

Viability of a Simple Dialogue Act Scheme for a Tactical Questioning Dialogue System

Ron Artstein Sudeep Gandhe Michael Rushforth David Traum

Institute for Creative Technologies, University of Southern California

13274 Fiji way, Marina del Rey, CA 90292, USA

<lastname>@ict.usc.edu

Abstract

User utterances in a spoken dialogue system for tactical questioning simulation were matched to a set of dialogue acts generated automatically from a representation of facts as ⟨object, attribute, value⟩ triples and actions as ⟨character, action⟩ pairs. The representation currently covers about 50% of user utterances, and we show that a few extensions can increase coverage to 80% or more. This demonstrates the viability of simple schemes for representing question-answering dialogues in implemented systems.

1 Introduction

Dialogue acts are often used as representations of the meaning of utterances in dialogue, both for detailed analyses of the semantics of human dialogue (e.g., Sinclair and Coulthard, 1975; Allwood, 1980; Bunt, 1999) and for the inputs and outputs of dialogue reasoning in dialogue systems (e.g., Traum and Larsson, 2003; Walker et al., 2001). There are many different taxonomies of dialogue acts, representing different requirements of the taxonomizer, both the kinds of meaning that is represented and used, as well as specifics of the dialogues and domain of interest (Traum, 2000). There are often trade-offs made between detailed coverage and completeness, simplicity for design of domains, and reliability for both manual annotation and automated recognition.

In this paper, we examine the adequacy for use in tactical questioning characters of a fairly simple dialogue act scheme in which the set of possible dialogue acts is automatically created by applying illocutionary force constructor rules to a set of possible semantic contents generated by an ontology of a domain. The advantage of this kind

of scheme is that a dialogue system is fairly easily authored by domain experts who work on the level of a simple ontology, without detailed knowledge of dialogue act semantics and transitions. The disadvantage is that it (intentionally) has limited expressibility in that some dialogue functions are not directly expressible, and it is not so easy to represent multiple meanings of an utterance.

We evaluated the scheme as follows: first we created an initial version of the character by authoring the ontology and using this to automatically generate the set of dialogue acts that fit into designed protocols for tactical questioning dialogues. Initial Natural Language Understanding and Generation capabilities were also authored using a classification approach (Leuski and Traum, 2008). The complete system was then used to generate a corpus of man-machine dialogues by having people interact with the character. Finally, the user utterances in this corpus were annotated by multiple annotators according to the dialogue act taxonomy. We evaluated both the coverage of the dialogue act taxonomy and the reliability of the annotations. The reliability of the matching was 49% above chance and full agreement was reached for only 30% of the utterances, but a detailed analysis shows that coverage of the current representation is closer to 50%, and that a few extensions can bring it to 80% or more.

The rest of the paper is structured as follows. Section 2 describes the tactical questioning genre of dialogue, and the dialogue system architectures that have been used to create specific domains and characters for this genre, as well as the development process for creating characters. The domain specification and dialogue representation is described in section 3. Section 4 presents the specific experiments, with the results presented in section 5, and a detailed analysis of the coverage of the dialogue act representation in section 6.

2 The Tactical Questioning Domain

Tactical Questioning is an activity carried out by small-unit military personnel, defined as “the expedient, initial questioning of individuals to obtain information of immediate value” (U.S. Army, 2006). A tactical questioning dialogue system is a simulation training environment where virtual characters play the role of a person being questioned. Unlike typical question-answering systems, tactical questioning characters are designed to be non-cooperative at times. The character may answer some of the interviewer’s questions in a cooperative manner, but may refuse to answer other questions, or intentionally provide incorrect answers (lie). Some of the strategies that an interviewer may use in order to induce cooperation include building rapport with the character, addressing their concerns, making promises and offers, as well as threatening or intimidating the character; the purpose of the dialogue system is to allow trainees to practice these strategies in a realistic setting.

Building tactical questioning dialogue systems is an on-going project at Institute for Creative Technologies, which has evolved through a number of different architectures; see Traum et al. (2008) for a detailed overview. The third and current architecture introduces an intermediate representation for dialogue acts, a finite-state representation of local dialogue segments, a set of policies for engaging in the network, and a rule-based dialogue manager to update the context and choose dialogue acts to perform (Gandhe et al., 2008). This functionality allows for short subdialogues where the character can ask for and receive certain assurances (such as protection or confidentiality) and still remember the original question asked by the trainee.

With earlier tactical questioning systems, based on text-to-text classifiers, character development typically proceeds in a bottom-up fashion: we start by collecting a corpus of in-domain human-human dialogues through roleplays or Wizard-of-Oz sessions, and use this as a starting point for the implementation of a question-to-response mapping. This mapping is refined as the system goes through iterative test cycles: additional user questions are gathered and mapped to appropriate responses, and the character’s domain is expanded by authoring new responses. The use of an intermediate representation for dialogue acts requires

top-down authoring: the first step is specifying the domain, that is the set of facts that the character can be questioned about; dialogue acts are created automatically from the domain specification, and these represent what the character can understand. When iterative testing with users reveals deficiencies or gaps in the character’s understanding capabilities, expansion cannot take place at the textual level but must go back to the domain specification or the rules for creating dialogue acts.

Our tactical questioning system is designed for rapid prototyping and creation of multiple characters with shared knowledge about a specific domain (Gandhe et al., 2009). The representation language for dialogue acts is therefore fairly simple, unlike that of more complex systems (Traum and Hinkelman, 1992; Traum and Rickel, 2002; Keizer and Bunt, 2006). The core of the representation language rests on facts represented as $\langle \text{object, attribute, value} \rangle$ triples, and which constitute the material for questioning by the user. For the system to succeed, this impoverished representation must capture enough information about the users’ actual utterances.

3 Domain specification and dialogue acts

In the scenario for the experiment, the user plays the role of a commander of a small military unit in Iraq whose unit had been attacked by sniper fire. The user interviews a character named Amani who was a witness to the incident and is thought to have some information about the identity of the attackers (Figure 1). Amani’s knowledge about the incident is represented as facts which are $\langle \text{object, attribute, value} \rangle$ triples; each fact is either true or false – false facts are used by Amani when she wants to tell a lie. Table 1 gives some facts about the incident. For example, Amani knows that the name of the suspected sniper is Saif, and that he lives in the store. She can lie and say that she doesn’t know the suspect’s name. She does not

Table 1: Some facts about the incident

Object	Attribute	Value	T/F
strange-man	name	saif	true
strange-man	name	unknown	false
strange-man	location	store	true
brother	name	mohammed	true



Figure 1: Amani – A virtual human for Tactical Questioning. The figure sitting in the chair represents Amani’s brother, Mohammed, who is not an interactive character.

have an available lie about the suspect’s location, though she can always refuse to answer a question.

In addition to facts about the incident, the domain specifies certain attributes that are unique to the characters (both Amani and the user). Characters may have attitudes towards objects; they can perform actions such as offers, threats, admissions and suggestions; and they have a set of compliments and insults that they can use for building rapport with their interlocutors. All of these, together with the facts, are specified in an XML format that defines the domain of interaction (Gandhe et al., 2008; Gandhe et al., 2009).

The domain represents the character’s knowledge. It defines a space of dialogue acts which are the interpretations of language utterances; this is the level at which the character reasons about the conversation. Dialogue acts are automatically generated from the domain specification, by applying an illocutionary force (or *dialogue act type*) to a semantic content containing the relevant portion of the domain specification. Each fact generates 3 dialogue acts – an assertion of the fact by the character, a yes-no question by the user, and a wh-question by the user which is formed by abstracting over the value. For example, the fact ⟨strange-man, name, saif⟩ defines a dialogue act by Amani with a meaning equivalent to “the suspect is named Saif”, and two questions by the user, equivalent in meaning to “is the suspect named Saif?” and “what is the suspect’s name?” (note

Table 2: Dialogue acts in the Amani domain

Dialogue Act Type	Amani	User
accept	1	1
ack	1	1
apology	1	1
assert	36	
closing	3	3
compliment		3
elicit	6	
greeting	1	1
insult		2
offer		3
offtopic	1	1
pre_closing	3	3
refuse_answer	1	1
reject	1	1
repeat-back	10	10
request-repair-object	10	10
request_repair	1	1
response	54	3
thanks	1	1
unknown		1
whq		31
ynq		35

that distinct facts may give rise to identical question dialogue acts). Each user action generates a corresponding dialogue act, as well as forward-function (elicitation) and backward-function (response) dialogue acts by the character (Allwood, 1995; Core and Allen, 1997). Currently, elicitation are only defined for offers (so Amani can ask for a particular offer); responses of various kinds are defined for all of the user’s illocutionary acts (offers, threats, compliments, insults). Additionally, some generic dialogue acts are defined independently of the domain – these include greetings, closings, thanks, grounding acts (such as repeat-back or request-repair), and special dialogue acts that are designed to handle out-of-domain dialogue acts from the user. Table 2 shows the various dialogue act types used in the current tactical questioning architecture and the number of full acts of each type generated for the user and Amani, given Amani’s ontology. The full algorithm for generating dialogue acts is presented in Gandhe et al. (2009).

The link between dialogue acts and actual utterances is done via Natural Language Understand-

ing and Generation modules. The NLU uses a statistical language modeling text classification technique (Leuski and Traum, 2008), trained on pairings of user utterances to dialogue acts, to determine the appropriate dialogue act for novel text produced by the speech recognizer; if it cannot find a good match with high confidence, the classifier outputs a special “unknown” dialogue act which informs the dialogue manager that the user utterance has not been properly understood. A similar classifier, trained on mappings from character dialogue acts to text, is used for generation. A dialogue manager is responsible for the transition from user dialogue acts, provided by the NLU module, to character dialogue acts which are passed to the NLG module. The dialogue manager is based on the information state model (Traum and Larsson, 2003). It uses rules described in State Chart XML (Barnett et al., 2008) to keep track of obligations (Traum and Allen, 1994), questions under discussion, offers and threats; similar rules track the character’s emotional state (Roque and Traum, 2007) as well as grounding (Roque and Traum, 2009). The main responsibilities of the dialogue manager are to update the information state of the dialogue and use it to select the contents of the response.

The dialogue manager drives the character’s interaction and is responsible for all of its reasoning, and it works at the level of dialogue acts. But users have their own mental models of what can be said to the system, and are not aware of what distinctions the system can represent. We therefore need to determine whether the dialogue act representation – intentionally designed to be simple – is rich enough to capture the meaning in user utterances. To answer this question we carried out an experiment with actual user utterances.

4 Experiment

To test how well the automatically generated dialogue acts capture the meaning of actual user utterances, we performed a matching experiment. First, we collected a corpus of interactions of users with the initial version of Amani. The dialogue participants were all staff members at ICT; they had experience talking to virtual characters, including question-answering characters, but were not familiar with the Amani scenario prior to the dialogues, nor had any experience talking to a third-generation question-answering charac-

ter. Dialogue participants were given an instruction sheet with some information about the incident, the character, and suggestions for interaction (e.g. the possibility of making offers) – similar to the instruction sheet a trainee would receive. The instructions did not include guidance about particular language to use with the character. We collected a total of 261 user utterances from 16 dialogues, which varied in length from 2 to 40 utterances.

User utterances from interactions with the system were transcribed, and then matched to the existing user dialogue acts by 3 experienced annotators. The annotators were all involved with the project: they included the first and third authors, and a student annotator. The purpose of the study was to find out how adequate the current domain representation was, what extensions it needed, and what systematic problems arose that might require not only changes to the domain specification but to the way dialogue acts are defined. Since this study was of an exploratory nature, the instructions were very simple and given in a single sentence: “Match each user utterance to the most appropriate player speech act; if none is appropriate, match to ‘unknown’.”

Annotators matched utterances to dialogue acts using the domain creation tool (Gandhe et al., 2009). We proceeded under the assumption that each utterance text is mapped to a single dialogue act, not taking into account context that would disambiguate different dialogue acts for the same text appearing at different times. This was not a major concern with our corpus, because the vast majority of utterance texts occur only once (224 distinct utterance texts), and of the 7 utterance texts with frequency of 3 or more, 6 are greetings or closings. The analysis below is therefore on utterance texts, ignoring how many times these utterances appeared.¹

5 Reliability

As a means of checking that the annotators had a similar understanding of the task, we calculated inter-annotator reliability using Krippendorff’s α (Krippendorff, 2004).² Reliability cannot be taken

¹A more extensive study would have to look at the frequency of utterance texts and at the classification of text-identical user utterances to distinct dialogue acts when they occur in different contexts.

²Krippendorff’s α is a chance-corrected agreement coefficient, similar to the more familiar K statistic (Siegel and

Table 3: Inter-annotator reliability

	α	$A_o^{(a)}$	$A_e^{(a)}$
Dialogue act	0.489	0.545	0.109
Dialogue act type	0.502	0.585	0.166
Matches domain	0.383	0.741	0.580

^aKrippendorff’s α is defined in terms of observed and expected disagreement: $\alpha = 1 - D_o/D_e$. For expository purposes we have converted these into values representing observed and expected agreement: $A_o = 1 - D_o$, $A_e = 1 - D_e$.

as a measure of the reproducibility of the annotation procedure, since the annotators were not working from detailed written guidelines, and any shared understanding must therefore come from their previous experience. Rather, reliability is indicative of how straightforward the task is before implementing corrective measures such as detailed guidelines and domain and dialogue act improvements. Table 3 shows the results of the agreement study on three sets of data: the top row is the annotators’ mapping of utterances to individual dialogue acts; the middle row is derived from the actual annotation by replacing each dialogue act with its type; and the bottom row treats “unknown” as one category and collapses all the other dialogue acts into a second category, marking a decision of whether the utterance fits at all to any of the existing dialogue acts.

Reliability was substantially above chance, though not as high as typically accepted norms; it can definitely be improved with clearer annotation guidelines (see section 6 below). An important source of disagreement was whether an utterance was a good enough match for an existing dialogue act: while observed agreement on this distinction is necessarily higher than on the dialogue act or dialogue act type, reliability (or chance-corrected agreement) is substantially lower, due to the fact that much higher agreement is expected by chance. Choosing the threshold for matching an utterance to a dialogue act is a known problem for the clas-

Castellan, 1988). Like K, α ranges from -1 to 1 , where 1 signifies perfect agreement, 0 obtains when agreement is at chance level, and negative values show systematic disagreement. The main difference between α and K is that α takes into account the magnitudes of the individual disagreements; in this study we treated all disagreements as equivalent, so α is essentially equivalent to K except that α employs a small correction for sample size. For additional background, definitions and discussion of agreement coefficients, see Artstein and Poesio (2008).

sifier, which uses a single threshold that represents the optimal balance between false positives (inappropriate matches above threshold) and false negatives (appropriate matches below threshold); the study shows that this is a difficult task for human judges as well. One judge marked 89 utterances as “unknown”, another marked 79, while the third judge marked only 33 utterances as “unknown”.

The study also shows that when annotators agreed on the dialogue act type, they typically also agreed on the on the dialogue act itself: observed agreement on dialogue act types is not much higher than on dialogue acts, and reliability (or chance-corrected agreement) shows an even smaller difference. To make the analysis simpler, we proceed with the analysis of the individual utterances using the dialogue act type alone.

6 Utterance analysis

A total of 72 user utterances were marked with an identical dialogue act type (other than “unknown”) by all the annotators. These included some straightforward greetings (such as *Hello Amani*), compliments (*You have a beautiful home*), thanks (*Thank you that helps a lot*), closings (*Goodbye madam*), offers – both explicit (*I promise to keep this discussion secret*) and implicit (*Everything you tell me is in confidence*), and questions (*What is the name of the man with the large gun*). While these account for just under 30% of the total utterance types, this shows that the existing dialogue act representation already provides for substantial coverage of what users say.

Some additional disagreements are fairly easily fixed. There are 24 disagreements on question type, of which 15 include the phrase *do you know* or *can you tell/describe*, for example *Do you know the name of the sniper?* These are formally yes/no questions but carry the impact of a wh-question, and a cooperative positive response would provide the sought-after information; the difference between asking a *can you tell/do you know* question and a direct wh-question is that the former allows a “no” response (or a non-cooperative “yes”), whereas the latter requires a phrase or sentence as a response. However, in order to make communications clearer, our tactical questioning characters are designed to always give fuller answers than a simple yes or no, so the distinction is immaterial. We could extend the dialogue act representation to represent *can you tell/do you know* questions,

but even though this type of question is rather frequent, distinguishing it from direct wh-questions would have little impact on the system, so a better guideline would be to treat these as wh-questions.

Other disagreements between question types are related to the domain specification. For example, the question *Have you seen him around lately* is clearly a yes/no question, but it is not an exact match to an existing dialogue act. The domain does specify the fact ⟨strange-man, last-seen, yesterday⟩, which all annotators found to be a close enough match to the user utterance. However, one annotator matched it to the wh-question derived from this fact (equivalent in meaning to “when did you last see him?”), whereas the two others matched it with the corresponding yes/no question (equivalent to “did you last see him yesterday?”). It is not clear what sort of guidelines would bring uniformity to this type of disagreements, but like the previous type, this is not expected to affect system performance.

Certain greetings were also the cause of disagreement that can probably be reconciled with more explicit annotation guidelines. There was confusion as to how to mark formulaic greetings which are literally questions (e.g. *How are you?*) or statements (*it's nice to meet you*). This can be solved through an explicit guideline to mark them as greetings, or by adding corresponding facts to the domain specification and matching these utterances to the literal dialogue acts. The first solution would be more useful for affecting the character's emotion and rapport (since she will understand these as greetings), while the second would allow more specific responses.

Other disagreements that can probably be alleviated to some extent result from confusion among the annotators about the distinctions between certain pairs of dialogue acts – accept and acknowledge, closing and pre-closing, request-repair and repeat-back. These, together with the greetings and questions discussed above, constitute 55 utterances; together with the utterances on which there is full agreement there are 127 user utterances (57% of all utterance types) which can be classified properly into dialogue acts using the current domain specifications.

The remaining user utterances are not covered by the existing dialogue acts. However, simple extensions can account for many of them. The most common utterances in this class are questions

about an object but without a specific attribute, such as *Can you tell me about the shooter?* Our corpus contains 26 such questions, that is almost 12% of all question types. To deal with these utterances we added a new type of dialogue act – a wh-question with just an object and no attribute. These dialogue acts are generated automatically for all objects in the domain, and corresponding policies have been added to the dialogue manager.

An additional 16 user utterances (7%) are simply not in the domain: for example, the question *Do you own a gun?* does not have a corresponding fact, but it would be very easy to add one, and an appropriate dialogue act would be generated automatically. A small number of user questions cannot be represented through existing dialogue acts even though the relevant facts exist in the domain specification. For example, the user utterance *Can you tell me who lives on top of Assad's shop?* is fully answered by the fact ⟨strange-man, location, store⟩ – but we do not generate dialogue acts that ask which object has a known attribute and value. Since such questions are relatively rare in our corpus (only 4), we decided against generating this type of dialogue act, opting instead to represent the questions that do arise as independent facts, so the above fact is now also represented as ⟨the-shop, occupant, strange-man⟩. This is a compromise solution, because the character is not aware that these two facts in the domain are essentially identical in content. The advantage of this duplication of facts is in keeping the domain simple, without generating an inflated space of dialogue acts which are rarely encountered in practice.

Overall, almost 50 user utterances fall into the above classes – utterances that can be represented using the ⟨object, attribute, value⟩ scheme by either adding facts to the domain or extending the dialogue acts generated from these facts. Together with the utterances discussed previously, these account for nearly 80% of the user utterances.

The remaining utterances are a mixed bag. Sometimes a user asks Amani to clarify an elicitation request, as in *Which promises do you want to hear?* or *Are you worried about your safety?* The system used in the experiment had no corresponding dialogue acts, but these have since been added. Several compound utterances correspond to more than one dialogue act – the utterance *Amani, if I offer you and your family protection can you lead me to the sniper?* contains a conditional offer and

a question. These will be dealt with using a separate utterance segmenter which is under development. Some utterances are inherently vague (perhaps intentionally). For example, when the user says *Your safety is very important to us* in response to a request for a guarantee of safety, it is not clear whether an offer has been made (there are 10 such utterances in our corpus). Some utterances contain rather obscure references; for example, in response to Amani's assertion that many Iraqis have guns, the user says *Wanna see mine?* which should probably be understood as a threat. The question *Can you tell me something useful?* was taken to be an insult by one annotator. One utterance, *Hello Mohammed*, is addressed to Amani's brother who is not an interactive character. Each of these types of utterances would require a different strategy in order to allow the character to understand it. Developing such capabilities for all of these utterances would be beyond the scope of the tactical questioning system, but this is not really necessary: there will always be some utterances that the character cannot understand, and the dialogue manager is designed to deal with this situation by providing off-topic responses or allowing the character to take initiative. The study shows that the vast majority of user utterances can be understood using the simple dialogue act representation language, and this is sufficient for tactical questioning characters.

7 Conclusion

This study has shown that from a simple representation of facts as ⟨object, attribute, value⟩ triples and actions as ⟨character, action⟩ pairs we can automatically generate dialogue acts that provide substantial coverage for interpreting user utterances spoken to a tactical questioning dialogue character. We have identified a few deficiencies in the dialogue act generation process, most notably requiring additional types of questions, which have been corrected in subsequent development. An extended system with an expanded domain and additional dialogue act types has been recently tested in the field with a large number of new users, and we are currently working on analyzing the results. We expect this new study to give a more accurate estimate of the proportion of user utterances covered by the representation.

One limitation that emerges from the current study is the linking of only one dialogue act per

utterance, which makes it more difficult to capture the multifunctionality of dialogue. For example, many utterances which have an illocutionary effect such as greetings, threats, and insults can be phrased in the form of a question which may also be relevant in the domain. Some functions can be computed automatically from the main dialogue act applied to the context, but some inferences are more challenging and would be better served by labelling multiple acts directly, which would complicate both the authoring and annotation tasks. Representing multiple facets of such an utterance without implementing to a full inference chain which calculates implicatures and illocutionary force from literal meanings remains a challenge for future research.

Acknowledgments

We wish to thank Nicolle Whitman and Sarrah Ali, who helped in developing the initial tools and authoring the initial Amani scenario. Thanks also to 3 anonymous reviewers for their thoughtful comments and suggestions. We alone are responsible for any errors and omissions.

The project or effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

- Jens Allwood. 1980. On the analysis of communicative action. In M. Brenner, editor, *The Structure of Action*. Basil Blackwell. Also appears as Gothenburg Papers in Theoretical Linguistics 38, Dept of Linguistics, Göteborg University.
- Jens Allwood. 1995. An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg.
- Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.
- Jim Barnett, Rahul Akolkar, R. J. Auburn, Michael Bodell, Daniel C. Burnett, Jerry Carter, Scott McGlashan, Torbjörn Lager, Mark Helbing, Rafah Hosn, T. V. Raman, and Klaus Reifensath. 2008. State Chart XML (SCXML): State machine notation for control abstraction. <http://www.w3.org/TR/scxml/>, May.

To appear in the proceedings of *DiaHolmia, the 13th Workshop on the Semantics and Pragmatics of Dialogue, Stockholm, Sweden, June 2009*.

- Harry C. Bunt. 1999. Dynamic interpretation and dialogue theory. In Michael M. Taylor, F. Néel, and Don G. Bouwhuis, editors, *The Structure of Multimodal Dialogue, Volume 2*. John Benjamins, Amsterdam.
- Mark G. Core and James F. Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, Cambridge, Massachusetts, November. AAAI.
- Sudeep Gandhe, David DeVault, Antonio Roque, Bilyana Martinovski, Ron Artstein, Anton Leuski, Jillian Gerten, and David Traum. 2008. From domain specification to virtual humans: An integrated approach to authoring tactical questioning characters. In *proceedings of Interspeech 2008*, Brisbane, Australia, September.
- Sudeep Gandhe, Nicolle Whitman, David Traum, and Ron Artstein. 2009. An integrated authoring tool for tactical questioning dialogue systems. In *6th Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Pasadena, California, July.
- Simon Keizer and Harry Bunt. 2006. Multidimensional dialogue management. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 37–45, Sydney, Australia, July. Association for Computational Linguistics.
- Klaus Krippendorff, 2004. *Content Analysis: An Introduction to Its Methodology*, chapter 11, pages 211–256. Sage, Thousand Oaks, California, second edition.
- Anton Leuski and David Traum. 2008. A statistical approach for text processing in virtual humans. In *Proceedings of 26th Army Science Conference*, Orlando, Florida, December.
- Antonio Roque and David Traum. 2007. A model of compliance and emotion for potentially adversarial dialogue agents. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 35–38, Antwerp, September.
- Antonio Roque and David Traum. 2009. Improving a virtual human using a model of degrees of grounding. In *Proceedings of IJCAI 2009*, Pasadena, CA, July.
- Sidney Siegel and N. John Castellan, Jr, 1988. *Non-parametric Statistics for the Behavioral Sciences*, chapter 9.8, pages 284–291. McGraw-Hill, New York, second edition.
- John McHardy Sinclair and Malcolm Coulthard. 1975. *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford University Press.
- David R. Traum and James F. Allen. 1994. Discourse obligations in dialogue processing. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 1–8, Las Cruces, New Mexico, June. Association for Computational Linguistics.
- David R. Traum and Elizabeth A. Hinkelman. 1992. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599. Special Issue on Non-literal language.
- David R. Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In Jan van Kuppevelt and Ronnie W. Smith, editors, *Current and New Directions in Discourse and Dialogue*, chapter 15, pages 325–353. Kluwer, Dordrecht.
- David R. Traum and Jeff Rickel. 2002. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the first International Joint conference on Autonomous Agents and Multiagent systems*, pages 766–773.
- David Traum, Anton Leuski, Antonio Roque, Sudeep Gandhe, David DeVault, Jillian Gerten, Susan Robinson, and Bilyana Martinovski. 2008. Natural language dialogue architectures for tactical questioning characters. In *Proceedings of 26th Army Science Conference*, Orlando, Florida, December.
- David Traum. 2000. 20 Questions on Dialogue Act Taxonomies. *J Semantics*, 17(1):7–30.
- U.S. Army. 2006. Police intelligence operations. Field Manual FM 3-19.50, U.S. Army. Appendix D: Tactical Questioning.
- Marilyn A. Walker, Rebecca Passonneau, and Julie E. Boland. 2001. Quantitative and qualitative evaluation of darpa communicator spoken dialogue systems. In *ACL '01: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 515–522, Morristown, NJ, USA. Association for Computational Linguistics.