Conversation Analysis Structured Dialogue for Multi-Domain Dialogue Management

Nathan Duran
University of the West of England
Bristol, UK
nathan.duran@uwe.ac.uk

Steve Battle
University of the West of England
Bristol, UK
steve.battle@uwe.ac.uk

ABSTRACT

Dialogue state tracking is a vital component of task-oriented dialogue systems. Often, dialogue states are constrained by their target domains entity's, slots and values. Adding new domains and knowledge may require laborious hand-crafting or re-training using new corpora and training data. This makes the development of multi-domain dialogue systems a considerable challenge.

To address this problem, we propose a method of structuring dialogue that is independent of domain, and closely related to constructs defined by the sociological research of Conversation Analysis, a study of human interaction in conversations. First, we summarise the applicable theories of Adjacency Pairs and Dialogue Acts and their relevance to dialogue systems. We then introduce a schema for structuring dialogue and an accompanying corpus, that utilise the Conversation Analysis inspired constructs, and discuss their potential advantages in moving towards domain agnostic dialogue management.

CCS CONCEPTS

• **Human-centered computing** → *HCI theory, concepts and models*;

KEYWORDS

Conversation Analysis, Adjacency Pairs, Dialogue Act, Dialogue Management, State Tracking

ACM Reference Format:
Nathan Duran and Steve Battle. 2018. Conversation Analysis Structured Dialogue for Multi-Domain Dialogue Management. In *Proceedings of Workshop at the sixth international conference on Human-agent Interaction (HAI) (DEXAHAI'18).* ACM, New York, NY, USA, 4 pages.

1 INTRODUCTION

A key component of dialogue systems is the ability to determine a user's goal or intent, based on the current dialogue state, in order to select an appropriate system response or action. Current approaches to *dialogue state tracking* (DST) are often constrained to specific domains as a result of the dialogue corpora they are trained with, and the target domains dictate an *ontology* of relevant *slots* and *values*, or in the case of Reinforcement Learning (RL) state-action space, which inform the set of possible dialogue states [17].

Some recent approaches to DST include end-to-end neural dialogue systems, which learn a mapping from dialogue history to system response in a supervised fashion [18] [25]. However, while these systems are promising, and reduce the need for a more traditional modular approach, they struggle to generalise to previously unseen out-of-domain data [28] and tend to generate generic or repetitive responses [19]. RL based approaches learn a *policy model* which selects system actions based on the current dialogue state and have proven to be effective at managing long-term goals and handling uncertainty in user utterances. However, even within well-defined static domains, the state-action space for RL systems can grow infeasibly large, making them difficult to adapt to new or changing domains [33]. More recently, Deep Reinforcement Learning (DRL) has been gaining increased attention as a method to reduce the need for manual hand crafting of large state-action spaces and allow the agent to simultaneously learn relevant state-action feature representations and dialogue policy [9]. While DRL has the favourable qualities of combining the advantages of neural end-to-end systems with those of RL [21] [34], adapting these systems to new domains, or creating truly multi-domain dialogue systems, remains an open and interesting research problem [10].

To address this problem, we propose a method of structuring dialogue, for dialogue management systems, that is inspired by the study of human conversation. Specifically, Conversation Analysis (CA) which is an area of sociological research that aims to define, and analyse, constructs that facilitate turn-taking in human conversations [29]. Much of the theory defined by CA can be applied to any conversation and therefore offers some insights into methods of capturing the structure of dialogue that may be used for domain agnostic dialogue systems. Section 2.1, outlines the applicable concepts from CA and their relevance to dialogue systems. Section 2.2 describes our approach to defining a CA structure for dialogue, including the creation of a multi-domain corpus for developing dialogue systems that use CA as a basis for dialogue management. In Section 3 we discuss previous work related to capturing or analysing dialogue structure using CA (or similar) concepts and highlight their use in dialogue systems. Finally, we conclude and discuss future work in section 4.

2 CONVERSATION ANALYSIS FOR DIALOGUE SYSTEMS

2.1 Adjacency Pairs and Dialogue Acts

A key aspect of CA, is the concept of Adjacency Pairs (AP). AP are the base units of sequence-construction in talk, and in their basic unexpanded form comprise of two turns by different speakers that take place one after the other. The initial turn is called the *first pair*

A:	Do you know the directions?	***FPPbase***
B:	Are you driving or walking?	***FPPinsert***
A:	Walking.	***SPPinsert***
B:	Get on the subway . . .	***SPPbase***

Figure 1: Example of AP labelled dialogue.

part (FPP) and initiates an exchange, the second turn is a *second pair part* (SPP) which are responsive to the prior FPP. To account for more complex dialogue structures, AP also include the concept of expansion which allows the construction of sequences of talk that are made up of more than one AP, while still contributing to the same basic action [22], i.e. a question (FPP-base) could be followed by a question (FPP-insert), to elicit information required to better answer the initial question (see fig 1).

AP are also 'type related', for example, a question and an answer [30]. This pair-type relation has the useful property of limiting the range of possible SPP responses to a given FPP, i.e. a question should be followed by an answer (positive or negative), but not a greeting [22]. For dialogue systems, this has the advantageous effect of reducing the set of all possible responses to just a few valid types, while also ensuring the response makes sense in conversational terms.

Though not strictly part of CA, Dialogue Acts (DA) are closely related to AP-types and are a method of labelling the *semantic content* and *communicative function* of an utterance, such as, setQuestion or request; facilitating the computational modelling of communicative behaviour in dialogue [7]. For dialogue systems, DA play a key role in the challenge of interpreting the meaning of user utterances that is central to dialogue management and is not dependant on topic or domain [5].

AP then, can act as formal set of 'rules' which describe how an interaction can be structured and that apply to all types of conversation, irrespective of the current topic or objectives of the interaction. The inclusion of DA as semantic type-labels for adjacency-pair-parts support this structure, by enabling the identification and selection of subsequent pair-parts (i.e. a question should be followed by an answer, etc) and capturing an exchange into a single structure [4].

2.2 Dialogue Structure with Conversation Analysis

The focus of the proposed approach is to develop a definition of dialogue structure that facilitates the creation of dialogue systems that use AP with DA-types to interpret dialogue, and track dialogue state, in a domain agnostic fashion. The assumption, is that typed-AP represent a particular dialogue state, which in turn, informs the selection of appropriate response types that are constrained by the previous dialogue turns. This is analogous to the 'slot filling' approach commonly employed in dialogue systems, with the exception that the dialogue manager does not need to consider the specific *entities*, *slots* or *values* of the target domain. Instead, the dialogue manager must only consider higher level information, such as whether short or long-term goals have been met, whether more information is required from the user or if there is a need to inform the user with updated information. In essence, it is a separation of the domain specific knowledge from the task of dialogue management.

2.2.1 Schema. The first step, is to determine how typed-AP can be used to describe the structure of any dialogue, so that a system can be developed to interpret and manage dialogue using this structure. To this end, we have developed an AP and DA labelling schema that defines 11 AP and 35 DA. The set of AP include FPP and SPP for *base*, *pre*, *post* and *insert-expansions* as described in [22]. Because dialogue does not always contain even numbers of utterances, there are also single-labels (*pre*, *post* and *insert*) for utterances that do not belong to conventional AP. These are closely related to the idea of *minimal-expansions* [30], in that they are not designed to project any further sequences of talk, but rather open, close or add to sequences respectively.

The set of DA are derived from the Dialogue Act Markup Language (DiAML) as defined in ISO 24617 [6]. DiAML was developed as an empirically and theoretically well founded, application independent, DA annotation scheme and is also intended to be used by both human annotators and automatic annotation methods. The two components of each DA are the *semantic content* and *communicative function*. The semantic content specifies objects, propositions and events that the dialogue act is about; the communicative function specifies of the way an addressee should use the semantic content to update the information state [7].

While many DA labelling schema exist [31] [20], they are often subtly different, and DA labels used for dialogue systems frequently contain approach specific elements [23] [2], which causes difficulties with compatibility and re-use [8]. DiAML is therefore ideal for the purposes of our schema, by extending (and maintaining compatibility with) an existing ISO standard, and supporting a move towards the standardisation of DA annotation schemes used for dialogue systems. The DA and AP labels defined by the schema combine to form the typed-AP previously described, and together, they can be used to specify both the structure of a dialogue and the semantic meaning of its constituent utterances that is independent of the dialogues domain.

2.2.2 Conversation Analysis Corpus. In conjunction with developing the CA labelling schema we developed a CA labelled corpus. Primarily, to ensure that the labelling schema is applicable to a large number of dialogues across several domains and facilitate the development of automated methods of DA and AP recognition within unlabelled dialogue data. Secondarily, it will provide training and test data for a dialogue system developed to use the CA constructs.

While DA are commonly used in dialogue systems [15] [14] [10], and there are numerous corpora annotated with DA [20] [32] the use of AP is much less common. Currently, the ICSI Meeting Recorder Dialog Act (MRDA) Corpus [31] is the only AP and DA annotated dialogue corpus. However, the MRDA is only annotated with the *'base'* type of AP and many utterances between a *base* FPP and SPP (i.e. *expansions*) are not labelled, making it time consuming to adapt to our schema. Instead we chose to label the recently created Key-Value Retrieval Networks (KVRET) corpus [12]. KVRET is a multi-domain, task-oriented corpus set in the in-car personal assistant space. The corpus contains ~3000 dialogues that are split between the tasks of calendar scheduling, Navigation and Weather

USER: Is it going to rain at all in the upcoming week?
FPPbase - propositionalQuestion

SYS: What city shall I check for rain?
FPPinsert - setQuestion

USER: Redwood city.
SPPinsert - answer

SYS: Chance of rain on Saturday in Redwood City.
SPPbase - confirm

USER: Ok thank you.
FPPbase - thanking

SYS: No problem.
SPPbase - acceptThanking

Figure 2: Example of KVRET dialogue with CA Schema.

enquiries, each with distinct slot types and values. This is a particularly advantageous feature for developing CA labelling schema and multi-domain dialogue system. Firstly, it ensures the CA structures described by the schema are applicable to several task-oriented dialogue domains. Secondly, it enables a domain agnostic dialogue system to be developed using, for example, two dialogue domains for training and testing with the previously unseen third, with its own *vocabulary*, *entities*, *slots* and *values*. The intention here, is that the dialogue manager remains unchanged and the knowledge from the previously unseen domain is incorporated into the existing ontology. An example of a dialogue annotated using the schema can be found in figure 2. The user initiates the dialogue by asking a propositional question about the weather in the coming week. The system requires more information (i.e. it is missing a *value* for a required *slot*) and so initiates an *insert-expansion* AP. The users answer then closes the *insert-expansion* and provides the system with the necessary information to close the *base* AP.

3 RELEVANT WORK

Here we briefly discuss previous work that use DA or AP to capture or predict the structure of dialogue. We also highlight a number of approaches to dialogue systems that use DA to inform the selection of subsequent dialogue turns.

3.0.1 Dialogue Structure. In the work of Alexandersson and Reithinger [1], DA are used as a basis to model the structure of dialogue using a tree-like 'intentional structure'. The intentional structure, which is an abstraction of different levels of dialogue (*dialogue*, *phase* and *turn*), is then used to predict turn classes. Similarly, DA are used in [3] to automatically create 'parse-tree-like' task/subtask structures in task-oriented dialogues. They show that the 'incremental evolution of dialogue structure' can, at least in part, be predicted based on the sequence of its DA. In their work on segmenting dialogue for verification of AP, Midgley, Harrison and Macnish [24], concluded that AP yield information about 'what is likely to happen', not just for the next utterance but also later in the dialogue. They performed a chi-squared test on their dialogue segments and found that different typed-AP do indeed occur as expected, giving 'empirical justification for Sacks and Schegloff's AP'. With this data they were also able to produce a directed acyclic graph which illustrates the common relationships between different types of AP. These works suggest that AP and DA do indeed carry information not just about the structure of dialogues, but also have some predictive qualities for what is likely to come next at a particular state of the conversation.

3.0.2 Dialogue Management. The utility of categorising dialogue utterance into fixed sets, to aid selection of subsequent utterances and simplify the decision space, is explored in the work of Gunasekara et al., [16]. They use clustering on vector representations of dialogue utterances, where each cluster represents semantically similar utterances that are analogous to DA. Once the utterances were clustered they found that a conversation can be represented as a sequence of clusters and resulted in more accurate conversations regarding the particular goal of interest. In [14] DA are divided into DA-type and slot-parameter, where the former is domain independent and the latter is domain specific. A RNN was then used to estimate the most likely next user and system DA-types to inform a RL based dialogue managers decision process. Similarly, Frampton and Lemon [13], use DA as 'high level' contextual information in the information state update for a RL dialogue system. Cuayáhuitl et al., [10], use individual DA as system actions and together represented the action-space for their DRL dialogue system. These approaches indicate the advantages of DA in reducing the size of the state-action space for RL based dialogue systems, while also carrying some semantic meaning to, inform the dialogue manager in identifying the current dialogue state and select appropriate actions.

4 CONCLUSIONS

In this paper we have proposed a method of structuring dialogue, to aid the development of multi-domain dialogue systems, that is inspired by the sociological research of CA. We presented a dialogue labelling schema that incorporates the concept of AP [29] and DA based on ISO 24617 [6]. This schema has the advantage of being entirely domain agnostic, and is therefore theoretically able to describe both the structure and meaning of any dialogue. Additionally, by remaining consistent with an established ISO standard, the schema supports a move to standardise the labelling of dialogue for computational modelling. We also introduced a multi-domain dialogue corpus, labelled with the schema, and based on the KVRET corpus [12]. This corpus begins to establish the schemas applicability to multiple dialogue domains, and in future, will aid in the development of automated approaches for detecting CA structures in dialogue and the development of multi-domain dialogue systems that use it.

We have two avenues of future work for this concept. Firstly, the exploration of automated approaches for the identification or labelling of dialogues using the CA schema. This would enable the CA dialogue structure to be applied to a wider range of dialogue corpora without the need for time consuming manual labelling. The approach of capturing the relationship between dialogue utterances as latent variables, as in [19] [27], shows promise in also being applied to AP. So too, does the unsupervised pre-training of language models, followed by supervised fine-tuning, for next sentence prediction [26] [11]. Secondly, development of a multi-domain dialogue system that uses the constructs defined by our schema to manage dialogue in a domain agnostic fashion. Key to this work, will be determining methods of adding new domains and knowledge with minimal, or no, modification of the dialogue management system.

REFERENCES

[1] Jan Alexandersson and Norbert Reithinger. 1997. Learning Dialogue Structures from a Corpus. *Sci. Technol.* (1997), 3–6.

[2] Layla El Asri, Hannes Schulz, Shikhar Sharma, Jeremie Zumer, Justin Harris, Emery Fine, Rahul Mehrotra, and Kaheer Suleman. 2017. Frames: A Corpus for Adding Memory to Goal-Oriented Dialogue Systems. In *Proc. SIGDIAL 2017 Conf.* Association for Computational Linguistics, 207–219.

[3] Srinivas Bangalore, Giuseppe Di Fabbrizio, and Amanda Stent. 2008. Learning the Structure of Task-Driven HumanâĂŞHuman Dialogs. *IEEE Trans. Audio. Speech. Lang. Processing* 16, 7 (2008), 1249–1259. https://doi.org/10.1109/TASL.2008.2001102

[4] Kristy Elizabeth Boyer, Eun Young Ha, Robert Phillips, Michael D Wallis, Mladen A Vouk, and James Lester. 2009. Inferring Tutorial Dialogue Structure with Hidden Markov Modeling. In *Proc. Fourth Work. Innov. Use NLP Build. Educ. Appl. - EdAppsNLP '09*. 19–26. https://doi.org/10.3115/1609843.1609846

[5] Kristy Elizabeth Boyer, Eun Young Ha, Robert Phillips, Michael D Wallis, Mladen A Vouk, and James Lester. 2010. Dialogue Act Modeling in a Complex Task-Oriented Domain. In *Proc. SIGDIAL 2010 11th Annu. Meet. Spec. Interes. Gr. Discourse Dialogue.* 297–305.

[6] British Standards Institution. 2012. ISO 24617-2: Language Resource Management - Semantic Annotation Framework (SemAF) Part 2: Dialogue acts.

[7] Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-belis, and David Traum. 2012. ISO 24617-2 : A Semantically-based Standard for Dialogue Annotation. In *Proc. Lr. 2012.* 430–437.

[8] Harry Bunt, Alex C Fang, and Xiaoyue Liu. 2013. Issues in the Addition of ISO Standard Annotations to the Switchboard Corpus. In *Proc. 9th Jt. ISO ACL SIGSEM Work. Interoper. Semant. Annot.* Association for Computational Linguistics, 67–78.

[9] Heriberto Cuayáhuitl, Simon Keizer, and Oliver Lemon. 2015. Strategic Dialogue Management via Deep Reinforcement Learning. In *NIPS Deep Reinf. Learn. Work.* arXiv:1511.08099

[10] Heriberto Cuayáhuitl, Seunghak Yu, Ashley Williamson, and Jacob Carse. 2016. Deep Reinforcement Learning for Multi-Domain Dialogue Systems. In *NIPS Work. Deep Reinf. Learn.* 1–9. arXiv:arXiv:1611.08675v1

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2018).

[12] Mihail Eric and Christopher D Manning. 2017. Key-Value Retrieval Networks for Task-Oriented Dialogue. *arXiv* (2017). arXiv:1705.05414

[13] Matthew Frampton and Oliver Lemon. 2005. Reinforcement Learning Of Dialogue Strategies Using The User's Last Dialogue Act. In *Proc. IJCAI Work. Knowl. Reason. Pract. Dialogue Syst.* 83–90.

[14] Wendong Ge and Bo Xu. 2015. Dialogue Management Based on Multi-domain Corpus. In *Annu. Spec. Interes. Gr. Discourse Dialogue.* 364–373.

[15] David Griol, L Hurtado, Encarna Segarra, and Emilio Sanchis. 2008. A Statistical Approach to Spoken Dialog Systems Design and Evaluation. *Speech Commun.* 50, 8-9 (2008), 666–682. https://doi.org/10.1016/j.specom.2008.04.001

[16] R. Chulaka Gunasekara, David Nahamoo, Lazaros C. Polymenakos, David Echeverría Ciaurri, Jatin Ganhotra, and Kshitij P. Fadnis. 2018. Quantized Dialog âĂŞ A General Approach for Conversational Systems. *Comput. Speech Lang.* 54 (2018), 17–30. https://doi.org/10.1016/j.csl.2018.06.003

[17] Matthew Henderson. 2015. Machine Learning for Dialog State Tracking: A Review. In *Proc. First Int. Work. Mach. Learn. Spok. Lang. Process.*

[18] Matthew Henderson, Blaise Thomson, and Steve Young. 2014. Robust Dialog State Tracking Using Delexicalised Recurrent Neural Networks and Unsupervised Adaptation. In *2014 IEEE Work. Spok. Lang. Technol. SLT 2014 - Proc.* 360–365. https://doi.org/10.1109/SLT.2014.7078601

[19] Yangfeng Ji, Gholamreza Haffari, and Jacob Eisenstein. 2016. A Latent Variable Recurrent Neural Network for Discourse Relation Language Models. In *NAACL-HLT 2016.* Association for Computational Linguistics, 332–342. arXiv:1603.01913

[20] Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. 1997. *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual.* Technical Report. 97–102 pages.

[21] Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Daniel Jurafsky. 2016. Deep Reinforcement Learning for Dialogue Generation. In *Proc. 2016 Conf. Empir. Methods Nat. Lang. Process.*

[22] Anthony J. Liddicoat. 2007. *An Introduction to Conversation Analysis.* Continuum. 319 pages.

[23] Carlos D. Martínez, Emilio Sanchis, F. García, and Pablo Aibar. 2002. A Labelling Proposal to Annotate Dialogues. In *Proc. Third Int. Conf. Lang. Resour. Eval.* 1566–1582.

[24] T Daniel Midgley, Shelly Harrison, and Cara Macnish. 2006. Empirical Verification of Adjacency Pairs Using Dialogue Segmentation. In *Proc. 7th SIGdial Work. Discourse Dialogue.* Association for Computational Linguistics, 104–108.

[25] Nikola Mrkšić, Diarmuid Ó Séaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve Young. 2017. Neural Belief Tracker: Data-Driven Dialogue State Tracking. In *Proc. 55th Annu. Meet. Assoc. Comput. Linguist. (Volume 1 Long Pap.* Association for Computational Linguistics, 1777–1788. https://doi.org/10.18653/v1/P17-1163 arXiv:1606.03777

[26] Nicole Novielli and Carlo Strapparava. 2009. Towards Unsupervised Recognition of Dialogue Acts. In *Proc. NAACL-HLT Student Res. Work. Dr. Consort.* Association for Computational Linguistics, 84–89.

[27] Kechen Qin, Lu Wang, and Joseph Kim. 2017. Joint Modeling of Content and Discourse Relations in Dialogues. In *Proc. 55th Annu. Meet. Assoc. Comput. Linguist.* Association for Computational Linguistics, 974–984. https://doi.org/10.18653/v1/P17-1090 arXiv:1705.05039

[28] Abhinav Rastogi, Dilek Hakkani-Tür, and Larry Heck. 2017. Scalable Multi-Domain Dialogue State Tracking. In *2017 IEEE Autom. Speech Recognit. Underst. Work.*

[29] Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language (Baltim).* 50, 1 (1974), 696–735.

[30] Emanuel A. Schegloff. 2007. *Sequence Oranization in Interaction: A Primer in Conversation Analysis I.* Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

[31] Elizabeth Shriberg, Raj Dhillon, Sonali Bhagat, Jeremy Ang, Hannah Carvey, and C S U Hayward. 2003. The ICSI Meeting Recorder Dialog Act (MRDA) Corpus. In *SIGDIAL 2004.* 97–100.

[32] Jason D. Williams, Antoine Raux, and Matthew Henderson. 2016. The Dialog State Tracking Challenge Series: A Review. *Dialogue & Discourse* 7, 3 (2016), 4–33. https://doi.org/10.5087/dad.2016.301

[33] Steve Young, Milica Gašić, Blaise Thomson, and Jason D. Williams. 2013. POMDP-based Statistical Spoken Dialog systems: A review. *Proc. IEEE* 101, 5 (2013), 1160–1179. https://doi.org/10.1109/JPROC.2012.2225812

[34] Tiancheng Zhao and Maxine Eskenazi. 2016. Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning. In *SIGDIAL 2016.* 1–10. https://doi.org/10.18653/v1/W16-3601 arXiv:1606.02560