

A Computational Feature Analysis for Multilingual Character-to-Character Dialogue

Charles Callaway

North Carolina State University
Department of Computer Science
P.O. Box 7534 Raleigh, NC 27695
cbcallaw@eos.ncsu.edu

<http://multimedia.csc.ncsu.edu/imedia/index.html>

Abstract. Natural language generation systems to date have concentrated on the tasks of explanation generation, tutorial dialogue, automated software documentation, and similar technical tasks. A largely unexplored area is *narrative prose generation*, or the production of texts used in stories such as novels, mysteries, and fairy tales. We present a feature analysis of one complex area of NLG found in narrative prose but not in technical generation tasks: character-to-character dialogue. This analysis has enabled us to modify a surface realization system to include the necessary features that dialogue requires and thus to write the types of texts found in narratives.

1 Introduction

Recent years have witnessed an intense interest in the use of interactive language technologies in educational and entertainment applications. Their successful integration requires a number of support technologies that are currently under development in order to effectively engage learners and game players. Character-to-character dialogue in particular is one of the support technologies necessary for the realization of interactive multiagent education and entertainment. By creating computational models of these phenomena, it becomes possible to create intelligent systems for both literacy tutoring and producing interactive drama and novels. Furthermore, the capability of multilingual dialogue would allow those systems to reach a much wider audience.

Creating an interactive literacy tutoring system would be a key step in achieving software that could have a significant impact on the education of large numbers of children. First, such a system could function as a one-on-one reading tutor [15]. By allowing students to select key aspects of a story's content, those students could be immersed in a narrative world of their own origination. Such a world might contain their own characters, props, and action sequences. By allowing students to make such directorial choices, they will be more interested in the resulting outcome and more likely to avidly read what they have helped to create.

Second, students could engage in *exploratory narration* [11, 22]. For example, if a student had access to the raw materials for their favorite Jack London novella, they could request the same story content, but in a different style, such as Ernest Hemingway's chopppiness, or James Joyce's stream-of-consciousness style. This kind of capability encourages students to approach text in a different way, to understand the difference between content and presentation, and to want to create their own literary works with an appreciation for the world of potential design choices.

Third, students could engage in a new form of foreign language education. Students could elect to generate the narrative prose in any of a number of languages [1, 2], inviting them to make comparisons between different language versions of the same story. With conversational capability, this kind of interactive narrative system would enable students to have a one-on-one foreign language tutor.

Finally, this type of real-time interactive communication opens the way for intelligent tutors to actively explore 3D learning environments with students. Current such knowledge-centered learning environments [13, 21] interact with learners by sequencing prerecorded audio files. Once appropriate responses are determined, the corresponding audio files containing the pieces of spoken text are put into the proper order and sent to the audio device. The vast time and space requirements this method requires can be avoided by instead using a narrative prose generation system in combination with a high-quality voice synthesizer with prosody control [4, 6].

Although character-to-character dialogue is found in both spoken and written forms, the research described here discusses only the computational aspects of written dialogue, *i.e.*, the range of linguistic phenomena that comprise it and a computational feature analysis which will enable the creation of feature-based grammars for the generation of character dialogue. The phenomena of direct character-to-character dialogue is comprised of five major parts: the structure and function of the quoted material, the structure and function of the external elements, the utterer of the dialogue, the utterance matrix relation, and the orthography of dialogue punctuation. Although several of these elements have been discussed in the literature, no effort has yet been made to integrate them and thus create a comprehensive computational model of character dialogue.

The computational feature analysis of multilingual character-to-character dialogue described here is part of a long-term research program into educational and entertainment applications that make use of knowledge-based learning environments, natural language generation, and animated 3D agents in immersive 3D worlds. The model of dialogue described here has been implemented in a narrative prose generation system [3] which writes Little Red Riding Hood stories (cuentos de Caperucita Roja). By combining all of these technologies, we will be able to create 3D environments populated with engaging animated 3D characters capable of teaching literacy skills. This paper describes our model of character-to-character dialogue as part of the natural language generation module of such a literacy tutor.

2 Character-to-Character Dialogue

Written character-to-character dialogue is an aspect of narrative that takes the form of turn-taking conversational interactions between one or more agents. The written form lacks prosodic and accentual features found in spoken dialogue; instead, the substitution of orthographic punctuation marks serves as clues allowing the reader to reconstruct the spoken conversation.

Aside from the topic of narration, several researchers have worked on the topic of punctuation in (computational) linguistics, including Quirk *et al.*'s book [18] governing distribution of punctuation in discourse, Geoffrey Nunberg's book [16], Christine Doran's dissertation [8] that discuss the theory of punctuation in regard to particular grammatical theories, and Bernard Jones' dissertation [10] that uses distributional and frequency analysis to describe a functional model of punctuation incorporating both syntax and semantics. Other researchers such as Longacre [14] discuss conversational dialogue without reference to punctuation or its computational analysis.

However with regard to narration, there is very little extant research. Christine Doran [7] discusses reported speech in newspaper article corpora, provides a preliminary structural analysis, and describes how an LTAG grammar can be augmented for parsing applications. However, newspaper articles rarely contain true turn-taking conversational dialogue and are usually so formal that they do not contain as wide a range of variation as that found in fictional narratives (novels, mysteries, short stories, etc.)

In order for literacy education, conversational tutoring, or interactive fiction to become a reality, we must ensure that the conventions found in literature can be followed by natural language generation systems. Thus if particular linguistic phenomena are found in actual short stories and novels, they need to be computationally reproducible. The first stage in this process is to analyze in detail character-to-character dialogue that occurs naturally, then to model the phenomena (via the definition of features and rules that govern them in the case of functional systemic natural language generation grammars), and finally to write a grammar that uses the newly defined features in the actual production of narrative prose. In addition, if we wish our system to be *multilingual* so that it applies to as broad an audience as possible, we must repeat this process for each desired language.

Typical English Dialogue Patterns. Written English dialogue in direct form¹ has a small set of realizable forms, consisting of four main types: **Utterance-Only** where the speaker and the manner of speech must be inferred, **Utterance-Speaker** where the identity of the speaker follows the quoted material, **Speaker-Utterance** where the opposite is true, and **Utterance-Speaker-Utterance** where the utterances are either a single sentence split along an appropriate linguistic boundary

¹ Note that we present here only positive examples of direct speech, ignoring ungrammatical cases. For instance, the utterer sequence "said he" is considered to be archaic. These cases are discussed elsewhere in the literature.

or are separate sets of (possibly multiple) sentences. Each utterance may be a statement, an exclamation, or a question; each speaker includes a relation and an optional modifying clause describing how the utterance was delivered by the speaker. Additionally, utterances may be complete sentences, or split where one half of a coordinated phrase is in one utterance while the other half plus an optional sentential component are in the other half. A representative sample is shown below.

- “<sentence>.?!”
(1) “That scarf belongs to Patty.”
- “<nounphrase>.?!” John said <adverbial-manner>.
(2) “I couldn’t catch up to him,” John said breathlessly.
- “<sentence-np1>,” John said <time-manner>, “and <np2>.?!”
(3) “I usually drink coffee,” John said when he got up, “and orange juice.”
- John smiled, “<sentence>.?!” (nonverbal speaking relation)
(4) John smiled, “Were you always so ticklish?”
- “<sentence>.?! <sentence>.?! . . .”
(5) “I think it’s great. But you have to decide for yourself.”
- “<sentence>.?!” John said slowly, “<sentence>.?!”
(6) “Get out of here,” John said slowly, “I want to be alone.”
- “<sentence1>,” John said dryly, “but <sentence2>. <sentence3>.”
(7) “It’ll do,” John said dryly, “but I don’t like it. It smells bad.”
- “<nounphrase>.?!” “<prepphrase>.?!” . . . (In answer to a question.)
(8) “The table.” “On the shelf?” “Now!”
- “<np1>,” John said, “and <np2>.?!”
(9) “A hammer,” said John, “and nails.”
- “<interjection>.?!” John yelled, “<sentence>.?!”
(10) “No!” John yelled, “It’s mine!”
- “<circumstantial>.?!” John spoke up suddenly, “<sentence>.?!”
(11) “Before the sun goes down,” John spoke up suddenly, “Get some matches!”
- “<sentence>.?!” John exclaimed, “<preposition>.?!”
(12) “I want the garage door up,” John exclaimed, “up.”

Typical Spanish Dialogue Patterns. Written Spanish dialogue in direct form² is for the most part similar to English except in punctuation and style (determinable by distributional statistics) and that due to verb morphology, the

² Again, only grammatical examples are considered here. Spanish also has idiosyncrasies which I have not been able to find described in the literature. For example, the utterer sequence “Marcelo dijo” seems to be dispreferred.

identity of the speaker can be more easily inferred than in English by examining its ending. Spanish uses the long dash rather than quotation marks, and it only functions as a delimiter when a linguistic boundary must be given (*c.f.* examples (1) and (13)). A representative sample of Spanish character-to-character dialogue is shown below.

- - ¿i<sentence>!?
(13) -Anoche acabé mi nuevo invento. -¡Ya vamos, papá! -¿Ahora?
- - ¿i<sentence>!?-dijo Juan.
(14) -Está a punto de prueba-dijo, muy orgulloso.
(15) -¿Cuál es la sorpresa?-quiso saber Juan.
- -¿i<sentence>!?-dijo Juan <manner>.
(16) -Os dejasteis la puerta sin cerrar-dijo Juan, entrando en la tienda.
(17) -¿Qué querran los ladrones?-comentó Juan, mirando el desorden.
- -¿i<sentence>!?-dijo Juan-. ¿i<sentence>!?
(18)-¡Espere, espere!-pidió Juan, intentando tranquilizar a la asustada ancianita-.
¿No vive aquí ningún hombre?
(19) -Está bien-dijo Juan en voz alta-. ¡Veamos qué hacéis con ellos!
(20) -Ésta es la calle de la Nuez-dijo Juan, al volver una esquina-. El número dieciséis se halla en la acera de enfrente.
- -¿i<sentence-part>-dijo Juan-<sentence-remainder>!?
(21) -¿Por qué-preguntó Juan, cogiéndole el horario-has rodeado todas las clases de Diego con corazoncitos?
- -¿i<sentence-part>-dijo Juan-, <sentence-remainder>!?
(22) -La policía está trabajando ahí dentro-dijo la señora Martínez-, y nos han pedido que esperásemos fuera.
- Juan sonreió y dijo-: ¿i<sentence>!?
(23) Pero Juan le hizo una seña y cuchicheó-: No te pelees con ellos.

By comparing the requirements for multilingual direct character-to-character dialogue in a feature-based way, we can create a single natural language generation system capable of composing dialogue in a multitude of languages.

3 Features of Character-to-Character Narrative Dialogue

A computational feature analysis for some linguistic phenomenon begins with an analysis of its overall variability (or coverage). Once commonalities in the data can be observed, it becomes possible to assign a finite set of features and values which those features can take on. For example, one feature for nouns is *gender*, whose values can be *masculine*, *feminine*, or *neuter*. For character-to-character dialogue there are six relevant high-level features:

- The structure and function of the quoted material
- The utterer (speaker) of the dialogue
- The utterance matrix clause
- The structure and function of the surrounding paragraphs
- Whether it is direct or indirect
- The orthography of dialogue punctuation

We discuss each of these in turn.

3.1 The Quoted Material

Basic direct quotations can be of any syntactic structure: **S**, **NP**, **PP**, etc. (*c.f.* 1 and 8), especially when the quotation is a direct response to a previous question in the dialogue. However, there are several modifications to the basic sentential structure that authors may employ:

- *Interjections*: To express shock, surprise, and pain among a number of other feelings, characters will frequently utter interjections such as “Oh!”, “Aha!”, or “Ouch!”. These interjections are frequently fronted inside a complete sentence.
 - (24) “Ouch, you must have just come from the dentist!”
 - (25) “Oh my, perhaps you should sit down.”
- *Addressee*: In face to face dialogue, characters will frequently address each other by name, nickname, or some other appellation:
 - (26) “Carol, don’t forget to pick up some milk.”
 - (27) “My dear, all the better to see you with!”
- *Written stylistic effects*: In order to show emotions and have characters express themselves in nonstandard ways, authors write dialogue that includes onomatopoeiatic effects, simulated yelling, regional dialects, ellipsis, vocal handicaps, or vowel lengthening. Current surface realizers are ill-equipped to deal with these types of textual effects.
 - (28) “Ewwww, I can’t believe you just did that!”
 - (29) “Mom, you CAN’T do that!”
 - (30) “I’d really ’preciate it if you’d be gettin’ along now.”
 - (31) “But Molly . . .”
 - (32) “B-b-but, it’s so s-s-scary!”
 - (33) “Nooooooooo!”
- *Combinations*: Furthermore, these modifications to traditional sentential structures can be used in combination:
 - (34) “Wow, John, you REALLY hit that ball!”

Before a traditional grammar can be used to generate narratives that contain these types of utterances, its sentence-level syntactic category must be augmented to include interjections and addressees, and surface realizers that are accustomed to producing only precisely grammatical utterances must be modified to create the types of stylistic effects that are commonly found in existing narrative corpora.

Finally, these types of effects must be available in a multilingual environment. Natural language researchers must come to realize that some types of effects are available in some languages but not possible in others. In addition, these effects should be expressed semantically to allow surface realizers to produce their best syntactic interpretation for a specific language.

3.2 The Utterer and Matrix Clause

Although direct quotes can stand alone (and often do in fast-fire exchanges), more frequently they are accompanied by another unquoted clause which specifies the speaker and the manner in which the speaker conveyed the utterance. In these cases, the particular language (English, Spanish, etc.) strongly affects the way the utterer and matrix clause can be realized. There are six major features that influence these possibilities:

- *Communicative Act Ordering*: There are four positions (in English) that the utterer and the communicative relation can appear in comparison to the quoted dialogue: **utterance-only** as in “<sentence>”, **proposed** as in *John said*, “<sentence>”, **postposed** as in “<sentence>,” *John said*. and **interposed** as in “<sentence1>,” *John said*, “<sentence2>.”
- *Utterer Ordering*: There are two positions (in English) that the utterer and the communicative relation can appear in comparison to each other: **utterer-first** as in “<sentence>,” *John said*. and **utterer-second** as in “<sentence>,” *said John*. Also see [18, Section 14.29]. The former possibility is not available in Spanish unless it is followed by a nonverbal relation immediately coordinated with the matrix clause (*c.f.* example 23). Also, certain combinations are not allowed in modern English, *e.g.*, *Said John*, “*I don’t think so.*”
- *Utterance Manner*: The matrix relation in pure dialogue is often modified with a manner phrase indicating accompanying action or style in which the communicative act was delivered (besides connotative nuances of the communicative relation itself), such as **adverbial** in “<sentence>,” *John said hurriedly*, with a **prepositional phrase** as in “<sentence>,” *John said with aplomb*, with a **gerundive clause** as in “<sentence>,” *John said fidgeting with his tie*, or with a **co-event clause** as in “<sentence>,” *John said as he pressed the button*. Our informal corpora analyses indicate that these modifications occur more frequently in Spanish in texts that have been translated from English rather than written originally in Spanish.
- *Utterance relation semantics*: The utterance relation is not restricted to the traditional notion of purely communicative acts. For example, emotive verbs

are frequently used as in “*That’s my favorite,*” *John smiled.* Additionally, the utterance manner can be indicated by the connotation of the utterance relation, e.g., *John whispered,* “*Do you think so?*” It is not clear from the available evidence however that the relation semantics should be a top level feature rather than some combination of low-level features; more investigation into similar phenomena in other languages is required.

- *Use of Pronouns:* English requires a speaker in direct speech if a communicative act is used; because Spanish verbs include person and number, a distinct speaker or pronoun is unnecessary. In addition, if a pronoun is used after a communicative act in English in direct speech, it is considered archaic; a postposed pronoun in indirect speech is prohibited. These restrictions must be considered by a surface realizer when creating text.
- *Segmentation:* Most notably in replies to questions [14], quoted dialogue does not need to be a complete sentence. There are a large number of utterances that are fragmentary phrases due to the informal nature of dialogue [18, Chapter 11]. For example, “*On the chair,*” *said John,* “*and near the door.*”

3.3 The Surrounding Paragraphs

Necessarily, turn-taking dialogue interacts with expository text in a narrative and with dialogue initiated by a different character in the story. In addition, narrative passages are told from a particular point of view, either that of a particular character or some narrator who is not a participating character. Whenever an author switches between embodied characters or narrators in a turn-taking dialogue, there is a required paragraph break, which functions as a whitespace delimiter. Consider the following example from a recent Harry Potter book [20]:

The Dursleys often spoke about Harry like this, as though he wasn’t there — or rather, as though he was something very nasty that couldn’t understand them, like a slug.

“What about what’s-her-name, your friend — Yvonne?”

“On vacation in Majorca,” snapped Aunt Petunia.

“You could just leave me here,” Harry put in hopefully.

Aunt Petunia looked as though she’d just swallowed a lemon.

On the other hand, when an author switches between dialogue and expository text where the same character is both the speaker and main actor, the author may choose either to delimit via a paragraph break or not. In addition, a particular author or publisher will almost always follow a consistent editing scheme in a particular publication. The following example demonstrates the choice of delimiting between expository text and dialogue when the actor and speaker are the same:

“What’s this?” he asked Aunt Petunia. Her lips tightened as they always did if he dared ask a question.

“Your new school uniform,” she said.

Harry looked in the bowl again.

“Oh,” he said, “I didn’t realize it had to be so wet.”

3.4 Direct vs. Indirect Speech

An author has a wide range of choices concerning the choice between characters' use of direct and indirect speech³. The text can range from pure exposition (with no communicative acts) to exposition with indirect communication to fully interleaved exposition and direct dialogue to pure dialogue. Longacre [14] cites examples along this range, some of which are adapted slightly here:

- *Pure Dialogue:*
 “It’s a beautiful night, Mary,” John said.
 “Yes, John, and how softly the moon is lying on the water,” replied Mary.
 “I love you, Mary,” John said.
 “At your age, John?” Mary asked.
- *Interleaved Exposition and Dialogue:*
 It was a beautiful night. The moon lay softly on the water. Then John said to Mary, “I love you.” But Mary said, “At your age, John?”
- *Exposition and Indirect Dialogue:*
 It was a beautiful night. The moon lay softly on the water. Then John told Mary that he loved her. She said she couldn’t believe he would ask her such a thing.
- *Exposition with No Communicative Acts:*
 It was a beautiful night. The moon lay softly on the water. Then John talked with Mary about his love for her. Mary refused to believe him.

In addition, the entire narrative can contain a mixture of these four types of character interactions as the author sees fit, taking into account various stylistic factors.

3.5 Orthography

The actual rules used to insert punctuation into direct character-to-character dialogue vary highly by language, as shown in the previous enumeration of English and Spanish examples. For English, Christine Doran’s work [7, 8] on the incorporation of punctuation into lexicalized tree-adjoining grammar describes how this can be accomplished at the deep structure level. However, her work does not investigate the complete range of coverage found in narrative texts. The features we found necessary are:

- Quotation marks: The actual orthographical punctuation marks expressed in the text, and whether they are used as delimiters or not. These are typically different for different languages.
- Capitalization inside quotations: In English, when a question or exclamation occurs as one part of the direct quotation, it is treated as one entire sentence and is capitalized accordingly; in Spanish, the question or exclamation is marked and not capitalized:
 “Well, boy, do you know what I’m telling you?”
 –Bueno, muchacho, ¿sabes qué te digo?

³ See [8] for more details on indirect speech.

- Modifications to terminal punctuation: There are important differences between punctuation in quoted dialogue and punctuation within expository text. First, normally terminating punctuation such as question marks and exclamation points can become nonterminal if they do not end the utterance, as in “*Hey!*” *John yelled*. However in English, periods always remain terminal and must be replaced with a comma if they do not terminate the utterance, regardless of whether the termination is for the communicative relation or the quoted text. Also see [8].
- Multiple sentences in one quotation: Multiple sentences within the same dialogue turn are combined into one large paragraph where adjacent quotation marks or other dialogue delimiters are removed.

4 NLG Aspects

The rules of character-to-character dialogue described above must be integrated into a natural language generation system in order to allow it to create narrative texts. Previous NLG systems have focused on explanation generation [12], tutorial systems [19], and automatic software documentation [17] among other domains. Deep-level NLG systems typically have a pipelined architecture where semantic input specifications are processed successively until text is finally produced (Figure 1).

The process begins when the system receives a high-level semantic specification to communicate information. In the case of writing narratives with characters and a plot, the source is usually a high-level narrative planner. Once a narrative plan is received, a *sentence planner* maps out which conceptual elements should be sentential subjects, sentential objects, verbal relations, locations, etc. The concepts are then assigned those roles, instantiated into a deep structure, and lexicalized by retrieving the appropriate items from the lexicon.

Because thematic roles map well across multiple languages, it is relatively easy to perform a “direct” translation by replacing lexical items at all levels of the deep structure produced by the sentence planner. The main key is to ensure that the interface between the sentence planner and the surface realizer is constant across all supported languages and to ensure that the surface realizer accepts semantic rather than syntactic input. However, it is not necessarily the case that the direct translation is the best translation.

Typically every element of the pipeline from the sentence planner on down concentrates solely on individual sentences. Thus, the features described in the previous section should be divided into two categories: those that are extrasentential, which require some knowledge on the part of the narrative planner to model, and those that are intrasentential, which can be handled solely by the lower levels of the NLG system. However, since many surface features of the previously described phenomena concern punctuation that must be integrated with the surface realization phase, it is usually easier for the narrative planner to supplement the sentential planner with enough additional semantic information to make informed choices about extrasentential phenomena.

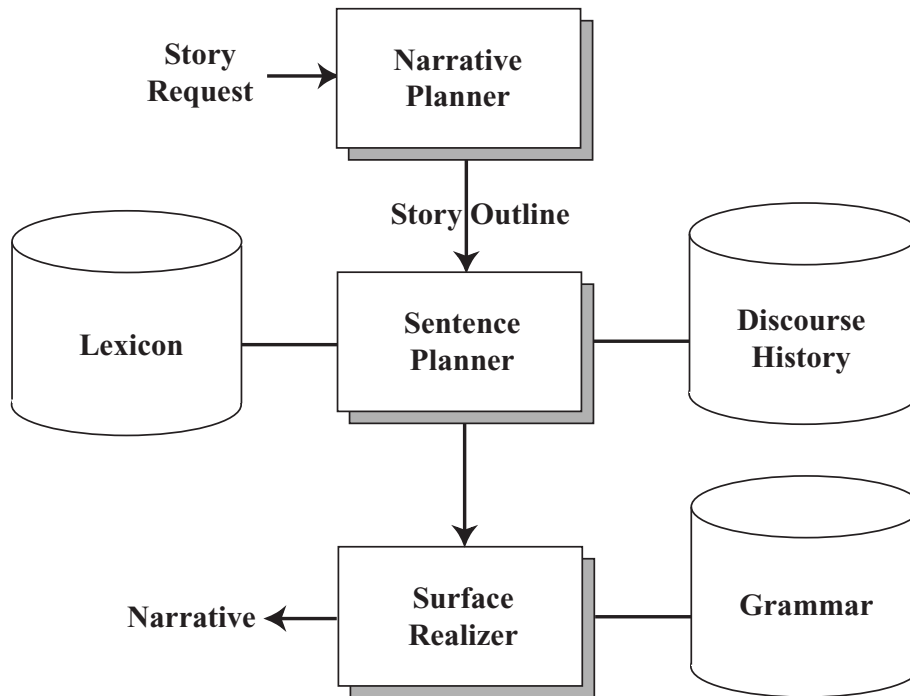


Fig. 1. A Generic Deep Generation Architecture

5 Implementation

The linguistic phenomena of narrative character-to-character dialogue and the model described above have been implemented in the full-scale narrative prose generator STORYBOOK, described in [3]. STORYBOOK takes a linearized narrative plan produced by a simple LISP based narrative planner and, using knowledge of character-to-character dialogue along with more traditional expository natural language generation techniques, writes publishable-quality narratives complete with formatting. The “rules” that the narrative planner uses to create the plot are currently hand-coded in a standard knowledge representation language. STORYBOOK uses the publically available FUF/SURGE [9] surface realization system with modifications to its systemic-functional grammar to implement the low-level dialogue features described in the previous section:

- Paragraph formatting
- Interjections
- Addressee
- Communicative Act Ordering
- Utterer Ordering
- Utterance Manner

In addition to these features, our implementation of the additions to the FUF/SURGE English grammar allows for the proper low level punctuation that readers expect from narrative prose. As an example, consider the following excerpt from the grammar for dealing with punctuation in dialogue consisting of two of the five necessary alternative branches. The first branch governs the case when only a direct utterance is given (marked as `utt1`) but no speaker or utterance relation is given. In all such cases, the grammar capitalizes the first word, adds a quotation mark before the sentence, and the terminal punctuation and a second quotation mark to the end of the sentence.

The second branch governs the case when the direct quotation consists of a single utterance and the speaker and utterance relation are expressed with the quote preceding the speaker and utterance relation. In this case, as the grammar shows, if the terminal punctuation for the quote is the period, it is replaced with a comma. Additionally, the grammar expresses the constraint that the linearization module must force the quote to appear before the remainder of the sentence. Other branches of the main grammar add the terminal punctuation to the utterance relation. The additional three branches of the quotation module of the grammar deal with cases where the quotation follows the utterance relation, the utterance relation is surrounded by two quoted elements, or the default case where no quotation is expressed. As a final note, the Spanish grammar version contains similar branches for dealing with quotations but with rules, constraints, and lexicalizations particular to the Spanish language.

```
(alt ordering
  (((sayer none) (utt2 none) (utt1 given) (pattern (utt1))
    (utt1 ((alt (((cat clause) (mood \#(under interrogative))
      (mood given) (question-mode direct)
      (punctuation ((before ") (capitalize yes)
        (after ?")))))
      ((punctuation ((before ") (capitalize yes)
        (after ."))))
      ((punctuation ((before ") (capitalize yes)
        (after !")))))))))))
  ((sayer given) (utt1 given) (utt2 none)
    (sayer ((order second))) (pattern (utt1 sayer))
    (utt1 ((alt (((cat clause) (mood #(under interrogative))
      (mood given) (question-mode direct)
      (punctuation ((before ") (after ?")
        (capitalize yes))))
      ((punctuation ((before ") (after ,")
        (capitalize yes))))
      ((punctuation ((before ") (after !")
        (capitalize yes))))))))))
  ...
```

The modifications to the FUF/SURGE grammar described here are used by the STORYBOOK implementation to write 2 to 3 page fairy tales in the Little Red

Riding Hood domain [3]. A sample of STORYBOOK's narrative prose is shown below.

The road to grandmother's house led through the dark forest. Little Red Riding Hood was not afraid, instead she was quite happy. Birds sang her their sweetest songs while squirrels ran up and down tall trees. Now and then, a rabbit would cross her path. She had not gone far when she met a wolf. "Hello," greeted the wolf.

The wolf was a cunning looking creature. He asked, "Where are you going?"

"I am going to my grandmother's house," replied Little Red Riding Hood.

"Well then, take care. There are many dangers."

6 Conclusion

Natural language generation systems to date have concentrated on the more technical aspects of text: explanation generation, tutorial dialogue, and automated software documentation, among other areas. One relatively unexplored area is *narrative prose generation*, or the production of texts used in stories such as novels, mysteries, and fairy tales. We have presented a feature analysis of character-to-character dialogue, which is extensively used in narrative texts. This analysis has enabled us to modify a widely-used surface realization system to include the necessary features that dialogue requires. This surface realizer was used in an implemented system that wrote fairy tales in the domain of Little Red Riding Hood. The resulting narrative prose displayed the correct punctuation and formatting expected of human-produced narrative prose.

References

1. Bateman, J., Sharoff, S.: Multilingual grammars and multilingual lexicons for multilingual text generation. In Proceedings of the ECAI Workshop on Multilinguality in the Lexicon-II. Brighton, UK (1998) 1–8
2. Callaway, C., Daniel, B., Lester, J.: Multilingual Natural Language Generation for 3D Learning Environments. In Proceedings of the 1999 Argentine Symposium on Artificial Intelligence. Buenos Aires, Argentina (1999) 177–190
3. Callaway, Charles: Narrative Prose Generation. Ph.D. Dissertation. North Carolina State University. Raleigh, NC (2000)
4. Cassell, J., Stone, M., Yan, H.: Coordination and Context-Dependence in the Generation of Embodied Conversation. International Natural Language Generation Conference. Mitzpe Ramon, Israel (2000)

5. Dale, Robert: Exploring the Role of Punctuation in the Signalling of Discourse Structure. In Proceedings of the Workshop on Text Representation and Domain Modelling. T. U. Berlin (1991) 110–120
6. Daniel, B., Bares, W., Callaway, C., Lester, J.: Student-Sensitive Multimodal Explanation Generation for 3D Learning Environments. In Proceedings of the Sixteenth International Conference on Artificial Intelligence. Orlando, FL (1999) 114–120
7. Doran, Christine: Punctuation in Quoted Speech. In Notes of the ACL Workshop on Punctuation in Computational Linguistics. Santa Cruz, CA (1996) 9–18
8. Doran, Christine: Incorporating Punctuation into the Sentence Grammar: A Lexicalized Tree Adjoining Grammar Perspective. Ph.D. Dissertation. University of Pennsylvania. Philadelphia, PA (1998)
9. Elhadad, Michael: Using Argumentation to Control Lexical Choice: A Functional Unification Implementation. Ph.D. Dissertation. Columbia University. New York, NY (1992)
10. Jones, Bernard: What's the Point? A (Computational) Theory of Punctuation. Ph.D. Dissertation. The University of Edinburgh. Edinburgh, Scotland (1996)
11. Laurel, Brenda: Toward the Design of a Computer-Based Interactive Fantasy System. Ph.D. Dissertation. The Ohio State University. Columbus, OH (1986)
12. Lester, James C., Porter, Bruce W.: Developing and Empirically Evaluating Robust Explanation Generators: The KNIGHT Experiments. *Computational Linguistics* **23** 1:65–101 (1997)
13. Lester, James C., Stone, Brian A.: Increasing Believability in Animated Pedagogical Agents. In Proceedings of the First International Conference on Autonomous Agents. Marina del Rey, CA (1997) 16–21
14. Longacre, Robert E.: *The Grammar of Discourse*. Plenum Press, New York, NY (1983)
15. Mostow, J., Aist, G.: The Sounds of Silence: Towards Automated Evaluation of Student Learning in a Reading Tutor that Listens. In Proceedings of the Fourteenth National Conference on Artificial Intelligence. Providence, RI (1997) 355–361
16. Nunberg, Geoffrey: *The Linguistics of Punctuation*. Center for the Study of Language and Information. Stanford, CA (1990)
17. Paris, Cécile L., Vander Linden, Keith: Building Knowledge Bases for the Generation of Software Documentation. In Proceedings of the 16th International Conference on Computational Linguistics. Copenhagen, Denmark (1996) 734–739
18. Quirk, R., Greenbaum, S., Leech, G., Svartvik, J.: *A Comprehensive Grammar of the English Language*. Longman Publishers, London (1985)
19. Rosenblum, James A., Moore, Johanna D.: Participating in Instructional Dialogues: Finding and Exploiting Relevant Prior Explanations. In Proceedings of the World Conference on AI and Education. August (1993)
20. Rowling, J. K.: *Harry Potter and the Sorcerer's Stone*. Scholastic, Inc. (1998)
21. Towns, S., Callaway, C., Voerman, J., Lester, C.: Coherent Gestures, Locomotion, and Speech in Life-Like Pedagogical Agents. Proceedings of the Fourth International Conference on Intelligent User Interfaces. San Francisco, CA (1998) 13–20
22. Walker, Marilyn A., Cahn, Janet E., Whittaker, Stephen J.: Improvising Linguistic Style: Social and Affective Bases of Agent Personality. Proceedings of the First International Conference on Autonomous Agents. Marina del Rey, CA (1997)