Dialogue Generation in Character-based Interactive Storytelling

Marc Cavazza and Fred Charles

School of Computing, University of Teesside,
Middlesbrough, TS1 3BA, United Kingdom.
{ m.o.cavazza, f.charles }@tees.ac.uk

Abstract

In recent years, there has been significant progress in developing Interactive Storytelling systems, in particular in terms of the underlying AI techniques. There has been an emerging consensus on the AI approach, in particular the use of planning systems. However, these have concentrated on the generation of a sequence of narrative actions, staged through the behaviour of virtual actors. In order to achieve the long-term objective of implementing interactive media that would reproduce the aesthetic qualities of traditional films, most interactive storytelling systems are still missing the ability to generate dialogues between characters. We describe an extension of our Interactive Storytelling approach which integrates dialogue generation within narrative situations. Our unit of generation is the dialogue act (pair of utterances) and one main objective is to reproduce realistic dialogue phenomena based on implicit forms of expression. The emphasis is also on adopting a unified approach relating narrative representations to the linguistic form of the generated utterances. From a representation of active narrative goals and the emotional affinity between characters, our system is able to derive a set of parameters governing various aspects of the linguistic form to be generated. We illustrate this approach on several implemented examples.

Introduction and Objectives

Dialogues between feature characters are an essential component of traditional media, such as films and plays. In that respect, interactive narratives featuring virtual humans as actors have not yet fully addressed this aspect. Most work in interactive narratives remains centred on the generation of meaningful actions staged as 3D animations. While a certain consensus is emerging on the types of techniques (Cavazza et al. 2002) (Riedl and Young 2004) (Young 2004), which can produce sequences of narrative actions, it is important to investigate how dialogue can be generated as part of these actions. This is a natural extension of the narrative description, as many actions actually consist in communicative actions, as well as an essential aspect of the actual staging of the generated narrative considering the endeavour of obtaining film-like media.

Interactive Storytelling (IS) techniques should be made compatible with the requirements of dialogue generation to support some of the aesthetic components of traditional media (intended as including humour, witticisms, etc., so characteristic of real-world media productions such as plays, dramas, sitcoms, cartoons …).

While recognising that this constitutes a formidable challenge, in this paper we describe first steps for the principled integration of dialogues into plan-based interactive storytelling. Taking as a starting point our approach for the generation of virtual narratives in the sitcom genre, we investigate how narrative formalisms can be properly related to linguistic structures supporting dialogue generation.

IS can be described as comprising two levels. The first level is dedicated to the generation of narrative actions, some of which will consist of communicative actions between characters. The second level corresponds to the proper dramatisation of narrative actions. Physical actions and situations are staged essentially through camera placement. Communicative actions, on the other hand, are dramatised by the generation of appropriate dialogues. The generation of dialogue lines should be appropriate to the narrative genre considered, in particular in terms of linguistic phenomena.

Our main objective is to integrate dialogue generation within our Interactive Storytelling technology (Cavazza et al., 2002) in a way which should be consistent with the story genre and the overall dramatisation of narrative actions. In particular, the dialogue's linguistic contents should be able to reflect the emotional relations between the various characters in terms of style or use of dialogue moves.

Relation to Previous Work

Several researchers in IS have tackled the problem of dialogue, mostly between user and virtual actors. One of the most advanced systems in this regard is undoubtedly Façade (Mateas and Stern 2002) (Mateas and Stern 2004), a first-person IS system in which the user intervenes into a dispute between two characters through natural language input. Façade is one of the early attempts at relating narrative functions to dialogue acts in a principled fashion.

This is achieved through the fine-grained definition of dialogue acts some directly addressing narrative functions (such as "flirt", "pacify" or "advise"), while others correspond to generic communication functions (Mateas and Stern 2004). However the dialogue in Façade is more concerned with individual replies and dialogue progression than with narrative elements arising from the dialogue itself (such as witticisms, etc.). The latter aspect has been the major focus of our previous research (Charles et al. 2004) on the understanding of dialogue jousts in "James Bond" narratives.

Another approach taken in the MRE system consists in integrating strongly dialogue and narrative by resorting to a task-based dialogue (Traum et al. 2003), such a negotiation dialogue which naturally supports the mode of communication to be expected in taking decisions under critical conditions.

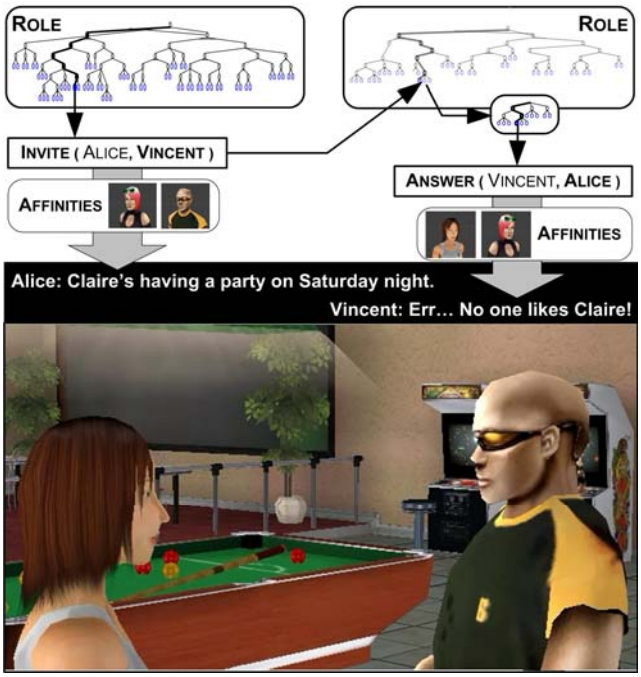

Figure 1: Characters' roles and the influence of their affinities towards other characters.

Context and Application

Our work on IS has been essentially based on the sitcom genre, which provides a manageable context for narrative generation. In addition, the importance of the notion of situation provides a framework for assessing the validity of the technical approach, as well as for early attempts at evaluating scalability (Charles and Cavazza 2004). We defined our approach as character-based interactive storytelling (Cavazza et al. 2002) to emphasise the specific stance it takes towards the duality of character and plot. The baseline plot is described through a set of independent roles for each character, which are formalised as HTN plans. During system operation, each character acts independently following its own plan. The dynamic interaction between characters and their competing for action resources (including other characters' availability) generates narrative situations whose sequence produces the story. One central aspect of the system, where dramatisation and generation mechanisms converge, corresponds to characters' actions failure. These are dramatised to constitute narrative elements, and at the same time lead to re-planning which is the basis for story generation and variability. This re-planning process also supports user intervention, as it influences the emerging story mostly through failure of characters' actions. This results in a combinatorics of plot elements, where a set of basic role definitions can generate several hundred different narratives.

Another advantage of the character-based approach is its ability to integrate various relevant representations around the plan-based representation of a character's role. In particular, it can include specific emotional representations describing the affinities between characters, which are often central to that narrative genre (Figure 1). Affinities between characters play a central role in selecting contextual response to a situation (for instance, joining or interrupting a conversation between characters). As we shall see, they will play an important role in the generation of dialogues between characters as well.

IS systems rely on a cast of virtual actors whose actions constitute the narrative. For many reasons, including historical ones, IS technology has developed by staging physical actions first, with very few exceptions (Mateas and Stern 2002). However, even when taking an overall physical stance, key narrative events are often better staged through dialogue.

Relations between Narrative and Dialogue

The interaction between characters constitutes most of the narrative action and as such should be not only meaningful, but should be staged with the same aesthetic preoccupations which characterise the sequencing of narrative actions to constitute a story. This means that the dialogue itself should be staged through a choice of linguistic expressions which should display the properties of a real dramatic dialogue (Figure 2).

Traditional narratology works at a level of description which stands above that of linguistic expression. It was Bremond (Bremond 1973) who first thoroughly investigated the relations between narrative roles and their translation in terms of linguistic expression as part of inter-character dialogue. This was a consequence of Bremond's narrative theory, which is centred on a classification of characters' roles which often hints at communicative aspects, as with the influencer role. Bremond proposed to map the narrative role onto rhetoric functions which will constitute the linguistic embodiment, and at the same time the dramatisation of a narrative action whose nature is communicative. This can be briefly illustrated by

considering Bremond's role of influencer (which is self-explanatory). A character wishing to influence another will use different rhetoric functions, such as advice, threats, seduction, for expressing that influence, depending on his understanding of the recipient's goals (Bremond 1970). To some extent, Bremond's proposal was the first one to attempt to properly map narrative functions to communicative acts. However, it did not address the conversational aspects of the characters' dialogue, nor did it specify the semantic content of corresponding utterances.

Figure 2: An invitation to a party.

In the next section, we propose an approach inspired from Bremond's model, but adapted to a computational approach. Firstly, we will use affinities between characters in the narrative in place of the various types of motivations described by Bremond. Secondly, instead of rhetorical functions (that characterise isolated utterances) we will generate specific dialogue acts which are better adapted to the staged narrative.

System Implementation and Experiments

The objective of the dialogue generator is to produce appropriate dialogue acts involving two characters within a situation generated by the storytelling system. Several narrative actions actually constitute communicative actions, such as: enquiries, advice, invitation, etc. The main example situation which we will use to support our discussion is the invitation of one character to a party, which is part of our latest experimental interactive narrative (Charles and Cavazza 2004), featuring the organisation of a party and the "politics" behind it.

From the IS perspective, the invitation is a terminal action which should either succeed or fail, depending on the response received from the invitee. However, very much like physical actions need to be properly staged through animation and camera placement, a communicative action needs to be dramatised through the selection of linguistic expressions that carry some of the style of the narrative genre. This should hold both for the invitation and the reply if a believable dialogue is to take place.

The dramatic value of the dialogue act lies not only in its relevance but also in its rhetorical value, which in turn can be said to derive from the contrast between two actors' utterances.

We have to consider the narrative action of invitation both from the story perspective and from the communication perspective. As an element of the story, it constitutes one of the sub-tasks in the course of organising the party and is related to the event representation for the party. As a communicative action, it involves two characters in the specific context of a network of affinities and feelings relating the various actors.

The rationale for our approach is to establish a unified model for narrative representations and communicative actions, which would be easier to translate into linguistic representations. The other aspect of this unified representation consists in using the matrix of affinities between the various characters to determine the linguistic form of utterances. This is also where specific properties of the sitcom genre should be taken into account. While it may not be possible to generate the kind of witticisms so characteristic of sitcoms, it is important for the utterances to reflect the style of sitcom dialogues as much as possible. We have retained two elements of style, which are the prevalence of implicit utterances and the fact that dialogue varies, sometimes exaggeratedly, according to the relations between characters.

Dialogues are constituted of at least one pair of utterances. Communicative actions as implemented in our IS system only generate short dialogues at this stage, which are meant to determine the success of the action and present a proper dramatisation of the linguistic exchange. In addition, a distinction should be made between opening statements, whose content is determined by the narrative situation only, and replies to previous utterances, which should take into account their contents as well.

In the next sections, we describe this approach through the various steps governing the generation of dialogue elements from a narrative situation, using our invitation example. We will describe this process both for opening/initial queries and for subsequent replies. In this first implementation we have only addressed short dialogues corresponding to a single communicative action, which are composed of two or three utterances. The various steps of utterance generation for the case of a reply are presented on Figure 3.

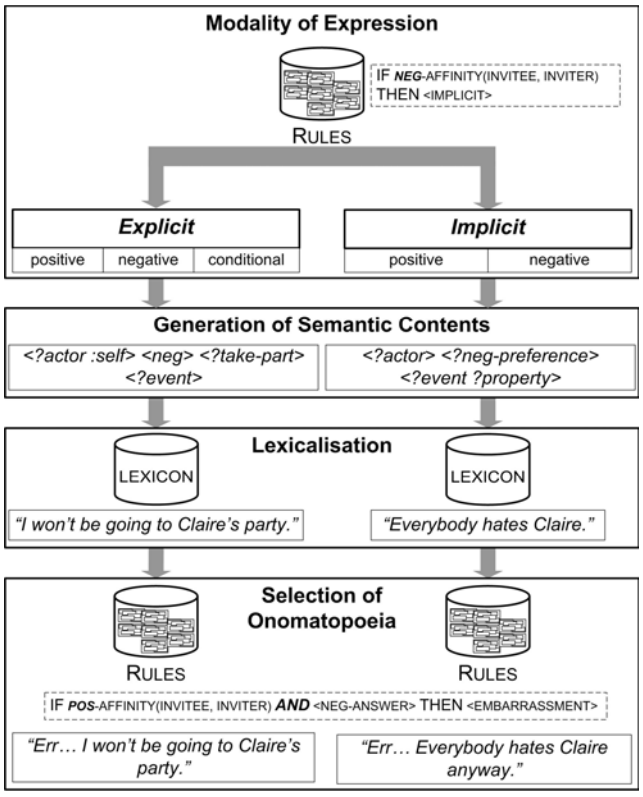

Figure 3: Steps for utterance generation in the case of a reply to an invitation to a party.

Affinity(Invitee, X)			**Modality of Expression**
Inviter	**Organiser**	**Guests**	
Positive	negative		***Explicit***
		negative	
	positive		
		positive	
Negative	negative		***Implicit***
		negative	
	positive		
		positive	
Very Positive	negative		***Justification***
		negative	
Inviter == Organiser		**Guests**	
Positive		Very negative	***Conditional***

Table 1: Some determinants for the selection of modality of expression.

The first step determines the modality of expression, in other words whether the utterance information will be conveyed explicitly (e.g. direct question or request) or implicitly (speech act). This choice applies both to opening statements and replies. Obviously, in the case of a reply, the first step should also determine the nature of the reply, i.e. positive, negative or conditional. These two aspects are determined by using a set of production rules which interpret the affinities between the various actors (inviter, organiser, invitee and other guests) in terms of the utterance parameters. Some of the determinants are illustrated in Table 1.

The next step should generate the actual semantic contents of the utterance, and is based on the event structure of the narrative action being considered. Work from Pustejovsky (Pustejovsky 1995) was the first one to relate the structure of an event to the lexical semantics of linguistic descriptions of that event. Without adopting his model in its entirety we will retain some of its elements to illustrate our approach. Let us consider the event of holding a party. The relation of each character to the party event depends on their role in the event: this corresponds to an agentive structure. For instance, guests are in a relation of participation (let us name this semantic relation <?take-part>) with the event, while organisers are in a relation of creating (similarly, <?make-happen>) the event. These relations can then be used to generate linguistic descriptions of that party event, by instantiating them with appropriate vocabulary.

One of our hypotheses is precisely that an implicit invitation can take the form of a partial description of that future event. These partial descriptions feature a given actor in its relation to the party event, such relation being determined by the event agentive structure (for instance a guest will "take part" in the party held). This gives us a semantic template such as [1]:

<?actor :type :guest> <?take-part> <?event :type :party> <?event ?property>

which would be lexicalised as "James will be coming to Claire's party", or

<?actor :type :organiser> <?make-happen> <?event :type :party> <?event ?property>

which would be lexicalised as "Claire is having a birthday party".

On the other hand, explicit invitations will be composed of an interrogative and part of the event's agentive structure. In the case of an invitation to a party:

<?interrogative> <?actor :type :addressee> <?take-part> <?event :type :party>

which would be lexicalised as "Would you come to Claire's party?"

When the utterance to be generated constitutes a reply to another character's occurrence, the basic principle consists

[1] Bracketed structures represent templates, symbols prefixed with "?" are variables (semantic categories) and symbols prefixed with ":" are static properties.

in producing a structure which complements the content of the previous occurrence.

This can be illustrated by two different approaches. One consists in responding by qualifying positively or negatively one of the requests' explicit topics (as in: "Claire is having her birthday party on Saturday" / "No-one likes Claire anyway"). The semantic templates for the implicit negative reply would thus look like:

<?actor> <?neg-preference> <?event ?property>

for instance: "I don't fancy parties" or "No-one likes Claire" where the event property is part of the contents of the request (constraint not represented here) (Figure 4).

Figure 4: Turning down an invitation (implicit refusal, see text).

Another form of implicit reply consists in producing an utterance whose semantic content is incompatible with the event structure considered. For instance if the implicit request is expressed as:

<?actor> <?take-part> <?event1 :type :party> <?event1 ?property>

generating invitations such as "Steve will be coming to Claire's party on Friday". A negative, implicit reply can be of the form:

<?actor :self> <?take-part> <?event2> ¦ where <?event2> = <?event1>

for instance, "I am going to a restaurant on Friday".

It can be noted that this determination of rejection by contrasting the semantic context of successive utterances has been previously described in human-computer dialogue (Walker 1996) (Cavazza 2003) for the analysis of dialogue acts and we are proposing here to adapt it to the dual problem of dialogue act generation.

The final step generates the surface form of the utterance. It does so by "lexicalising" the semantic categories that constitute the above semantic template. By lexicalising here, we also mean the attribution of relevant lexicon-centred syntactic structures, following a linguistic paradigm of lexicalised syntax, best exemplified by Tree-Adjoining Grammars (Joshi and Schabes 1992), which have been successfully used in natural language generation. The lexicalisation step i) guarantees the variability/diversity of generated utterances ii) adds some stylistic elements by matching the selected vocabulary to the actor's intention and iii) adds onomatopoeia and adverbs to achieve a more colloquial surface form.

Lexicalisation is the process by which correct syntactic structures are generated from the above semantic templates. The operations cannot always consist in a strict linear mapping from semantic units to lexicalised units. For instance the invitation

<?interrogative> <?actor> <?take-part> <?event :type :party> <?event ?property>

can be translated into "Would you like to come to Claire's party?". The most appropriate method for such generation is to use a lexicalised grammar, such as one derived from TAG (Joshi and Schabes 1992). Our current prototype uses ad hoc templates which compile some of the principles behind the TAG in terms of insertion of syntactic elements.

For those stylistic elements that confer additional believability to the dialogue, once again the nature and intensity of inter-personal relation are used to determine which forms should be generated: the choice of vocabulary for instance depends on the intensity of characters' affinities. In a similar fashion, onomatopoeias are selected depending on affinity equations. A character refusing an invitation but having a positive affinity with the actor inviting him will express embarrassment ("err…"). The insertion of other modifiers (e.g. adverbs, "really", "actually") follows similar principles (although through a small number of ad hoc rules).

The architecture of the dialogue module is represented in Figure 5. It comprises four main components: i) an interface to the HTN representation which contains representations for narrative events and situations (the latter in the form of sub-goals), as well as affinity between characters, ii) a rule-based engine that manages the template generation options iii) a template-based generator that produces semantic templates from the narrative generation and iv) a lexicalisation system that uses ad hoc lexical templates to produce the final linguistic utterance.

The dialogue system has been implemented and integrated as part of our IS environment. The utterances generated can be produced as text appearing on screen, or synthesised through a text-to-speech systems, in our implementation the Microsoft Speech API 5.1.

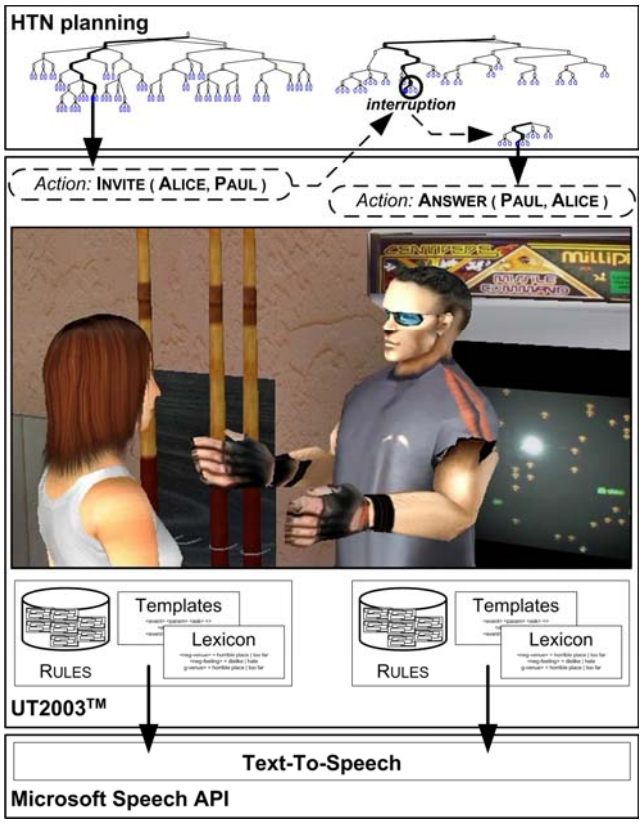

Figure 5: System architecture including Dialogue Generation.

Conclusions

We have proposed an approach integrating dialogue within an IS paradigm where staged cinematic action remains at the centre of the narrative. At this stage of prototype development, we have only defined linguistic resources for a small but representative set of situations, such as invitations, requests for resources (borrowing objects, etc.) or requests for assistance. While these account for a significant fraction of narrative events, they are far from being exhaustive: it is nevertheless the underlying principles, which unifies the various representations, narrative and linguistic, that we claim to be generic enough.

One of those unifying principles consists in using the relations between characters to relate the progression of the narrative to the generation of dialogue. The narrative actions are largely determined by the affinities between characters: hence, by providing compatible principles for the generation of dialogues (in particular, contextual rules governing the expression of acceptance or rejection, see Figures 2 and 4) we have identified a possible strategy for incorporating dialogue into character-based interactive storytelling.

The approach we proposed remains compatible with user intervention through similar linguistic principles (see Charles et al. 2004), although the paradigm for user involvement would probably have to be modified, embodying the user in one of the virtual characters.

Acknowledgements

We would like to thank Bob Berry and Vic DeLeon from Digitalo Design Inc. for the use of the character models from the game “Devastation”.

References

Bremond, C., 1970. Le rôle d'influenceur. In *Recherches Rhétoriques*, Communications 16, Paris: Seuil.

Bremond, C., 1973. *Logique du récit*, Paris: Seuil.

Cavazza, M., Charles, F., and Mead, S.J., 2002. Character-based Interactive Storytelling. In *IEEE Intelligent Systems*, special issue on AI in Interactive Entertainment, pp. 17-24.

Cavazza, M., 2003. An Empirical Study of Speech Recognition Errors in Human Computer Dialogue. In *Current and New Directions in Discourse and Dialogue*, Series: Text, Speech and Language Technology, Vol. 22 Van Kuppevelt, J.C.J; Smith, R.W. (Eds.), Kluwer.

Charles, F., and Cavazza, M., 2004. Exploring the Scalability of Character-based Storytelling. in *Proceedings of the 3rd International Conference on Autonomous Agents and Multi Agent Systems*, New York, USA.

Charles, F., Martin, O., Cavazza, M., Mead, S.J., Nandi, A., and Marichal, X., 2004. Compelling Experiences in Mixed Reality Interactive Storytelling. In *Proceedings of the 1st International Conference on Advances in Computer Entertainment Technology*, Singapore.

Joshi, A. K., and Schabes, Y., 1992. Tree-adjoining grammars and lexicalized grammars. In Maurice Nivat and Andreas Podelski, editors, *Tree Automata and Languages*, Elsevier Science.

Mateas, M., and Stern, A., 2002. A Behavior Language for Story-based Believable Agents. In *IEEE Intelligent Systems*, special issue on AI and Interactive Entertainment.

Mateas, M., and Stern, A, 2004. Natural Language Understanding in Façade: Surface-text Processing. In *Proceedings of the 2nd Technologies for Interactive Digital Storytelling and Entertainment* (TIDSE’04), Darmstadt, Germany.

Pustejovsky, J., 1995. *The Generative Lexicon*, MIT Press.

Riedl, M., and Young, R. M., 2004. An intent-driven planner for multi-agent story generation. In *Proceedings of the 3rd International Conference on Autonomous Agents and Multi Agent Systems*, New York, USA.

Traum, D. R., Rickel, J., Gratch, J., and Marsella, S., 2003. Negotiation over tasks in hybrid human-agent teams for simulation-based training. In *Proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems*, Sydney, Australia, pp. 441-448.

Walker, M., 1996. Inferring Acceptance and Rejection in Dialogue by Default Rules of Inference. In *Journal of Language and Speech*, 39-2.

Young, R. M., Riedl, M., Branly, M., Martin, R.J. and Saretto, C.J., 2004. An architecture for integrating plan-based behavior generation with interactive game environments. In *Journal of Game Development*, vol.1 issue 1, pages 51-70.