An Annotated Corpus of Film Dialogue for Learning and Characterizing Character Style

Marilyn A. Walker, Grace I. Lin, Jennifer E. Sawyer

University of California Santa Cruz
Natural Language and Dialogue Systems Lab, Computer Science Dept.
maw|glin|jsawyer@soe.ucsc.edu

Abstract

Interactive story systems often involve dialogue with virtual dramatic characters. However, to date most character dialogue is written by hand. One way to ease the authoring process is to (semi-)automatically generate dialogue based on film characters. We extract features from dialogue of film characters in leading roles. Then we use these character-based features to drive our language generator to produce interesting utterances. This paper describes a corpus of film dialogue that we have collected from the IMSDb archive and annotated for linguistic structures and character archetypes. We extract different sets of features using external sources such as LIWC and SentiWordNet as well as using our own written scripts. The automation of feature extraction also eases the process of acquiring additional film scripts. We briefly show how film characters can be represented by models learned from the corpus, how the models can be distinguished based on different categories such as gender and film genre, and how they can be applied to a language generator to generate utterances that can be perceived as being similar to the intended character model.

Keywords: dialogue, expressive, natural language generation, film, archetype theory

1. Introduction

Conversation is an essential component of social behavior, one of the primary means by which humans express emotions, moods, attitudes and personality. Thus a key technical capability for interactive narrative systems (INS) is the ability to support natural conversational interaction. To do so, natural language processing can be used to process the user's input to allow users flexibility in what they say to the system (Johnson et al., 2005; Mateas and Stern, 2003; Louchart et al., 2005). However, in most interactive narrative systems to date, character dialogue is highly handcrafted. Although this approach offers total authorial control and produces high quality utterances, it suffers from problems of portability and scalability (Walker and Rambow, 2002), or what has been called the *authoring bottleneck* (Mateas, 2007). Moreover, handcrafting makes it difficult, if not impossible, to personalize the dialogue interaction, but personalization leads to perceptions of greater player agency (Murray, 1997; Hayes-Roth and Brownston, 1994; Mott and Lester, 2006; Thue et al., 2010).

Expressive Natural Language Generation (ENLG) promises a solution to these problems, but the ENLG engine must be able to produce variations in linguistic style that clearly manifest differences in dramatic character. Therefore the first requirement for building an ENLG for dialogue for dramatic characters, is a method or a theory that systematically and comprehensively quantifies the most important individual and stylistic differences in behavior, the way they affect linguistic output in dialogue, and the predicted effect on the perceptions of the listener. Previous work on ENLG has explored parameters and models based on Brown and Levinson's theory of politeness, the Big Five theory of personality, and dramatic theories of archetypes, (Piwek, 2003; André et al., 2000; Mairesse and Walker, 2010; Gupta et al., 2007; Walker et al., 1997; Wang et al., 2005; Rowe et al., 2008; Cavazza and Charles, 2005) *inter alia*. Here we describe a new annotated corpus of film dialogue and how we have used it to learn character models of linguistic style that incorporate some concepts from the dramatic theory of archetypes. We believe that the stylized, crafted aspects of film dialogue are actually useful for our purposes because it is authored deliberately in order to convey the feelings, thoughts and perceptions of the character being portrayed. Furthermore, the screenplay often specifies the emotion of an utterance with psychological state descriptors. In addition, the dialogue is constructed to reveal or focus the viewer's attention on the character's personality and the key plot events involving a character and their perceptions, especially in dramatic films as opposed to action.

In the first section we will describe the content of the film corpus and the methods we used to create it. Next we will discuss some of our recent applications of the corpus. We have used the corpus to train character models for generating expressive dialogue with the PERSONAGE generator and performed a perceptual study indicating that subjects were able to discern similar personality traits between original film dialogue utterances and generated dialogue for another domain. Finally, we discuss possible ways in which we could augment the film corpus for future work.

2. Corpus Description

Our corpus consists of 862 film scripts from The Internet Movie Script Database (IMSDb) website (http://www.imsdb.com/), representing 7,400 characters, with a total of 664,000 lines of dialogue and 9,599,000 tokens. Our snapshot of IMSDb is from May 19, 2010. Figure 1 provide example dialogues in the corpus.

We used The Internet Movie Database (IMDB) ontology to define groupings of character types according to the following attributes: GENRE, DIRECTOR, YEAR, and CHARACTER GENDER. See Table 1. Previous work suggests that females and males in each genre might have different lin-

ANNIE HALL SCENE: Lobby of Sports Club	INDIANA JONES SCENE: Marion's Bar on Fire
ALVY: Uh ... you-you wanna lift? **ANNIE:** *Turning and aiming her thumb over her shoulder* Oh, why-uh ... y-y-you gotta car? **ALVY:** No, um ... I was gonna take a cab. **ANNIE:** *Laughing* Oh, no, I have a car. **ALVY:** You have a car? *Annie smiles, hands folded in front of her* **ALVY:** So ... *Clears his throat.* **ALVY:** I don't understand why ... if you have a car, so then-then wh-why did you say "Do you have a car?"... like you wanted a lift?	**INDY:** Let's get out of here! **MARION:** Not without that piece you want! **INDY:** It's here? *Marion nods, kicks aside a burning chair. Another burning beam falls from the roof. Indy pulls Marion close to him protectively.* **INDY:** Forget it! I want you out of here. Now! *He begins dragging her out.* **MARION:** *pointing.* There! *She breaks away from him, darts back and picks the hot medallion up in the loose cloth of her blouse.* **INDY:** Let's go! **MARION:** (looking around) You burned down my place! **INDY:** I owe you plenty! **MARION:** You owe me plenty! **INDY:** *smiles* You're something! **MARION:** I am something. And I'll tell you exactly what - *She holds up the medallion possessively.* I'm your partner!

Figure 1: Scenes from *Annie Hall* and *Indiana Jones and the Raiders of the Lost Ark*.

Group	Categories
Genre	drama, thriller, crime, comedy, action, romance, adventure
Gender	male, female
Film Year	year>2000, 1995>year<=2000, 1990>year<=1995, 1985>year<=1990, 1980>year<=1985, older
Film Director	Michael Mann, WesCraven, Steven Spielberg, Stanley Kubrick, Ridley Scott, Frank Capra, Steven Soderbergh, David Fincher, Alfred Hitchcock, Robert Zemeckis, David Lynch, James Cameron, Joel Coen, Martin Scorsese, Quentin Tarantino

Table 1: Film Categories

guistic styles (Ireland and Pennebaker, 2011), so we used the Names Corpus, Version 1.3 (see website of Kantrowitz and Ross 1994) to label common gender names and hand-annotated the remaining characters. Note also that most films belong to multiple genres. For example, *Pulp Fiction* belongs to crime, drama, and thriller. This allows for characters to be grouped in multiple categories.

Each script was parsed to extract dialogic utterances, producing output files for each individual character from the film that containing only their lines. For example, *pulp-fiction-vincent.txt* contains all of the lines for the character, Vincent, from *Pulp Fiction*.

Next we annotated the corpus with various linguistic reflexes. A summary of these are given in Table 2. In some cases, we use tools that have been used previously for personality or author recognition or as useful as indicators of a person's personality, gender or social class (Mairesse et al., 2007; Furnham, 1990; Pennebaker and King, 1999; Ireland and Pennebaker, 2011). We have also written new linguistic inference methods and trained a dialogue act tagger for the corpus.

Basic: We assume that how much a character talks and how many words they use is a primitive aspect of character. Therefore, we count number of tokens and turns. These, especially when considered in tandem with other features may indicate traits such as introversion, overall verbosity, and linguistic sophistication.

Polarity: Positive and negative polarity are determined using SentiWordNet 3.0 (`http://sentiwordnet.isti.cnr.it/`). It assigns to each synset of WordNet three sentiment scores: positivity, negativity, and objectivity. After using Stanford's POS Tagger, we convert Penn tags to WordNet tags. Then we approximate the sentiment value of a word with a label (no word sense disambiguation) using weights. For example, if there are three values $(v1, v2, v3)$, where $v1$ is associated with the most common sentiment value, associated with a particular word, then the score is calculated as $\frac{(1)*v1+(1/2)*v2+(1/3)*v3}{(1)+(1/2)+(1/3)}$. For more than one word (in a sentence or entire dialogue), simply average the scores. The polarity is assigned based on the range defined in Table 3.

Feature	Description
Basic	Number of sentences, sentences per turn, number of verbs, number of verbs per sentence, etc.
Polarity	Overall polarity, polarity of sentences, etc.
Dialogue Act	Trained with *NPS Chat Corpus* with 15 dialogue act types such as "Accept", "Clarify", "Emotion", and "ynQuestion".
First Dialogue Act	Look at the dialogue act of the first sentence of each turn.
Merge Ratio	Use regular expression to detect the merging of subject and verb of two propositions.
Passive Sentence Ratio	Using a third party software (see text) to detect passive sentences.
Concession polarity	Polarity for concessions
LIWC Word Categories	Word categories from the Linguistic Inquiry and Word Count (LIWC) text analysis software.
Pragmatic Markers	Word categories and examples: taboo (fuck, shit, hell, damn), sequence (first, second, third), opinion (think, feel), aggregation (with, also, because), soft (somewhat, quite, around), emphasis (really, basically, actually), acknowledge (yea, right, ok), pauses (i mean, you know), concession (but, yet, although, even though, on the other hand), concede (although, but, though, even if), justify (because, since, so), contrast (while, but, however, on the other hand), conjunction (for, and, nor, but, or, yet, so), etc.
Tag Question Ratio	Amount of tag questions
Word Length	Average content word length
Verb Strength	Averaged sentiment values of verbs

Table 2: Automatically Annotated Linguistic Features

Polarity assigned	Range of score (s)
String Positive	$s \geq 2/3$
Positive	$1/3 < s < 2/3$
Weak Positive	$0 < s < 1/3$
Neutral	$s == 0$
Weak Negative	$-1/3 \leq s < 0$
Negative	$-2/3 \leq s < -1/3$
Strong Negative	$s \leq -2/3$

Table 3: Polarity score with *SentiWordNet*

Dialogue Act: Different types of characters use different dialogue acts to take the initiative or in response. Dialogue act type is detected with a dialogue act tagger trained on the NPS Chat Corpus 1.0 (Forsyth and Martell, 2007).
First Dialogue Act: The Dialogue Act of the first sentence of each turn.
Merge Ratio. To detect merging of sentences (merge of subject and verb of two propositions), we use a grammar that looks for verb+noun+conjunction+noun.
Passive Sentence Ratio. Passive sentences are detected using scripts from `http://code.google.com/p/narorumo`, under `source/browse/trunk/passive`. These scripts implement the rule that if a to-be verb is followed by a non-gerund, the sentence is probably in passive voice.
Concession Polarity. Find the polarity for concession part of the sentence, if exists, using the Polarity feature set.
LIWC Word Categories: The Linguistic Inquiry Word Count (LIWC) tool provides a lexical hierarchy that tells us how frequently characters use different types of words such as words associated with anger or happiness, as well as more subtle linguistic cues like the frequent use of certain pronouns. Examples of the LIWC word categories are given in Table 4. These features may correspond to particular themes that a character pursues in their discussions, or whether the character fits within a particular archetypal style. For example, one prediction would be that the archetype SHADOW would use more negative emotion and more anger words.

LIWC Category	Sample words
Anger words	hate, kill, pissed
Metaphysical issues	God, heaven, coffin
Physical state/function	ache, breast, sleep
Inclusive words	with, and, include
Social processes	talk, us, friend
Family members	mom, brother, cousin
Past tense verbs	walked, were, had
References to friends	pal, buddy, coworker

Table 4: Examples of LIWC word categories and sample words

Pragmatic Markers: Since pragmatic markers are particularly important part of linguistic style, we develop features to count them (Brown and Levinson, 1987). These include both categories of pragmatic markers and individual word count/ratio.
Tag Question Ratio. Tag questions are detected by using regular expressions to parse sentences.
Average Content Word Length. Use WordNet's tag to find content words (noun, adjective, adverb, and verb), then average the length of words (number of letters).
Verb Strength. Average sentiment scores of all verbs.
We have also carried out an annotation study on a number of characters and scenes in our IMSDb (Internet Movie Script Database) corpus. The idea was to first classify film characters into particular archetypes, and then derive corpus-based models from the archetypes. We asked three annotators to classify 17 film characters into one of the 13 archetypes described in (Faber and Mayer, 2009). The list of film characters and archetypes are in Table 5.
One advantage of this approach is that it lets us indirectly incorporate observations about types of characters from Archetype Theory. (Faber and Mayer, 2009). Archetype Theory provides a number of stock characters, such as HERO, SHADOW, or CAREGIVER, who have typical roles and personalities that can be re-used in different types of narrative (Rowe et al., 2008).

Film Characters (17): Bruce: *Batman Returns*, Rae: *Black Snake Moan*, Neil: *Dead Poets Society*, Costello: *The Departed*, Tyler: *Fight Club*, Carter: *Final Destination*, Hooper: *Jaws*, Scott Smith: *Milk*, Furious: *Mystery Men*, Pete: *O Brother, Where Art Thou?*, Morris: *Purple Rain*, Paul: *Rachel Getting Married*, Plato: *Rebel without a cause*, Agnis: *The Shipping News*, Rose: *Titanic*, Goose: *Top Gun*, Spud: *Transpotting*
Archetypes (13): Caregiver, Creator, Everyman/Everywoman, Explorer, Hero, Innocent, Jester, Lover, Magician, Outlaw, Ruler, Sage, Shadow

Table 5: Annotation Task Film Characters and Archetypes

3. Application of the Film Corpus: Learning Character Models

We utilize the film corpus in our work (Lin and Walker, 2011) and (Walker et al., 2011) to develop statistical models of character linguistic style and use these models to control the parameters of the PERSONAGE generator (Mairesse and Walker, 2011; Mairesse and Walker, 2010). We find that the models learned from film dialogue are generally perceived as being similar to the character that the model is based on. Our experimental method can be summarized as follows:

1. Collect movie scripts from The Internet Movie Script Database (IMSDb).
2. Parse each movie script to extract dialogic utterances, producing an output file containing utterances of exactly one character of each movie (e.g., *pulp-fiction-vincent.txt* has all of the lines of the character Vincent).
3. Select characters from those with more than 60 turns of dialogue.
4. Extract features representing the linguistic behaviors of each character.
5. Learn models of character linguistic styles based on these features.
6. Use character models to control parameters of the PERSONAGE generator.
7. Evaluate human perceptions of dialogic utterances generated using the character models.

The extracted features can be used to train models which represent individual film characters or groups of characters. To represent a group of characters, we can use machine learning techniques to distinguish groups such genre, gender, directors, and film period. Selected top results for discriminating distinct classes of two-class GENRE X GENDER, five-class DIRECTOR, five-class GENDER X DIRECTOR, and five-class GENDER X FILM PERIOD, are shown in Table 6. The results show that we can discriminate two-class GENRE X GENDER categories of characters using binary classification models with accuracies over 70% as opposed to baselines of around 50%.

Group: Categories	Selected	Test Case	Size	Baseline	Accuracy
Genre: drama, thriller, crime, comedy, action, romance, adventure	**Genre, Gender**	Drama Female vs. Adventure Male	813	50.43%	74.05%
		Family Male vs. Biography Male	181	49.72%	74.03%
		Western Male vs. Animation Male	78	48.72%	71.79%
Directors: Mann, Craven, Spielberg, Kubrick, Scott, Capra, Soderbergh, Fincher, Hitchcock, Zemeckis, Lynch, Cameron, Coen, Scorsese, Tarantino	**Five Directors**	Mann vs. Hitchcock vs. Lynch vs. Cameron vs. Tarantino	108	18.35%	64.22%
		Mann vs. Lynch vs. Hitchcock vs. Kubrick vs. Zemeckis	103	19.42%	53.40%
	Gender, Director	Male: Mann, Capra, Fincher, Cameron, Tarantino	87	22.99%	66.67%
		Female: Scott, Capra, Fincher, Cameron, Coen	34	29.40%	50.00%
Film Period: now–2005, 2005–2000, 2000–1995, 1995-1990, 1990–1985, 1985–1980, before 1980	**Gender, Years**	Male: now–2005, 2005–2000, 2000–1995, 1995–1990, before 1980	4041	20.29%	83.37%
		Female: now–2005, 2005–2000, 2000–1995, 1995–1990, before 1980	1134	20.28%	76.37%

Table 6: Top Classification Results for Character Styles Learned Using J48 Decision Trees

The five-way discriminatory models for combinations of directors, gender and years are much more complex, and the accuracies are amazingly high, given baselines around 20%. We can easily develop distinct character models for different directors and gender/director combinations. Also, interestingly, the results show that the year of the film has a large impact on style, and that combinations of gender and time period can be discriminated with accuracies as high as 83%.

To represent individual characters, we derive distinctive features for that character by normalizing these feature counts against a representative population. For each feature x_i, the normalized value z-score, z_i, is calculated as:

$$z_i = \frac{x_i - \overline{x_i}}{\sigma_{xi}} \qquad (1)$$

There are many choices for the population of characters used for normalization. For example, for a female character, we could use all female characters or all female action characters. For our work we chose the gender population of character. Any z-score greater than 1 or less than -1 is more than one standard deviation away from the mean. We consider all features with z-score > 1 or < -1 as being significant, and these features are mapped to one or more PERSONAGE generation parameters.

Sample character models derived from the procedure above are provided in Table 7. Table 8 illustrates the result of applying these models of character to a different story domain, *SpyFeet* (Reed et al., 2011), and shows some of the variation that we are currently able to produce.

We wanted to test the character models and mappings as described above. The simplest way to do this is to ask human participants to rate a set of utterances produced using different models in terms of their similarity of linguistic style to the mimicked character. This is carried out in (Walker et al., 2011). We use six film characters for this study: Alvy and Annie from *Annie Hall*, Indy and Marion from *Indiana Jones - Raiders of the Lost Ark*, and Mia and Vincent from *Pulp Fiction*.

For each film character model, we generate a page showing the user (1) selected original film scenes with dialogue for each character; and (2) **all** of the generated utterances using all of the film character models. Then we ask users to judge on a scale of 1. . .7 how **similar** the generated utterance is to the style of the film character as illustrated in the three scenes. Users are instructed to use the whole scale, and thus effectively **rank** the generated utterances for similarity to the film character.

Parameter	Description	Annie
Content Planning		
Verbosity	Control num of propositions in the utterance	0.78
Content Polarity	Control polarity of propositions expressed	0.77
Polarization	Control expressed pol. as neutral or extreme	0.72
Repetition Polarity	Control polarity of the restated propositions	0.79
Concessions	Emphasize one attribute over another	0.83
Concessions Polarity	Determine whether positive or negative attributes are emphasized	0.26
Positive Content First	Determine whether positive propositions - including the claim - are uttered first	1.00
Syntactic Template Selection		
First Person in Claim	Control the number of first person pronouns	0.6
Claim Polarity	Control the connotation of the claim	0.57
Claim Complexity	Control the syntatic complexity (syntatic embedding)	0.31
Aggregation Operations		
Period	Leave two propositions in their own sents	0.04
With cue word	Aggregate propositions using *with*	0.51
Conjunction	Join two propositions using a conjunction, or a comma if more than two propositions	0.21
Merge	Merge subject and verb of two propositions	0.87
Also-Cue Word	Join two propositions using *also*	0.05
Contrast-Cue word	Contrast two propositions using *while, but, however, on the other hand*	0.85
Justify-Cue Word	Justify proposition using *because*, *since*, *so*	0.48
Merge with Comma	Restate proposition by repeat only the object	0.42
Pragmatic Markers		
Stuttering	Duplicate first letters of a name	0.54
Pronominalization	Replace occurrences of names by pronouns	1.00
Softener Hedges	Insert syntactic elements to mitigate strength of a proposition	1.00
Emphasizer Hedges	Insert syntactic elements to strengthen a proposition	1.00
Acknowledgments	Insert an initial back-channel	1.00
Filled Pauses	Insert syntactic elements	1.00
Tag Question	Insert a tag question	1.00
Lexical Choice		
Lexicon Frequency	Control average freq of use of each content word, according to BNC frequency counts	0.19
Lexicon Word Length	Control average number of letters of each content word	0.13
Verb Strength	Control the strength of the verbs	0.59

Table 7: Sample Learned Character Model. Only nonzero parameters are shown.

Table 9 shows the average similarity score judgments between utterances produced with a particular character model and the utterances of that character in the original film. For example Row 1 shows the judgments for the similarity of utterances generated with each character model to the utterances of the Alvy character in the original *Annie Hall* screen play. The strongest possible result would be a diagonal matrix with 7's along the diagonal and 0's in all the other cells, i.e. a only utterances generated with a particular character's model would be judged as being at all similar to that character. In general, what we are looking

Film, Character, and Generated Utterances	
***Annie Hall*: Alvy**	***Indiana Jones*: Indy**
- I don't know. People say Cartmill is st-strange, alright? Err... on the other hand, I don't rush to judgment.	- I don't rush to judgment, but people say Cartmill is strange.
- Right, I am not sure, would you be? I will tell something you because you br-brought me cabbage.	- I will tell something you since you brought me cabbage.
- Oh I am not sure. Wolf wears a hard shell. On the other hand, he is ge-ge-gentle, isn't he?	- Wolf is gentle but he wears a hard shell.
- I see, I don't know. I respect Wolf, wouldn't you? He, however, isn't my close friend.	- Wolf isn't my close friend. But I respect him.
- Yeah, I don't know. Sparrow conveys excitement to my life, so I am fr-fr-friends with her.	- I am friends with Sparrow since she brings excitement to my life.
***Annie Hall*: Annie**	***Pulp Fiction*: Vincent**
- Come on, I don't know, do you? People say Cartmill is strange while I don't rush to um.. judgment.	- Basically, I don't rush to judgment. On the other hand, people say Cartmill is strange, he is strange.
- I don't know. I think that you brought me cabbage, so I will tell something to you, alright?	- Yeah, I can answer since you brought me cabbage that.
- Yeah, I am not sure, would you be? Wolf wears a hard shell but he is really gentle.	- Everybody knows that Wolf wears a hard shell. He, however, is gentle.
- I see, I am not sure. Obviously, I respect Wolf. However, he isn't my close friend, is he?	- I respect Wolf. However, he isn't my damn close friend.
- Come on, I am not sure. Because Sparrow brings excitement to my life, I am friends with her, you see?	- Oh God I am friends with Sparrow because she brings excitement to my life.

Table 8: Utterances for *SpyFeet* generated using Film Character Models

for is a matrix with the highest values along the diagonal. From the similarity scores we can see that Alvy, Annie, Indy, and Vincent were all being perceived as being similar to the film characters they represented originally. On the other hand, Marion seems to be confused with Mia or Vincent, and Mia seems to be confused with Indy and Vincent. One possible reason could be that Mia and Marion are the strong female types, which can be perceived as being male, if we solely rely on text utterances.

Our work demonstrated the development of models of character linguistic style from examples, specifically using character utterances in film scripts. Our results are encouraging, showing that utterances generated in a different domain recognizably display important subtext for character personality as well as style that is more similar to the modeled character than to others.

Character	Alvy	Annie	Indy	Marion	Mia	Vincent
Alvy	5.2	**4.2**	**2.1**	**2.6**	**2.8**	**2.3**
Annie	4.2	4.3	**2.8**	**3.4**	3.9	**2.9**
Indy	**1.4**	**2.2**	4.5	**2.8**	**3.3**	**3.8**
Marion	**1.6**	**2.8**	3.7	3.1	**4.1**	**4.2**
Mia	**1.7**	**2.4**	4.3	3.2	3.6	4.3
Vincent	**2.1**	**3.2**	4.5	**3.5**	**3.6**	4.6

Table 9: Mean Similarity Scores between Characters and Character Models. Significant differences between the designated character and each other character are shown in **bold**.

4. Future Augmentation

We would like to augment our current corpus with dialogue from long running television series. This would allow us to collect enough dialogue to learn very detailed models. With scripts from television series, we could also investigate whether the same character, when scripted by different authors as often happens in a television series, differs stylistically and to what degree.

Additionally, we would like to more thoroughly evaluate the accuracy of our automatically generated annotations. For the purposes of our initial generation experiments, precise annotation was not essential, however this data would be valuable for future work and for anyone wishing to use our corpus.

5. Conclusion

We have presented a new annotated corpus of film dialogue and how we have used it to learn character models for generating expressive dialogue. In our perceptual study using generated utterances from these character models, we found that subjects were able to discern similar personality traits between original film dialogue and generated dialogue for another domain.

We believe that our current work on identifying character styles in film, as well as our continuing work on expressive dialogue generation, take important steps towards building tools to assist in the creative process which will help alleviate the authoring bottleneck for content rich applications such as interactive stories. These techniques could also be applied to other domains, such as task-oriented dialogue systems or recommender systems. We hope that by releasing our film corpus, we may enable others to explore the possibilities in their respective domains of interest. Our corpus will be released at `http://nlds.soe.ucsc.edu/software`.

6. References

E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. 2000. The automated design of believable dialogues for animated presentation teams. In J. Sullivan S. Prevost J. Cassell and E. Churchill, editors, *Embodied conversational agents*, pages 220–255. MIT Press, Cambridge, MA.

P. Brown and S. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge University Press.

M. Cavazza and F. Charles. 2005. Dialogue generation in character-based interactive storytelling. In *AAAI First Annual Artificial Intelligence and Interactive Digital Entertainment Conference, Marina del Rey, California, USA*.

M.A. Faber and J.D. Mayer. 2009. Resonance to archetypes in media: There's some accounting for taste. *Journal of Research in Personality*, 43(3):307 – 322.

E.N. Forsyth and C.H. Martell. 2007. Lexical and discourse analysis of online chat dialog. *IEEE Computer Society*.

A. Furnham. 1990. Language and personality. In H. Giles and W. Robinson, editors, *Handbook of Language and Social Psychology*. Winley.

S. Gupta, M.A. Walker, and D.M. Romano. 2007. How rude are you?: Evaluating politeness and affect in interaction. In *Proceedings of ACII*, pages 203–217.

B. Hayes-Roth and L. Brownston. 1994. Multiagent collaboration in directed improvisation. Technical Report

KSL 94-69, Knowledge Systems Laboratory, Stanford University.

M.E. Ireland and J.W. Pennebaker. 2011. Authors' gender predicts their characters' language. *In submission*.

L. Johnson, R. Mayer, E. André, and M. Rehm. 2005. Cross-cultural evaluation of politeness in tactics for pedagogical agents. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED)*.

G.I. Lin and M.A. Walker. 2011. All the world's a stage: Learning character models from film. In *Proceedings of the Seventh AI and Interactive Digital Entertainment Conference*, AIIDE '11. AAAI.

S. Louchart, R. Aylett, J. Dias, and A. Paiva. 2005. Unscripted Narrative for affectively driven characters. *Proc. First International conference on Artificial Intelligence and Interactive Digital Media*.

F. Mairesse and M.A. Walker. 2010. Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction*, pages 1–52.

F. Mairesse and M.A. Walker. 2011. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*.

F. Mairesse, M.A. Walker, M.R. Mehl, and R.K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research (JAIR)*, 30:457–500.

M. Mateas and A. Stern. 2003. Façade: An experiment in building a fully-realized interactive drama. In *Proceedings of the Game Developers Conference, Game Design track*.

M. Mateas. 2007. The authoring bottleneck in creating AI-based interactive stories. In *Proceedings of the AAAI 2007 Fall Symposium on Intelligent Narrative Technologies*.

B.W. Mott and J.C. Lester. 2006. U-director: a decision-theoretic narrative planning architecture for storytelling environments. *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pages 977–984.

J.H. Murray. 1997. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. The Free Press New York, NY, USA.

J.W. Pennebaker and L.A. King. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77:1296–1312.

P. Piwek. 2003. A flexible pragmatics-driven language generator for animated agents. In *Proceedings of Annual Meeting of the European Chapter of the Association for Computational Linguistics (EACL)*.

A. Reed, B. Samuel, A. Sullivan, R. Grant, A. Grow, J. Lazaro, J. Mahal, S. Kurniawan, M. Walker, and N. Wardrip-Fruin. 2011. A step towards the future of role-playing games: The spyfeet mobile rpg project. In *Proc. of the Seventh AIIDE Conf.*. AAAI.

J. Rowe, E. Ha, and J. Lester. 2008. Archetype-Driven Character Dialogue Generation for Interactive Narrative. In *Intelligent Virtual Agents*, pages 45–58. Springer.

D. Thue, V. Bulitko, M. Spetch, and T. Romanuik. 2010. Player Agency and the Relevance of Decisions. *Interactive Storytelling*, pages 210–215.

M.A. Walker and O. Rambow. 2002. Spoken language generation. *Computer Speech and Language, Special Issue on Spoken Language Generation*, 16(3-4):273–281.

M.A. Walker, J.E. Cahn, and S.J. Whittaker. 1997. Improvising linguistic style: Social and affective bases for agent personality. In *Proceedings of the 1st Conference on Autonomous Agents*, pages 96–105.

M.A. Walker, R. Grant, J. Sawyer, G.I. Lin, N. Wardrip-Fruin, and M. Buell. 2011. Perceived or Not Perceived: Film character Models for Expressive NLG. *In Interactive Storytelling: Fourth Joint Conference on Interactive Digital Storytelling*, ICIDS 2011.

N. Wang, W. Lewis Johnson, R.E. Mayer, P. Rizzo, E. Shaw, and H. Collins. 2005. The politeness effect: Pedagogical agents and learning gains. *Frontiers in Artificial Intelligence and Applications*, 125:686–693.