

# Finding Structure in Conversations with PhotoPal

**Genevieve Gorrell**

The University of Sheffield,  
Department of Information Studies,  
Regent Court, 211 Portobello,  
Sheffield, UK, S1 4DP,  
g.gorrell@dcs.shef.ac.uk

**Peter Wallis**

The University of Sheffield,  
Department of Computer Science,  
Regent Court, 211 Portobello,  
Sheffield, UK, S1 4DP,  
pwallis@acm.org

## Abstract

Creating natural-sounding dialogue management in a conversational system is a challenging problem, as flexible sharing of initiative and topic switching is required. We present an analysis of a corpus of dialogues collected via Wizard of Oz experiments in a conversational domain. Repetitive structural components are uncovered. These components might be used as the basis of a conversational dialogue manager.

## 1 Introduction

Dialogue systems often fall into two categories. Conversational systems (ALICE, 2008; Weizenbaum, 1966) are broad-ranging but shallow—typically little attempt is made to interpret the user’s utterance usefully. Task-focused systems aim for complete understanding but channel the user into a highly restricted interaction (Walker et al., 2001). PhotoPal is a system proposed by the Companions consortium (Companions, 2008), which will discuss the user’s photographs with them. The system asks the user to introduce the people in the picture, discuss the relevance of the picture to them personally and similar, with the intention that the user should enjoy the conversation and the system should learn the life narrative of the user. Thus the conversation is natural and broad-ranging, and real semantic interpretation is attempted.

Attempts along these lines (Batacharia et al., 1999) run into trouble with flexible initiative sharing and topic detection. Therefore it is useful to study such dialogues and discover their structure. Structure may allow us to predict topic change and allow the system to share the initiative in the manner of a human being. We present findings from analysis of a corpus of Wizard of Oz dialogues with PhotoPal. Repetitive patterns are uncovered, which may be automatable.

## 2 Method

Two corpora were selected, with different “wizards”. In each case, the interviewer has been briefed regarding how the dialogue should be conducted, so there is no attempt here to study naturally-occurring dialogues. The intention is to study somewhat artificial but domain-relevant dialogues under contrasting conditions. The first dialogue sample comprises 111 dialogue turns, where a turn is a complete utterance by one speaker. A dialogue turn may have been divided into multiple sections, where each section is tagged with a different dialogue act, so the total number of annotations amounts to a greater number than 111. The second dialogue sample comprises 82 dialogue turns.

The dialogue act set used in this work is a subset of DAMSL dialog act tags (DAs) (Jurafsky et al., 1997). The DAMSL tags are intended for task driven dialogues and as such not all of them were relevant to the casual nature of conversation with PhotoPal. Below is a description of the set of domain-appropriate tags used in the analysis process.

ASSERT—“This is my sister”. Also yes and no.
OFFER—“Shall we look at another picture?”
COMMIT—“Okay I’ll do that”
EXPRESSION—Social expressions such as “wow!” and “great!”
INFORMATION REQUEST—open question
CONFIRMATION REQUEST—yes/no question
REPEAT REQUEST—“Pardon?”
ACTION DIRECTIVE—“Show me another one.” Imperatives.
OPEN OPTION—“We could look at another picture.”
OPENING—“Hi”
CLOSING—“Goodbye”
ANSWER—An answer is invariably also an assert.
BACKCHANNEL—“Uhuh”
REPEAT REPHRASE—Paraphrasing
COMPLETION—Completing the utterance of the other speaker
NON-UNDERSTANDING—“I don’t understand”
CORRECTION—An assertion that corrects a previous assertion
ACCEPT—Accepting a proposal
REJECT—Rejecting a proposal

Table 1: Dialogue Act Set

In tables 2 and 3 an analysis of the frequency of each dialogue act in each corpus is presented.

In dialogue corpus 1, the dominant dialogue act is ASSERT, at 39%, as the subject predominantly communicates facts. ASSERT includes yes/no answers to CONFIRMATION REQUESTs. The next most common dialogue act is the INFORMATION REQUEST. The next two most common dialogue acts are ACCEPT and BACKCHANNEL. CONFIRMATION REQUEST appears only in fifth place, followed by OFFER. EXPRESSION comes last, with only two percent of the whole.

Dialogue Act	Ex	S	U	%
Assert	53	3	50	39
Information Request	29	29	0	21
Accept	18	1	17	13
Backchannel	13	11	2	10
Confirmation Request	11	10	1	8
Offer	10	10	0	7
Expression	2	2	0	2
Total	136	66	70	100

Table 2: Dialogue Act Incidence in Corpus 1

In the second corpus, again ASSERT is the most common

Dialogue Act	Tot	S	U	%
Assert	67	10	57	61
Confirmation Request	21	21	0	19
Information Request	17	17	0	16
Expression	1	1	0	1
Open Option	1	0	1	1
Backchannel	1	1	0	1
Offer	1	1	0	1
Total	109	51	58	100

Table 3: Dialogue Act Incidence in Corpus 2

Interchange lengths			
corpus I		corpus II	
Moves	Examples	Moves	Examples
9	1	20	1
6	3	19	2
5	5	10	2
4	3	5	1
3	7	4	1
2	21	3	3
1	6	1	8

Table 4: The length of structures relating to a root (in moves) and the number of examples of structures that size.

dialogue act, and in this case accounts for most of the data. In this case, CONFIRMATION REQUEST comes in second. In third place is the INFORMATION REQUEST. Other dialogue acts occur rarely in this corpus. The comparative absence of OFFER and ACCEPT might be accounted for by the reduced tendency on the part of this interviewer to overtly suggest to the subject that they move on to the next photo. Comparative absence of BACKCHANNEL might simply be a stylistic variation.

### 3 Structures in the Dialogue

The purpose of this paper is to show how topic-like structures can be found from a corpus in a systematic manner. In this paper, a dialogue structure is any sequence within which utterances respond to each other. An utterance that isn't a response to anything is considered a **root** and, potentially, introduces a new topic. For example, a root may be followed by a sequence of utterances that elaborate on the root. On the other hand an utterance may be surrounded by other utterances that do not form a part of it. For example, one speaker may attempt to introduce a new topic only to have the other speaker ignore the gambit and persist with the original topic.

The corpora have been divided into dialogue structures, as described earlier, and the structures sorted by length. Table 4 presents the number of dialogue structures of each length in corpus 1. Table 4 presents the number of dialogue structures of each length in corpus 2. The characteristics of corpus 2 are quite different to those found in the first corpus. Where the first corpus had more structures of length 2 than any other, the second corpus contains no structures of length 2. It is skewed toward much longer structures. Most common structures are ignored gambits, simply-answered questions and accepted offers. However, these short structures do not dominate the dialogues, because the rarer structures are also longer. The longer structures share some characteristics.

### 3.1 Anchors

In the second corpus, the interviewer made ASSERTs, INFORMATION REQUESTs (open questions) and CONFIRMATION REQUESTs (yes/no questions) in response to the immediately preceding system utterance around half the time. The rest of the time the interviewer referred to interviewee utterances of particular interest. Do these "anchors" share any features such that it might be possible for a machine to identify them as being of particular interest? If the anchors could be detected and the machine made to refer back to them frequently, then this behaviour might be interpreted by the user as human-like and pleasing. Ten utterances were found in the corpus which were returned to at least once (the immediately subsequent utterance does not count as a "return"). The most returns to an utterance was 4. Some preliminary observations were generated based on these "anchors". Mentioning a new person by name for the first time serves to attract returns by the interviewer, as more information is elicited about that person. Additionally, questions insufficiently addressed will be returned to by both parties. Utterances from the interviewee that do not supply new information, for example, mentioning people already introduced, do not serve to attract returns. As an illustration, the most returned-to utterance in the corpus was this: "We were four people, four of us. Me and my friend M a Japanese friend M P, and my Swedish friend E. Who was my Swedish friend who got kicked out of Sweden for being not a Swedish [inaudible 00:07:19] that's what he says! He lives in Sweden now."

## 4 Conclusion

We introduced the notion of roots and anchors, and their relationship to topics. Dialogue segments are groups of utterances linked by the response relationship. Roots are utterances that begin a new dialogue segments. Anchors are utterances within a segment that attract several responses, serving as foci of interest. We presented a small corpus study that illustrates a range of dialogue behaviours in natural free-ranging conversation focusing on photographs, and describe these behaviours with regard to roots and anchors. We suggest that the approach has the potential to be useful in the creation of natural-sounding artificial conversational agents.

## References

- The ALICE artificial intelligence foundation. 2008 <http://www.alicebot.org/documentation/>.
- The Companions Project. 2008 <http://www.companions-project.org/>.
- Dan Jurafsky, E. Shriberg, and D. Biasca. 1997. *Technical Report 97-01, University of Colorado Institute of Cognitive Science, Colorado, 1997.*
- Joseph Weizenbaum. ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), January 1966.
- B. Batacharia, D. Levy, R. Catizone, A. Krotov and Y. Wilks CONVERSE: a Conversational Companion *Machine Conversations* 1999.
- Walker et al. DARPA Communicator Dialog Travel Planning Systems: The June 2000 Data Collection *Proceedings of Eurospeech 2000.*