegloff, and G. Jefferson, 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735.
- (Sadek, 1991) M. D. Sadek, 1991. *Attitudes Mentales et Interaction Rationnelle : vers une théorie formelle de la Communication*. Phd Thesis, Université de Rennes I, France.
- (Schatzmann et al., 2007) J. Schatzmann, B. Thomson, K. Weilhammer, H. Ye, and S. Young, 2007. Agenda-based user simulation for bootstrapping a pomdp dialogue system. In Proceedings of *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*.
- (Schatzmann et al., 2006) J. Schatzmann, K. Weilhammer, M. Stuttle, and S. Young, 2006. A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. *The Knowledge Engineering Review* 21.

- (Schegloff, 1968) E. A. Schegloff, 1968. Sequencing in conversational openings. *American Anthropologist* 70, pp. 1075–1095.
- (Schlangen and Skantze, 2011) D. Schlangen and G. Skantze, 2011. A general, abstract model of incremental dialogue processing. *Dialogue and Discourse* 2, 83–111.
- (Schmitt et al., 2012) A. Schmitt, S. Ultes, and W. Minker, 2012. A parameterized and annotated spoken dialog corpus of the cmu let’s go bus information system. In Proceedings of *the Eight International Conference on Language Resources and Evaluation (LREC’12)*.
- (Schultz, 1998) W. Schultz, 1998. Predictive reward signal of dopamine neurons. *Neurophysiol.*
- (Schultz et al., 1995) W. Schultz, R. Romo, T. Ljungberg, J. Mirenowicz, J. R. Hollerman, and A. Dickinson, 1995. Reward-related signals carried by dopamine neurons. In Proceedings of *Models of Information Processing in the Basal Ganglia*, 233–248. MIT Press.
- (Searle, 1969) J. Searle, 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, UK.
- (Searle, 1968) J. R. Searle, 1968. Austin on locutionary and illocutionary acts. *The Philosophical Review*.
- (Seigel and Woodland, 2011) M. S. Seigel and P. C. Woodland, 2011. Combining information sources for confidence estimation with crf models. In Proceedings of *INTER-SPEECH*.
- (Selfridge and Heeman, 2010) E. Selfridge and P. A. Heeman, 2010. Importance-driven turn-bidding for spoken dialogue systems. In Proceedings of *ACL*, 177–185.
- (Selfridge et al., 2011) E. O. Selfridge, I. Arizmendi, P. A. Heeman, and J. D. Williams, 2011. Stability and accuracy in incremental speech recognition. In Proceedings of *the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Selfridge et al., 2012) E. O. Selfridge, I. Arizmendi, P. A. Heeman, and J. D. Williams, 2012. Integrating incremental speech recognition and pomdp-based dialogue systems. In Proceedings of *the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Selfridge and Heeman, 2012) E. O. Selfridge and P. A. Heeman, 2012. A temporal simulator for developing turn-taking methods for spoken dialogue systems. In Proceedings of *the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Sidner et al., 2013) C. Sidner, T. Bickmore, C. Rich, B. Barry, L. Ring, M. Behrooz, and M. Shayganfar, 2013. Demonstration of an always-on companion for isolated older adults. In Proceedings of *the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.

- (Singh et al., 2002) S. Singh, D. Litman, M. Kearns, and M. Walker, 2002. Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. *Journal of Artificial Intelligence Research*, 105–133.
- (Singh et al., 1999) S. P. Singh, M. J. Kearns, D. J. Litman, and M. A. Walker, 1999. Reinforcement learning for spoken dialogue systems. In Proceedings of *Advances in Neural Information Processing Systems (NIPS)*.
- (Skantze and Johansson, 2015) G. Skantze and M. Johansson, 2015. Modelling situated human-robot interaction using irstk. In Proceedings of *the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.
- (Skantze and Schlangen, 2009) G. Skantze and D. Schlangen, 2009. Incremental dialogue processing in a micro-domain. In Proceedings of *ACL*.
- (Strombergsson et al., 2013) S. Strombergsson, A. Hjalmarsson, J. Edlund, and D. House, 2013. Timing responses to questions in dialogue. In Proceedings of *INTERSPEECH Proceedings*.
- (Sullivan, 1947) H. Sullivan, 1947. *Conceptions of Modern Psychiatry*. Norton.
- (Sutton and Barto, 1998) R. S. Sutton and A. G. Barto, 1998. *Reinforcement Learning, An Introduction*. The MIT Press, Cambridge, Massachusetts, London, England.
- (Tabet and Boughazi, 2011) Y. Tabet and M. Boughazi, 2011. Speech synthesis techniques. a survey. In Proceedings of *Systems, Signal Processing and their Applications (WOSSPA), 2011 7th International Workshop on*.
- (Tanenhaus et al., 1995) M. K. Tanenhaus, M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy, 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634.
- (Thomson and Young, 2010) B. Thomson and S. J. Young, 2010. Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems. *Computer Speech & Language* 24(4), 562–588.
- (Thorndike, 1898) E. L. Thorndike, 1898. *Animal intelligence: An experimental study of the associative processes in animals*. Columbia University Press.
- (Vogel et al., 2013) A. Vogel, M. Bodoia, C. Potts, and D. Jurafsky, 2013. Emergence of gricean maxims from multi-agent decision theory. In Proceedings of *Conference of the North American Chapter of the Association of Computational Linguistics (NAACL)*.
- (Walker, 2000) M. Walker, 2000. An application of reinforcement learning to dialogue strategy selection in a spoken dialogue system for email. *Journal of Artificial Intelligence Research* 12, 387–416.
- (Walker et al., 1997) M. A. Walker, D. J. Litman, C. A. Kamm, and A. Abella, 1997. Paradise: a framework for evaluating spoken dialogue agents. In Proceedings of *the eighth conference on European chapter of the Association for Computational Linguistics*.

- (Wang et al., 2014) Z. Wang, H. Chen, G. Wang, H. Tian, H. Wu, and H. Wang, 2014. Policy learning for domain selection in an extensible multi-domain spoken dialogue system. In Proceedings of *the 19th Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- (Watkins, 1989) C. J. C. H. Watkins, 1989. *Learning from Delayed Rewards*. Phd Thesis, King's College.
- (Wei and Rudnicky, 1999) X. Wei and A. Rudnicky, 1999. n agenda-based dialog management architecture for spoken language systems. In Proceedings of *IEEE ASRU*.
- (Williams, 2008) J. D. Williams, 2008. The best of both worlds: unifying conventional dialog systems and pomdps. In Proceedings of *INTERSPEECH*, 1173–1176.
- (Williams and Young, 2007) J. D. Williams and S. Young, 2007. Partially observable markov decision processes for spoken dialog systems. *Computer Speech and Language Journal*.
- (Wirén, 1992) M. Wirén, 1992. *Studies in Incremental Natural Language Analysis*. Phd Thesis, Linköping University, Linköping, Sweden.
- (Witt, 2011) S. Witt, 2011. A global experience metric for dialog management in spoken dialog systems. *SemDial*.
- (Wlodarczak and Wagner, 2013) M. Wlodarczak and P. Wagner, 2013. Effects of talk-spurt silence boundary thresholds on distribution of gaps and overlaps. In Proceedings of *INTERSPEECH*.
- (Yngve, 1970) V. H. Yngve, 1970. On getting a word in edgewise. In Proceedings of *CLS-70*.
- (Young et al., 2010) S. Young, M. Gašić, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu, 2010. The hidden information state model: A practical framework for pomdp-based spoken dialogue management. *Computer Speech and Language* 24(2), 150–174.
- (Yu et al., 2011) K. Yu, H. Zen, F. Mairesse, and S. Young, 2011. Context adaptive training with factorized decision trees for hmm-based statistical parametric speech synthesis. *Speech Communication* 53(6).
- (Yuan et al., 2006) J. Yuan, M. Liberman, and C. Cieri, 2006. Towards an integrated understanding of speaking rate in conversation. In Proceedings of *INTERSPEECH Proceedings*.
- (Zhao et al., 2015) T. Zhao, A. W. Black, and M. Eskenazi, 2015. An incremental turn-taking model with active system barge-in for spoken dialog systems. In Proceedings of *the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIG-DIAL)*.



# Glossary

**Automatic speech recognition (ASR):** Module transforming the user's speech into an N-Best.

**Client:** Interface between the user and the system. In the case of spoken dialogue systems, it contains the ASR and the TTS.

**Confidence score:** Positive real number associated with an ASR or a NLU output hypothesis. The higher it is, the more likely this hypothesis is the correct one.

**Dialogue manager (DM):** Module computing the system's response in the form of concepts given a user's request.

**Dialogue system:** System communicating with users through natural interaction using natural language.

**Incremental dialogue system:** Dialogue systems which are able to process the user's speech as it is spoken, before the end of the sentence. They are able to take the floor while the user is still speaking and inversely, the latter can also do so while the system is speaking.

**Micro-turn:** Time interval between two updates of an incremental dialogue system.

**N-Best:** N ASR or NLU hypotheses that are associated with the N best confidence scores.

**Natural language generation (NLG):** Module transforming the concepts computed by the DM into text.

**Natural language understanding (NLU):** Module transforming the user's sentences in text format into concepts understandable by the DM.

**Reinforcement learning:** Machine learning framework where the learning agent learns by trial and error while directly interacting with an environment.

**Scheduler:** Turn-taking management module inserted between the Client and the Service. The set {Scheduler+Service} behaves like an incremental dialogue system from the Client's point of view.

**Service:** Backend part of a dialogue system. It contains the DM, it manages the back-end I/O (database access, etc.) and depending on the architecture, it may also contain the NLU and/or the NLG.

**Spoken dialogue system (SDS):** Dialogue system aimed to communicate through speech. It uses an ASR and a TTS.

**Turn-taking phenomenon (TTP):** Behaviours and mechanisms through which the conversation participants take turns in speaking during a conversation.

**Text-to-speech (TTS):** Module transforming text into a vocal signal. Also called speech synthesis.

**Turn-taking:** The act of taking the floor during a conversation.

**User simulator (US):** System aimed to simulate user behaviours. It makes it possible to quickly generate important dialogue corpora, which is especially interesting when building machine learning algorithms.