

CLASSiC

D4.4.1: Reinforcement Learning of optimal NLG policies using simulated users, for TownInfo and SelfHelp systems

Oliver Lemon, Verena Rieser, Srinivasan Janarthanam, Xingkun Liu

Distribution: Public

CLASSiC

Computational Learning in Adaptive Systems for Spoken Conversation
216594 Deliverable 4.4.1

December 2010



Project funded by the European Community
under the Seventh Framework Programme for
Research and Technological Development



The deliverable identification sheet is to be found on the reverse of this page.

Project ref. no.	216594
Project acronym	CLASSiC
Project full title	Computational Learning in Adaptive Systems for Spoken Conversation
Instrument	STREP
Thematic Priority	Cognitive Systems, Interaction, and Robotics
Start date / duration	01 March 2008 / 36 Months

Security	Public
Contractual date of delivery	M33 = November 2010
Actual date of delivery	December 2010
Deliverable number	4.4.1
Deliverable title	D4.4.1: Reinforcement Learning of optimal NLG policies using simulated users, for TownInfo and SelfHelp systems
Type	Prototype
Status & version	Final 1.0
Number of pages	17 (excluding front matter)
Contributing WP	4
WP/Task responsible	UEDIN
Other contributors	HWU
Author(s)	Oliver Lemon, Verena Rieser, Srinu Janarthanam, Xingkun Liu
EC Project Officer	Philippe Gelin
Keywords	Natural Language Generation, Reinforcement Learning, User Simulation, Information Presentation, Referring Expression Generation

The partners in CLASSiC are:

Heriot-Watt University	HWU
University of Cambridge	UCAM
University of Geneva	GENE
Ecole Supérieure d'Electricité	SUPELEC
France Telecom/ Orange Labs	FT
University of Edinburgh HCRC	EDIN

For copies of reports, updates on project activities and other CLASSiC-related information, contact:

The CLASSiC Project Co-ordinator:

Dr. Oliver Lemon
School of Mathematical and Computer Sciences (MACS)
Heriot-Watt University
Edinburgh
EH14 4AS
United Kingdom
O.Lemon@hw.ac.uk
Phone +44 (131) 451 3782 - Fax +44 (0)131 451 3327

Copies of reports and other material can also be accessed via the project's administration homepage,
<http://www.classic-project.org>

©2010, The Individual Authors.

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Contents

Executive Summary	1
1 RL for Information Presentation policies in the TownInfo domain	2
1.1 Previous work on Information Presentation in SDS	2
2 RL for Referring Expression Generation policies in the Self-Help domain	4
2.1 Related work	5
3 Description of Code associated with this deliverable	6
4 Abstracts of publications associated with this deliverable	7
4.1 “Optimising Information Presentation for Spoken Dialogue Systems” (ACL 2010)	7
4.2 “Learning to Adapt to Unknown Users: Referring Expression Generation in Spoken Dialogue Systems” (ACL 2010)	8
4.3 “Natural Language Generation as Planning Under Uncertainty for Spoken Dialogue Systems” (Book chapter)	9
4.4 “Learning Adaptive Referring Expression Generation Policies for Spoken Dialogue Systems” (Book chapter)	10
4.5 “Natural Language Generation as Planning Under Uncertainty for Spoken Dialogue Systems” (EACL 2009)	11
4.6 “A User Simulation Model for learning Lexical Alignment Policies in Spoken Dialogue Systems” (ENLG 2009)	12
4.7 “User simulations for online adaptation and knowledge- alignment in Troubleshooting dialogue systems” (SEMdial 2008)	13
4.8 “ Adaptive Natural Language Generation in Dialogue using Reinforcement Learning” (SEMdial, 2008)	14

Executive summary

This deliverable presents the prototypes developed in the CLASSiC project which use Reinforcement Learning (RL) methods to develop Natural Language Generation (NLG) policies. The prototypes use simulated users to train NLG policies, for Information Presentation policies in the TownInfo domain, and for Referring Expression Generation in the SelfHelp domain (internet connection setup).

The results of the research in this deliverable have been published in the following papers: [1, 2, 3, 4, 5, 6, 7, 8, 9], all of which are available at www.classic-project.org

The deliverable falls naturally into 2 parts: research on RL for Information Presentation policies in the TownInfo domain, and for Referring Expression Generation in the SelfHelp domain.

Chapter 1

RL for Information Presentation policies in the TownInfo domain

Work on evaluating SDS suggests that the Information Presentation (IP) phase is the primary contributor to dialogue duration [10], and as such, is a central aspect of SDS design. During this phase the system returns a set of items (“hits”) from a database, which match the user’s current search constraints. An inherent problem in this task is the trade-off between presenting “enough” information to the user (for example helping them to feel confident that they have a good overview of the search results) versus keeping the utterances short and understandable.

In the following we show that IP for SDS can be treated as a data-driven joint optimisation problem, and that this outperforms a supervised model of human ‘wizard’ behaviour on a particular IP task (presenting sets of search results to a user).

A similar approach has been applied to the problem of Referring Expression Generation in dialogue [9].

1.1 Previous work on Information Presentation in SDS

Broadly speaking, IP for SDS can be divided into two main steps: 1) IP strategy selection and 2) Content or Attribute Selection. Prior work has presented a variety of **IP strategies** for structuring information (see examples in Table 1.1). For example, the SUMMARY strategy is used to guide the user’s “focus of attention”. It draws the user’s attention to relevant attributes by grouping the current results from the database into clusters, e.g. [11, 12]. Other studies investigate a COMPARE strategy, e.g. [13, 14], while most work in SDS uses a RECOMMEND strategy, e.g. [15]. In a previous proof-of-concept study [16] we show that each of these strategies has its own strengths and drawbacks, dependent on the particular context in which information needs to be presented to a user. Here, we will also explore possible combinations of the strategies, for example SUMMARY followed by RECOMMEND, e.g. [17].

Prior work on **Content or Attribute Selection** has used a “Summarize and Refine” approach [11, 18, 19]. This method employs utility-based attribute selection with respect to how each attribute (e.g. price or food type in restaurant search) of a set of items helps to narrow down the user’s goal to a single item. Related work explores a user modelling approach, where attributes are ranked according to user preferences [12, 20]. Our data collection and training environment incorporate these approaches.

This research is the first to apply a data-driven method to this whole decision space (i.e. combinations of

Strategy	Example utterance
SUMMARY no UM	I found 26 restaurants, which have Indian cuisine. 11 of the restaurants are in the expensive price range. Furthermore, 10 of the restaurants are in the cheap price range and 5 of the restaurants are in the moderate price range.
SUMMARY UM	26 restaurants meet your query. There are 10 restaurants which serve Indian food and are in the cheap price range. There are also 16 others which are more expensive.
COMPARE by Item	The restaurant called Kebab Mahal is an Indian restaurant. It is in the cheap price range. And the restaurant called Saffrani, which is also an Indian restaurant, is in the moderate price range.
COMPARE by Attribute	The restaurant called Kebab Mahal and the restaurant called Saffrani are both Indian restaurants. However, Kebab Mahal is in the cheap price range while Saffrani is moderately priced.
RECOMMEND	The restaurant called Kebab Mahal has the best overall quality amongst the matching restaurants. It is an Indian restaurant, and it is in the cheap price range.

Table 1.1: Example realisations, generated when the user provided `cuisine=Indian`, and where the wizard has also selected the additional attribute `price` for presentation to the user.

Information Presentation strategies as well as attribute selection), and to show the utility of both lower-level features (e.g. from the NLG realiser) and higher-level features (e.g. from Dialogue Management) for this problem. Previous work has only focused on individual aspects of the problem (e.g. how many attributes to generate, or when to use a SUMMARY), using a pipeline model for SDS with DM features as input, and where NLG has no knowledge of lower level features (e.g. behaviour of the realiser). We also show that lower level features significantly influence users' ratings of IP strategies. In the research carried out within the CLASSiC project we use a Reinforcement Learning (RL) as a statistical planning framework [21] to explore the contextual features for making these decisions, and propose a new joint optimisation method for IP strategies combining content structuring and attribute selection.

Chapter 2

RL for Referring Expression Generation policies in the Self-Help domain

We present a reinforcement learning framework to learn user-adaptive referring expression generation (REG) policies from a data-driven user simulation. A user-adaptive REG policy allows the system to choose appropriate expressions to refer to domain entities in a dialogue setting. For instance, in a technical support conversation, the system could choose to use more technical terms with an expert user, or to use more descriptive and general expressions with novice users, and a mix of the two with intermediate users of various sorts (see examples in Table 2.1).

In natural human-human conversations, dialogue partners learn about the other person and adapt their language to suit their domain expertise [22]. This kind of adaptation is called *Alignment through Audience Design* [23, 24, 25]. Some current Spoken Dialogue Systems (SDS) incorporate user-adaptive behaviour. However, most such systems assume that the user’s domain knowledge is accurately known beforehand [26].

In contrast, we assume that the users interacting with SDS are mostly unknown to the system and therefore the SDS must be capable of observing the user’s dialogue behaviour, modelling his/her domain knowledge, and adapting accordingly, just like human interlocutors. As a dialogue progresses, the SDS must estimate the likely lexical knowledge of the user. We present a corpus-driven framework using which a user-adaptive REG policy can be learned, by employing reinforcement learning [27]. We show that these learned policies perform better than hand-coded policies by almost 9% in terms of accuracy of adaptation. We also compared the performance of policies learned using hand-coded and data-driven simulations and show that data-driven simulations produce better policies than hand-coded ones.

Jargon: Please plug one end of the broadband cable into the broadband filter.
Descriptive: Please plug one end of the thin white cable with grey ends into the small white box.

Table 2.1: Referring Expression examples for 2 entities, from the corpus

2.1 Related work

There are several ways in which NLG systems adapt to users. Some of them adapt to a user's goals, preferences, environment and so on. Our focus in this study is restricted to the user's lexical domain expertise. Several NLG systems adapt to the user's domain expertise at different levels of generation - text planning [28, 29], complexity of instructions [30], referring expressions [31], and so on. Some dialogue systems, such as COMIC, have also incorporated NLG modules that present appropriate levels of instruction to the user [26]. However, in all the above systems, the user's knowledge is assumed to be accurately represented in an initial user model using which the system adapts its language. In contrast to all these systems, our adaptive REG policy knows nothing about the user when the conversation starts. It gradually builds the user model and exploits it by re-estimating the users' expertise during the conversation.

Reinforcement Learning (RL) has been successfully used for learning dialogue management policies since [32]. The learned policies allow the dialogue manager to optimally choose appropriate dialogue acts such as instructions, confirmation requests, and so on, under uncertain noise or other environment conditions. Rieser and Lemon [33, 34] recently presented a model to learn information presentation strategies using reinforcement learning. In contrast, we present a framework that learns to choose appropriate referring expressions based on a user's domain knowledge. Similar work has been reported by [35]. However, this approach was not corpus-driven and only used a hand-coded user simulation.

Chapter 3

Description of Code associated with this deliverable

The learned policies were developed using the REALL toolkit [36, 37] as described in detail in [8, 9].

Once trained, the policies can be loaded into the REALL toolkit (java), and are then interfaced with our spoken dialogue systems using an interface to REALL. In the case of integration within CLASSiC System 1, a JNI interface to REALL was developed.

Within System 1, the NLG component reads in the dialogue act output by the DM component and generates natural language responses in the form of text plus Baratinoo TTS tags (e.g. for speech pitch and rate) using Edinburgh's NLG module. The EdinNLG module is a semantic output component which is switchable with other different semantic output components via the system config file. The module is comprised of a C++ interface component and a core NLG realiser which is a package of Java programs. The C++ component interacts with the core realiser via JNI (Java Native Interface).

Once the realiser gets the inputs from DM via JNI, it sends a request which contains the user filled slots, unfilled slots and the number of DB hits to the REALL server. REALL then outputs the NLG actions based on the trained policy and the input features. The realiser then generates the sentences based on the REALL policy and the DM inputs, and then sends generated sentences to the DM via the JNI interface.

In the final CLASSiC TownInfo System 1, the EdinNLG module is called whenever the dialogue act is an "Offer" or "FindAlt" act. It then uses a learned NLG policy, which has been trained using the simulated users described in D3.3, which were built using the Wizard-of-Oz data collected in D6.1.1.

An example input DA is:

```
inform(name="Cafe Adriatic", type=restaurant, near="Williams Art and Antiques", food="Takeaway pizza")
```

The fully trained NLG module from WP4 can choose combinations of Summary, Compare, and Recommend actions. In order to ensure compatibility with the output actions chosen by the DM component, the learned NLG component is restricted in System 1, so that its chosen NLG action always ends with the recommendation required by the DM.

Chapter 4

Abstracts of publications associated with this deliverable

The following publications [1, 2, 3, 4, 6, 7, 8, 9] are associated with the work presented in this deliverable and are available at www.classic-project.org

4.1 “Optimising Information Presentation for Spoken Dialogue Systems” (ACL 2010)

Authors: Verena Rieser and Oliver Lemon and Xingkun Liu

Publication venue: ACL 2010 (Annual Conference of the Association for Computational Linguistics, 2010)

Abstract:

We present a novel approach to Information Presentation (IP) in Spoken Dialogue Systems (SDS) using a data-driven statistical optimisation framework for content planning and attribute selection. First we collect data in a Wizard-of-Oz (WoZ) experiment and use it to build a supervised model of human behaviour. This forms a baseline for measuring the performance of optimised policies, developed from this data using Reinforcement Learning (RL) methods. We show that the optimised policies significantly outperform the baselines in a variety of generation scenarios: while the supervised model is able to attain up to 87.6% of the possible reward on this task, the RL policies are significantly better in 5 out of 6 scenarios, gaining up to 91.5% of the total possible reward. The RL policies perform especially well in more complex scenarios. We are also the first to show that adding predictive “lower level” features (e.g. from the NLG realiser) is important for optimising IP strategies according to user preferences. This provides new insights into the nature of the IP problem for SDS.

4.2 “Learning to Adapt to Unknown Users: Referring Expression Generation in Spoken Dialogue Systems” (ACL 2010)

Authors: Srinivasan Janarthanam and Oliver Lemon

Publication venue: ACL 2010 (Annual Conference of the Association for Computational Linguistics, 2010)

Abstract:

We present a data-driven approach to learn user-adaptive referring expression generation (REG) policies for spoken dialogue systems. Referring expressions can be difficult to understand in technical domains where users may not know the technical ‘jargon’ names of the domain entities. In such cases, dialogue systems must be able to model the user’s (lexical) domain knowledge and use appropriate referring expressions. We present a reinforcement learning framework in which the system learns REG policies which can adapt to unknown users online. Furthermore, unlike supervised learning methods which require a large corpus of expert adaptive behaviour to train on, we show that effective adaptive policies can be learned from a small dialogue corpus of non-adaptive human-machine interaction, by using a RL framework and a statistical user simulation. We show that in comparison to adaptive hand-coded baseline policies, the learned policy performs significantly better, with an 18.6% average increase in adaptation accuracy. The best learned policy also takes less dialogue time (average 1.07 min less) than the best hand-coded policy. This is because the learned policies can adapt online to changing evidence about the user’s domain expertise.

4.3 “Natural Language Generation as Planning Under Uncertainty for Spoken Dialogue Systems” (Book chapter)

Authors: Verena Rieser and Oliver Lemon

Publication venue: Empirical Methods in Natural Language Generation 2010 (Book chapter)

Abstract:

We present and evaluate a new model for Natural Language Generation (NLG) in Spoken Dialogue Systems, based on statistical planning, given noisy feedback from the current generation context (e.g. a user and a surface realiser). The model is adaptive and incremental at the turn level, and optimises NLG actions with respect to a data-driven objective function. We study its use in a standard NLG problem: how to present information (in this case a set of search results) to users, given the complex trade-offs between utterance length, amount of information conveyed, and cognitive load. We set these trade-offs in an objective function by analysing existing MATCH data. We then train a NLG policy using Reinforcement Learning (RL), which adapts its behaviour to noisy feedback from the current generation context. This policy is compared to several baselines derived from previous work in this area. The learned policy significantly outperforms all the prior approaches.

4.4 “Learning Adaptive Referring Expression Generation Policies for Spoken Dialogue Systems” (Book chapter)

Authors: Srinivasan Janarthanam and Oliver Lemon

Publication venue: Empirical Methods in Natural Language Generation 2010 (Book chapter)

Abstract:

We address the problem that different users have different lexical knowledge about problem domains, so that automated dialogue systems need to adapt their generation choices online to the users' domain knowledge as it encounters them. We approach this problem using Reinforcement Learning in Markov Decision Processes (MDP). We present a reinforcement learning framework to learn adaptive referring expression generation (REG) policies that can adapt dynamically to users with different domain knowledge levels. In contrast to related work we also propose a new statistical user model which incorporates the lexical knowledge of different users. We evaluate this framework by showing that it allows us to learn dialogue policies that automatically adapt their choice of referring expressions online to different users, and that these policies are significantly better than hand-coded adaptive policies for this problem. The learned policies are consistently between 2 and 8 turns shorter than a range of different hand-coded but adaptive baseline REG policies.

4.5 “Natural Language Generation as Planning Under Uncertainty for Spoken Dialogue Systems” (EACL 2009)

Authors: Verena Rieser and Oliver Lemon

Publication venue: EACL 2009 (European Conference of the Association for Computational Linguistics)

Abstract:

We present and evaluate a new model for Natural Language Generation (NLG) in Spoken Dialogue Systems, based on statistical planning, given noisy feedback from the current generation context (e.g. a user and a surface realiser). We study its use in a standard NLG problem: how to present information (in this case a set of search results) to users, given the complex trade-offs between utterance length, amount of information conveyed, and cognitive load. We set these trade-offs by analysing existing MATCH data. We then train a NLG policy using Reinforcement Learning (RL), which adapts its behaviour to noisy feedback from the current generation context. This policy is compared to several base-lines derived from previous work in this area. The learned policy significantly outperforms all the prior approaches.

4.6 “A User Simulation Model for learning Lexical Alignment Policies in Spoken Dialogue Systems” (ENLG 2009)

Authors: Srimi Janarthanam and Oliver Lemon

Publication venue: European workshop on Natural Language Generation 2009

Abstract:

We study the problem of lexical alignment between dialogue participants, using the practical example of troubleshooting dialogue systems. We address the problem that different users have different lexical knowledge about problem domains, so that automated dialogue systems need to adapt online to the different lexical choices of these users as it encounters them. We approach this problem using policy learning in a Markov Decision Process (MDP). In contrast to related work we propose a new statistical user model which incorporates the lexical knowledge of different users. We evaluate this user model by showing that it allows us to learn dialogue policies that automatically adapt their lexical choice online to new users, and that these policies are significantly better than threshold-based adaptive hand-coded policies for this problem. The learned policies are consistently between 2 and 8 turns shorter than a range of different hand-coded baseline lexical alignment policies.

4.7 “User simulations for online adaptation and knowledge-alignment in Troubleshooting dialogue systems” (SEMDial 2008)

Authors: Srinivasan and Oliver Lemon

Publication venue: Proceedings of SEMDial 2008

Abstract:

We study the problem of alignment between dialogue participants, using the practical example of troubleshooting dialogue systems. Recent work on troubleshooting concerns automated spoken dialogue systems which support users who need to repair their internet connection. We address the problem that different users have different types of knowledge of problem domains, so that automated dialogue systems need to adapt online to the different knowledge of these users as it encounters them. We approach this problem using policy learning in a Markov Decision Process (MDP). In contrast to related work we propose a new user model which incorporates the different conceptual knowledge of different users, together with an environment simulation. We show that this model allows us to learn dialogue policies that automatically adapt online to new users, and that these policies are significantly better than threshold-based adaptive hand-coded policies for this problem.

4.8 “Adaptive Natural Language Generation in Dialogue using Reinforcement Learning” (SEMdial, 2008)

Authors: Oliver Lemon

Publication venue: Proceedings of SEMdial 2008

Abstract:

This paper presents a new model for adaptive Natural Language Generation (NLG) in dialogue, showing how NLG problems can be approached as statistical planning problems using Reinforcement Learning. This approach brings a number of theoretical and practical benefits such as fine-grained adaptation, generalization, and automatic (global) optimization. We present the model and related work in statistical/trainable NLG, discuss its applications, and provide a demonstration of the approach, showing policy learning for adaptive information presentation decisions (Contrast, Cluster, or List items). An adaptive NLG policy learned in our framework shows a statistically significant 27% relative increase in reward over an RL-majority baseline policy for the same task. We thereby also show that such NLG problems should be approached in combination with dialogue management decisions, and we show how to jointly optimize NLG and dialogue management plans.

Bibliography

- [1] Oliver Lemon. Adaptive Natural Language Generation in Dialogue using Reinforcement Learning. In *Proceedings of SEMdial*, 2008.
- [2] Srinivasan Janarthanam and Oliver Lemon. User simulations for online adaptation and knowledge-alignment in Troubleshooting dialogue systems. In *Proceedings of SEMdial*, pages 149–156, 2008.
- [3] Verena Rieser and Oliver Lemon. Natural language generation as planning under uncertainty for spoken dialogue systems. In *EACL*, 2009.
- [4] Srinivasan Janarthanam and Oliver Lemon. A User Simulation Model for learning Lexical Alignment Policies in Spoken Dialogue Systems. In *European Workshop on Natural Language Generation*, 2009.
- [5] Srinivasan Janarthanam and Oliver Lemon. A Wizard-of-Oz environment to study Referring Expression Generation in a Situated Spoken Dialogue Task. In *European Workshop on Natural Language Generation*, 2009.
- [6] Verena Rieser and Oliver Lemon. Natural language generation as planning under uncertainty for spoken dialogue systems. In E. Krahmer and M. Theune, editors, *Empirical Methods in Natural Language Generation*, volume 5980 of *Lecture Notes in Computer Science*, pages 105–120. Springer, Berlin / Heidelberg, 2010.
- [7] Srinivasan Janarthanam and Oliver Lemon. Learning adaptive referring expression generation policies for spoken dialogue systems. In E. Krahmer and M. Theune, editors, *Empirical Methods in Natural Language Generation*, volume 5980 of *Lecture Notes in Computer Science*, pages 67–84. Springer, Berlin / Heidelberg, 2010.
- [8] Verena Rieser, Oliver Lemon, and Xingkun Liu. Optimising information presentation for spoken dialogue systems. In *Proceedings of ACL*, 2010.
- [9] Srinivasan Janarthanam and Oliver Lemon. Learning to adapt to unknown users: Referring expression generation in spoken dialogue systems. In *Proceedings of ACL*, 2010.
- [10] M. Walker, R. Passonneau, and J. Boland. Quantitative and qualitative evaluation of DARPA Communicator spoken dialogue systems. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2001.
- [11] Joseph Polifroni and Marilyn Walker. Intensional Summaries as Cooperative Responses in Dialogue Automation and Evaluation. In *Proceedings of ACL*, 2008.

- [12] Vera Demberg and Johanna D. Moore. Information presentation in spoken dialogue systems. In *Proceedings of EACL*, 2006.
- [13] Marilyn Walker, Amanda Stent, François Mairesse, and Rashmi Prasad. Individual and domain adaptation in sentence planning for dialogue. *Journal of Artificial Intelligence Research (JAIR)*, 30:413–456, 2007.
- [14] Crystal Nakatsu. Learning contrastive connectives in sentence realization ranking. In *Proc. of SIG-dial Workshop on Discourse and Dialogue*, 2008.
- [15] SJ Young, J Schatzmann, K Weilhammer, and H Ye. The Hidden Information State Approach to Dialog Management. In *ICASSP 2007*, 2007.
- [16] Verena Rieser and Oliver Lemon. Natural Language Generation as Planning Under Uncertainty for Spoken Dialogue Systems. In *Proc. of EACL*, 2009.
- [17] Steve Whittaker, Marilyn Walker, and Johanna Moore. Fish or Fowl: A Wizard of Oz evaluation of dialogue strategies in the restaurant domain. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*, 2002.
- [18] Joseph Polifroni and Marilyn Walker. Learning database content for spoken dialogue system design. In *Proc. of the IEEE/ACL workshop on Spoken Language Technology (SLT)*, 2006.
- [19] Grace Chung. Developing a flexible spoken dialog system using simulation. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2004.
- [20] Andi Winterboer, Jiang Hu, Johanna D. Moore, and Clifford Nass. The influence of user tailoring and cognitive load on user performance in spoken dialogue systems. In *Proc. of the 10th International Conference of Spoken Language Processing (Interspeech/ICSLP)*, 2007.
- [21] R. Sutton and A. Barto. *Reinforcement Learning*. MIT Press, 1998.
- [22] E. A. Issacs and H. H. Clark. References in conversations between experts and novices. *Journal of Experimental Psychology: General*, 116:26–37, 1987.
- [23] H. H. Clark and G. L. Murphy. Audience design in meaning and reference. In J. F. LeNy and W. Kintsch, editors, *Language and comprehension*. Amsterdam: North-Holland, 1982.
- [24] A. Bell. Language style as audience design. *Language in Society*, 13(2):145–204, 1984.
- [25] H. H. Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.
- [26] K. McKeown, J. Robin, and M. Tanenblatt. Tailoring Lexical Choice to the User’s Vocabulary in Multimedia Explanation Generation. In *Proc. ACL 1993*, 1993.
- [27] R. Sutton and A. Barto. *Reinforcement Learning*. MIT Press, 1998.
- [28] K. R. McKeown. *Text Generation*. Cambridge University Press, 1985.
- [29] C. L. Paris. *The Use of Explicit User Models in Text Generations: Tailoring to a User’s Level of Expertise*. Ph.D. thesis, Columbia University, 1987.

- [30] R. Dale. Cooking up referring expressions. In *Proc. ACL-1989*, 1989.
- [31] E. Reiter. Generating Descriptions that Exploit a User's Domain Knowledge. In R. Dale, C. Mellish, and M. Zock, editors, *Current Research in Natural Language Generation*, pages 257–285. Academic Press, 1991.
- [32] E. Levin, R. Pieraccini, and W. Eckert. Learning Dialogue Strategies within the Markov Decision Process Framework. In *Proc. of ASRU97*, 1997.
- [33] V. Rieser and O. Lemon. Natural Language Generation as Planning Under Uncertainty for Spoken Dialogue Systems. In *Proc. EACL'09*, 2009.
- [34] Verena Rieser and Oliver Lemon. Natural Language Generation as Planning Under Uncertainty for Spoken Dialogue Systems. In *Empirical Methods in Natural Language Generation*. 2010.
- [35] S. Janarthnam and O. Lemon. Learning Lexical Alignment Policies for Generating Referring Expressions for Spoken Dialogue Systems. In *Proc. ENLG'09*, 2009.
- [36] Shapiro, D. and Langley, P. Using background knowledge to speed reinforcement learning. In *Proceedings of the Fifth International Conference on Autonomous Agents*, 2001.
- [37] D. Shapiro and P. Langley. Separating skills from preference: Using learning to program by reward. In *Proc. ICML-02*, 2002.