All the World's a Stage: Learning Character Models from Film

Grace I. Lin and **Marilyn A. Walker**
Natural Language and Dialogue Systems Lab, Computer Science Dept.
University of California, Santa Cruz
1156 High Street
Santa Cruz, California 95064
glin|maw@soe.ucsc.edu
http://nlds.soe.ucsc.edu

Abstract

Many forms of interactive digital entertainment involve interacting with virtual dramatic characters. Our long term goal is to procedurally generate character dialogue behavior that automatically mimics, or blends, the style of existing characters. In this paper, we show how linguistic elements in character dialogue can define the style of characters in our RPG SpyFeet. We utilize a corpus of 862 film scripts from the IMSDb website, representing 7,400 characters, 664,000 lines of dialogue and 9,599,000 word tokens. We utilize counts of linguistic reflexes that have been used previously for personality or author recognition to discriminate different character types. With classification experiments, we show that different types of characters can be distinguished at accuracies up to 83% over a baseline of 20%. We discuss the characteristics of the learned models and show how they can be used to mimic particular film characters.

Introduction

Many forms of interactive digital entertainment involve interacting with virtual dramatic characters. While there has been tremendous progress on methods for procedural generation of character physical behaviors, character dialogue is still largely hand-crafted. Expressive natural language generation (ENLG) has the potential to support automatic rendering of character's linguistic behaviors, but there is still much work to be done. The primary technical aims in SpyFeet, an outdoor role-playing game (RPG) (Reed et al. 2011; Sullivan, Mateas, and Wardrip-Fruin 2009), is to integrate dynamic quest selection with an ENLG engine, so that (1) the player can choose to interact with **any** character to carry out any quest; and (2) the player's dialogue interaction with non-player characters (NPCs) is personalized to reflect the player choices, history and affinities. Thus we aim to integrate deep story representation with automatic expressive generation of surface utterances, an aim which we share with others working on interactive story (Riedl and Young 2004; Piwek 2003; André et al. 2000; Lester et al. 1997; Callaway and Lester 2001; Cavazza and Charles 2005; Rowe, Ha, and Lester 2008).

To achieve this goal, we are developing a new tool: Character Creator (CC). CC requires three components: (1) a language generation engine that provides a large number of CHARACTER-RELEVANT PARAMETERS that manipulate syntactic and pragmatic aspects of utterances, and can flexibly decide whether to include non-essential content evoking the player's history, choices and affinities with NPCs; (2) CHARACTER MODELS that specify how to combine the parameters in order to achieve particular effects on the user's perceptions of dramatic character; (3) AUTHORING TOOLS that expose these functionalities to authors in a way that makes sense to them. The question we examine in this paper is: How can we determine what we need in terms of character-relevant parameters and how can we construct character models to control them?

SCENE: LOBBY of Sports Club
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
So ... *Clears his throat*. 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?

Figure 1: Scene from Annie Hall.

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; Brown and Levinson 1987; Walker, Cahn, and Whittaker 1997; Wang et al. 2005; Rowe, Ha, and Lester 2008; Cavazza and Charles 2005) *inter alia*. While politeness and personality theories provide both character relevant parameters and models for controlling them, they do not, in any obvious way, map onto the way that authors of (interactive) stories think about character or dialogue. 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, Ha, and Lester (2008) produce heuristic models of character behavior using a taxonomy of 45 Master Archetypes (Schmidt 2007), and show how archetype models can be integrated with dialogue models. However, our perception was that taxonomies of character archetypes are difficult to operationalize; this is not surprising since their primary aim is to assist the writing practice of authors, rather than to offer a detailed inventory of parameters and models to control them in a computational framework. We concluded that it would be useful to examine how **authors** actually operationalize character when writing dialogue.

Here, we show how to define both character parameters and character models through an automatic corpus-based analysis of film screenplays, such as the example in Figure 1 from Woody Allen's *Annie Hall*. To our knowledge, no prior work has analyzed theatrical or film dialogue from a natural language processing perspective for the purpose of developing computational models of character (Oberlander and Brew 2000; Vogel and Lynch 2008; Ireland and Pennebaker 2011). In this paper, we show that we can learn at least two different kinds of models from film dialogue. First, for individual characters we learn models that indicate significant differences in linguistic behaviors between an individual character such as Annie in *Annie Hall* and other female characters. Second, we show that we can learn models for groups of characters with accuracies up to 83% over a baseline of 20% based on character gender, film genre and director. We describe how to use the individual models to set 10 to 30 parameters of the PERSONAGE generator. We leave to future work the application of group models and a perceptual test of both types of models.

Experimental Method

Our corpus consists of 862 film scripts from The Internet Movie Script Database (IMSDb) website, 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 provided an example of the corpus that we use to derive character models. We believe that the stylized, crafted aspects of film dialogue are actually useful for our purposes. Film dialogue is authored deliberately in order to convey the feelings, thoughts and perceptions of the character being portrayed, and the screenplay often specifies the emotion of an utterance with psychological state descriptors. In addition, the dialogue is deliberately constructed to 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.

We use 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 3. Previous work suggests that females and males in each genre might have different linguistic styles (Ireland and Pennebaker 2011), so we use 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. To summarize our method, we:

1. Collect movie scripts from 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 we wish to mimic; they must have at least 60 turns of dialogue; this is an arbitrary threshold we set to find leading roles within films;
4. Extract counts (features) reflecting particular linguistic behaviors for each character;
5. Learn models of character types based on these features;
6. Use models to control parameters of the PERSONAGE generator (Mairesse and Walker 2010).

Below we describe in detail Steps 4 to 6 of our method.

Extracting Film Dialogue Features

Procedurally generating interesting dialogue requires a large number of parameters for manipulating linguistic behavior. In step 4 of our method, in order to infer important parameters, we have to count features that correspond to them. Table 1 enumerates all our feature sets, which are described in detail below. We start by counting linguistic reflexes that have been useful in prior work characterizing individual differences in linguistic behavior due to personality and social class. While we believe that there are aspects of character not captured with this feature inventory, we attempt to quantify the extent to which they discriminate between different types of characters, and what the learned models tell us about differences in character types.

Set:Description
Basic: number of sentences, sentences per turn, number of verbs, number of verbs per sentence
LIWC Word Categories. Anger (hate, kill, pissed), Social processes (talk, us, friend), Friends (pal, buddy, coworker), Causation (because, know, ought), Discrepancy (should, would, could), Assents (yes, OK, mmhmm), Tentative (maybe, perhaps, guess), etc.
Dialogue Act: Accept, Bye, Clarify, Continuer, Emotion, Emphasis, Greet, No-Answer, Reject, Statement, Wh-Question, Yes-Answer, Yes-No-Question, Other
First Dialogue Act: Same as DA but only look at first sentence of each turn.
Pragmatic Markers: Word counts and ratios, plus word category counts: p-taboo, p-seq, p-opinion, p-aggregation, p-softeners, p-emphatics, p-ack, p-pauses, p-concession, p-concede, p-justify, p-contrast, p-conjunction, p-ingroup, p-near-swear, p-relative
Polarity: overall polarity, polarity of sentences, polarity for concessions
Merge Ratio: merging of subject and verb of two propositions
Tag Question Ratio: number of sentences with tag questions out of all sentences
Average Content Word Length: content words are noun, adjective, adverb, and verb; average words' length
Verb Strength: average sentiment values of verbs
Passive Sentence Ratio: number of passive sentences out of all sentences

Table 1: Summary of Feature Sets

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.

LIWC: The LIWC tool provides a lexical hierarchy that tells us whether characters use different types of words, such as positive and negative emotion words, or anger words. Examples of LIWC word categories are given in Table **??**. 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.

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.

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.

Polarity: Positive and negative polarity are determined by using SentiWordNet 3.0 (ref: *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 following range for score s: strong negative $s < -2/3$, negative $-2/3 \leq s < -1/3$, weak negative $-1/3 \leq s < 0$, neutral $s == 0$, weak positive $0 < s \leq 1/3$, positive $1/3 < s \leq 2/3$, and strong positive $2/3 < s$. For concession polarity, find the polarity for concession part of the sentence, if exists, using the Polarity feature set.

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.

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.

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.

Learning Character Models

In step 5 of our method, we first train models using vectors of features representing individual characters (learn from z-scores) and then train models representing groups of characters (learn from classification). The individual models are trained by normalizing the individual character model against a representative population; for example we normalize Annie in *Annie Hall* against all female characters. Any z-score greater than 1 or less than -1 is more than one standard deviation away from the mean. Z-scores greater and less than ± 1.96 are statistically significant differences of that character compared to other characters.

We train the classification models for groups of characters using Weka's *ZeroR* (majority class; used as baseline) and *J48* pruned decision tree, using 10-fold cross validation. We first select a subset of relevant features using the search method as best first, forward. The feature subset evaluator used is CFS (correlation-based feature subset). We report results for average classification accuracy over all folds.

Character Models from Z-Scores

We build character models by comparing individual characters to a population of same gender characters and extracting attributes with z-scores >1 or <-1, i.e. more than one standard deviation away from the mean. See Table 2. These high and low z-scores indicate the unique attributes that make particular characters stand out from his/her gender population.

For example, the model for Annie from *Annie Hall* in Table 2 shows that she uses many nonfluencies (LIWC-Nonfl) such as *Um* and *Uh*. She says *yes, yea* a lot, especially at the beginning of her utterances (Accept first ratio, LIWC-Assent, yeah ratio). She produces short sentences but she talks a lot. She uses a lot of tag questions (tag question ratio). She does not use long words (LIWC-6LTR and LIWC-Unique both below the mean. In addition, she utilizes pragmatic marker transformations for emphasis and hedging such as *really*, *sort of*, *I think* (really ratio, sort of ratio, I think ratio).

In contrast, the character Lisa from Rear Window produces long and complex sentences (LIWC-WPS, LIWC-Preps, verbs per sentence). She also uses many constructions indicating reasoning such as rhetorical relations for *because* and *even if*. She also starts off utterances with *no* and hedges her assertions with *it seems, kind of* and *I mean*. See Table 2.

For males, the model in Table 2 indicates that Col. Landa from *Inglourious Basterds* uses words like *oh well* and *however* more frequently than males in general. He also uses more words per sentence and longer content words, indicating a higher level of education. He also is less likely to hedge, to assent, to talk about himself, or say something negative, than other male characters.

We hypothesized that the specificity of our models would be a function of the size of the training corpus. Fig. 2 illustrates the effect of number of dialogue turns on the number of significant attributes in the model derived from z-scores. Using 3 male and 3 female characters, each with a relatively large number of turns, Fig. 2 shows an increasing trend for z>2 and z<-2, as well as z>3 and z<-3 as the number of sample utterances increases.

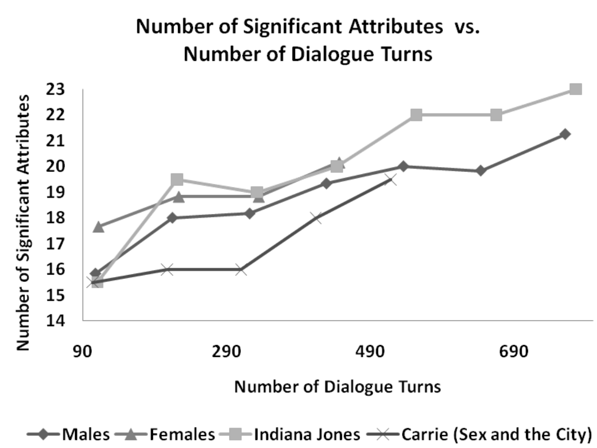

Figure 2: Effect of Corpus Size on Character Models

Character Models from Classification Models

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 3. The results show that we can discriminate two-class GENRE X GENDER categories

Gender	Director (Film)	Character	Z-Scores >1 or <-1
Female	Woody Allen (*Annie Hall*)	Annie	LIWC-Nonfl (10.6), Accept first ratio (8.1), LIWC-Assent (4.8), tag question ratio (3.3), polarity overall (3.1), num sentences (3.0), LIWC-WC (2.2), *really* ratio (1.6), *sort of* ratio (1.4), *yeah* ratio (1.2), LIWC-I (1.2), LIWC-Self (1.1), *I think* ratio (1.0), verbs per sents (-1.0), word length (-1.1), *just* ratio (-1.1), LIWC-Otheref (-1.3), LIWC-Sixltr (-1.3), concession polarity (-1.5), LIWC-Discrep (-1.6), LIWC-Unique (-2.2), LIWC-Preps (-2.7)
Female	Quentin Tarantino (*Inglourious Basterds*)	Bridget	*I see* ratio (8.6), category *with* ratio (5.8), LIWC-Sixltr (2.3), word length (2.0), Reject first ratio (1.5), LIWC-WPS (1.4), LIWC-Friends (1.9), num sents per turn (1.0), polarity overall (1.4), verb strength (1.2), LIWC-Self (-1.0), *around* ratio (-1.1), LIWC-Negemo (-1.1), *oh* ratio (-1.1), tag question ratio (-1.1), *I think* ratio (-1.1), concession polarity (-1.5), LIWC-Qmarks (-1.6), *right* ratio (-1.6) LIWC-You (-1.7), LIWC-Pronoun (-1.8), LIWC-Otheref (-1.9)
Female	Alfred Hitchcock (*Rear Window*)	Lisa	*because* ratio (3.3), Reject first ratio (2.1), *it seems* ratio (1.9), *even if* ratio (1.7), *I mean* ratio (1.6), LIWC-Discrep (1.4), *kind of* ratio (1.4), *even if* ratio (1.4), LIWC-Incl (1.3), LIWC-Preps (1.3), LIWC-WPS (1.3), verbs per sentence (1.3), *right* ratio (1.2), *just* ratio (1.2), LIWC-Assent (-1.0), LIWC-Period (-1.1), *really* ratio (-1.1), *so* ratio (-2.6)
Male	Quentin Tarantino (*Inglourious Basterds*)	Col. Landa	*oh well* ratio (9.0), *however* ratio (5.0), LIWC-WPS (3.6), *quite* ratio (3.3), *actually* ratio (3.2), LIWC-WC (2.5), word length (2.4), verbs per sent (2.2), *on the other hand* ratio (2.1), LIWC-Sixltr (2.1), *however* ratio (2.0), repeated verbs per sent (1.8), *oh well* ratio (1.7), *on the other hand* ratio (1.6), num sents per turn (1.5), LIWC-Preps (1.0), *I think* ratio (-1.1), *yeah* ratio (-1.1), LIWC-Pronoun (-1.2), LIWC-Self (-1.2), LIWC-Negate (-1.4), *though* ratio (-1.4), LIWC-Period (-1.7)
Male	Steven Spielberg (*Saving Private Ryan*)	Jackson	*it seems* ratio (7.6), LIWC-WPS (3.3), *I mean* ratio (2.8), Reject first ratio (2.8), verbs per sentence (2.2), category *with* ratio (2.1), *right* ratio (2.0), merge ratio (1.8), repeated verbs per sent (1.5), word length (1.4), LIWC-Preps (1.4), *kind of* ratio (1.3), LIWC-Sixltr(1.2), Reject first ratio (1.1), LIWC-Incl (1.1), LIWC-Unique (1.1), *just* ratio (1.0), LIWC-Discrep (-1.0), *yeah* ratio (-1.1), *around* ratio (-1.1), num sents per turn (-1.3), LIWC-Qmarks (-1.4), num sents (-1.4), LIWC-You (-1.4), *though* ratio (-1.4), *while* ratio (-1.4), LIWC-Negate (-1.5), concession polarity (-1.6), LIWC-Period (-1.6), LIWC-Pronoun (-1.6), LIWC-Cause (-1.6), *you know* ratio (-1.8), LIWC-Otheref (-2.1)
Male	Alfred Hitchcock (*The Birds*)	Mitch	*it seems* ratio (18.6), *right* ratio (1.4), LIWC-Family (1.4), tag question ratio (1.4), verb strength (1.1), *I think* ratio (1.1), LIWC-Certain (1.0), LIWC-Anger (-1.0), num sents per turn (-1.1), LIWC-Unique (-1.3), *though* ratio (-1.4)
Male	Clint Eastwood (*Gran Torino*)	Walt	LIWC-WC (5.4), num of sents (4.7), LIWC-Nonfl (2.1), LIWC-Incl (1.3), num of sents per turn (1.2), LIWC-Preps (1.2), *around* ratio (1.1), Reject first ratio (1.1), LIWC-Pronoun (-1.1), LIWC-Self (-1.2), *though* ratio (-1.4), concession polarity (-1.6), LIWC-Unique (-2.0)

Table 2: Z-Scores for Selected Characters

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

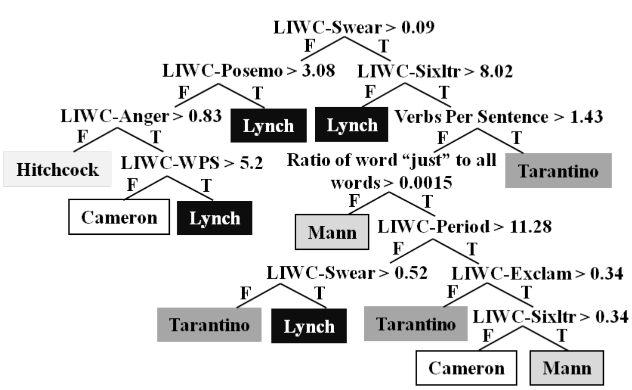

Figure 3: One Instance of Five Directors Model using PERSONAGE features

of character using binary classification models with accuracies over 70% as opposed to baselines around 50%. In many cases, the learned models focus on particularly salient stylistic differences. One such model indicates that Western males can be distinguished from Animation males by: 1) the use of shorter words (LIWC-Sixltr); 2) the use of causal process words (LIWC-Cause); and 3) less use of the phrase *I think*.

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%.

Since J48 decision trees can be quite large, we only discuss one of the more complex group character models. One decision tree (out of 10-fold cross-validation) of the five directors model for Mann (*Public Enemies*), Hitchcock (*Rear Window*), Lynch (*Dune*), Cameron (*Terminator*), and Tarantino (*Pulp Fiction*) is shown in Figure 3. The baseline accuracy is 18.35% while the accuracy of the J48 tree averaged over 10-fold cross-validation results is 64.22% (see Table 3). The size of the trees ranges from 23 to 31 nodes, with depth of up to 10 levels.

We first examine the number of paths that lead to each director (depth of each leaf node). Averaging them over 10 trees, we noticed that Hitchcock has the least number of paths (1.7), followed by Tarantino (2.2), Cameron (2.4), Mann (3), and Lynch (5.1). This suggests that Mann and Lynch characters are more diverse than those in Hitchcock or Tarantino films. Perhaps they should not be treated uniformly as interchangeably indicating director style.

To look at interesting characteristics, we examine the root node, depth to leaf nodes, and the nodes that split the leafs (directors). The root node can be important as it provides a big picture of character styles. We see that Mann and Cameron do not use complicated sentences (number of words per sentence), as indicated by six of the 10 trees. In contrast, other directors seem to be more varied. The root node of the remaining four trees is the use of swear words (LIWC-Swear), which take on a crucial role in social interactions for gender as well as age group. We note the generous usage of swear words in Tarantino's characters, and their almost non-existence in Hitchcock's characters.

We look at longer paths ($\geq$6 nodes) to leaves, as they indicate the the need for more attributes to distinguish the styles of two directors. The splitting directories include Lynch/Tarantino, Lynch/Cameron, Lynch/Mann, Mann/Cameron, and Tarantino/Mann. Notice that Hitchcock was not in any of the combinations. This shows that Hitchcock's style is probably the most unique among the five directors, while Lynch and Mann are more diverse in general.

Lastly we look at the leaf nodes that split directors. This could be beneficial as the final attribute is the determining factor that separates directors with similar characteristics. The attributed used most frequently for final splits of two directors are positive emotions (LIWC-Posemo), anger (LIWC-Anger), and swear words (LIWC-Swear). Some interesting splits include Tarantino uses less "friends" type of words (LIWC-Friends) than Mann; Mann uses more exclamation marks than Hitchcock; and Cameron uses more tentatives (LIWC-Tentat) than Mann. As expected, Hitchcock/Tarantino split at swear words (LIWC-Swear),

There is no split at the leaf nodes for the combinations of Cameron/Tarantino and Cameron/Hitchcock. This indicates Cameron is quite different from Tarantino and Hitchcock, resulting in fewer common attributes in the classification trees. We leave testing these classification-based character models to future work.

Personage Utterances from Character Models

Step 6 of our method maps each feature, or combinations of them, to the generation parameters of the PERSONAGE engine (Mairesse and Walker 2010). For example, the *Pragmatic Markers* features in Table 1 each correspond to aggregation parameters in PERSONAGE or pragmatic transformations, such as inserting emphasizers or hedges.

So far we tested our character models learned from z-scores in the context of SpyFeet to control parameters of the PERSONAGE generator. We start with a default character model that represents "neutral" personality. As each character model only has a subset of all possible attributes that are significant (Table 2), these attributes modify their corresponding PERSONAGE parameters. A sample character model for Annie from *Annie Hall* is shown in Table 4. Each attribute of the character model can be mapped to one or more PERSONAGE parameters, and vice versa. For example,

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 4: Sample Learned Character Model. Only nonzero parameters are shown.

Annie's non-fluencies (LIWC-Nonfl, see Table 2) is used to control PERSONAGE parameters *Filled Pauses* and *Stuttering*.

Table 5 illustrates the result of applying example character models to SpyFeet utterances, and some of the variations in style that we are currently able to produce. The refinement of the mapping and development of the generation dictionary for SpyFeet is work in progress.

Discussion and Future work

The demand for more engaging and personalized gaming experiences, along with the need for faster production of games will force the gaming industry to find ways of automating different creative aspects of games while retaining the same level of quality. Our current work on identifying character styles in film, and our continuing work on dialogue generation for interactive story and authoring tools take an important step to assist an author's creative process in constructing character dialogue.

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 5: Utterances for SpyFeet generated using Film Character Models

We have shown how we learn character models from film dialogue in order to define character types for SpyFeet. The models are based on features that can be extracted fully automatically from screenplays. The learned models identify features, and the corresponding generation parameters in PERSONAGE that can be used to produce utterances in dialogue whose style should match a particular character or group of characters. It may also be possible to generate even better models to better represent these characters, by creating additional relevant features and gathering additional film scripts from IMSDb.

Character models provide one target end point for rendering dialogue using methods for procedural dialogue generation for interactive story and games. In future work, we need to test these models and verify that they have the perceptual effects that we hypothesize they will have. In addition, we need to conduct experiments to test whether the idea of a character sample or example provides a useful interface to authoring characters that fits with an author's conceptual framework of character types.

Acknowledgments

We would like to thank the support of a GAANN Fellowship to Grace Lin, an NSF IIS-RI EAGER Award #RI-1046437 for *SpyFeet*, and NSF Creative IT Award #IIS-1002921 for *Character Creator*. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF).

References

André, E.; Rist, T.; van Mulken, S.; Klesen, M.; and Baldes, S. 2000. The automated design of believable dialogues for animated presentation teams. In *Embodied conversational agents*. MIT Press. 220–255.

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

Callaway, C., and Lester, J. 2001. Narrative prose generation. In *Proc. of the Seventeenth Int. Joint Conference on Artificial Intelligence*.

Cavazza, M., and Charles, F. 2005. Dialogue generation in character-based interactive storytelling. In *Artificial Intelligence and Interactive Digital Entertainment*.

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

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

Lester, J.; Converse, S.; Kahler, S.; Barlow, S.; Stone, B.; and Bhogal, R. 1997. The persona effect: affective impact of animated pedagogical agents. *Proc. of the SIGCHI conference on Human factors in computing systems* 359–366.

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

Oberlander, J., and Brew, C. 2000. Stochastic text generation. *Philosophical Transactions of the Royal Society of London* Series A, 358:1373–1385.

Piwek, P. 2003. A flexible pragmatics-driven language generator for animated agents. In *Proc. of Annual Meeting of the European Association for Computational Linguistics*.

Reed, A.; Samuel, B.; Sullivan, A.; Grant, R.; Grow, A.; Lazaro, J.; Mahal, J.; S, K.; Walker, M.; and Wardrip-Fruin, N. 2011. Spyfeet: An exercise rpg. In *Foundations of Digital Games*.

Riedl, M., and Young, R. M. 2004. An intent-driven planner for multi-agent story generation. In *Proc. of the 3rd Int. Conf. on Autonomous Agents and Multi Agent Systems*.

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

Schmidt, V. 2007. *45 Master characters*. Writers Digest Books.

Sullivan, A.; Mateas, M.; and Wardrip-Fruin, N. 2009. Questbrowser: Making quests playable with computer- assisted design. In *Proc. of Digital Arts and Culture*.

Vogel, C., and Lynch, G. 2008. Computational Stylometry: Whos in a Play? *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction* 169–186.

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

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