

Stylistic Variation in Television Dialogue for Natural Language Generation

Grace I. Lin and Marilyn A. Walker

Natural Language and Dialogue Systems Lab
University of California, Santa Cruz
{glin5, mawwalker}@ucsc.edu

Abstract

Conversation is a critical component of storytelling, where key information is often revealed by what/how a character says it. We focus on the issue of character voice and build stylistic models with linguistic features related to natural language generation decisions. Using a dialogue corpus of the television series, *The Big Bang Theory*, we apply content analysis to extract relevant linguistic features to build character-based stylistic models, and we test the model-fit through an user perceptual experiment with Amazon's Mechanical Turk. The results are encouraging in that human subjects tend to perceive the generated utterances as being more similar to the character they are modeled on, than to another random character.

1 Introduction

Conversation is an essential component of social behavior, one of the primary means by which humans express emotions, moods, attitudes, and personality. Conversation is also critical to storytelling, where key information is often revealed by what a character says, how s/he says it, and how s/he reacts to what other characters say. Here we focus on the issue of character voice. One way to produce believable, dramatic dialogue is to build stylistic models with linguistic features related to natural language generation (NLG) decisions. Television dialogue are exemplars of many different linguistic styles that were designed to express dramatic characters. Thus we construct a corpus of television character dialogue from *The Big Bang Theory* (BBT) and apply content analysis and language modeling techniques to extract relevant linguistic features to build character-based stylistic

models. We test the model-fit of character models through a generation experiment to test user perceptions of characters.

Our work can be applied to storytelling applications such as video games, interactive narrative, chatbots, or education systems where dialogue with personalities may improve user experience.

2 Related Work

Research from corpus linguistics include Bednarek's work on using *Gilmore Girls* to compare the genre dramedy to other types (Bednarek, 2011a), and Quaglio's work on using *Friends* with unscripted conversations (Quaglio, 2009). Other related research focuses on characterization through dialogue. For example, Bubel explored the friendship among characters in the *Sex and the City* (Bubel, 2005), and Bednarek analyzed linguistic stylistics shifts from characters from the *Gilmore Girls* (Bednarek, 2011b) and *The Big Bang Theory* (Bednarek, 2012).

Research from computational stylistics (or stylometry) focuses on the use of quantitative methods to study writing styles to characterize authors, which can be applied to many applications such as classical literary text, modern forensic text, and online reviews, just to name a few (Stamatatos, 2009). Principal component analysis is used to analyze the variations in words, focusing on the challenge of relating features and meanings in text, which is not fixed depending on the context (Schreibman et al., 2008).

There is an extensive amount of research in story generation (narrative content), which tends to focus on plots and character development to achieve narrative goals. One source of creating stories comes from crowd participants writing detailed descriptions for events, going into details

with characters’ intentions, facial expressions, and actions (Li et al., 2014). In addition, they used the Google N-Gram Corpus and Project Gutenberg to help select different types of sentences (most/least probable, most fictional, most interesting details) and different sentiments (most positive/negative). Our work is also related to character modeling from film dialogue for NLG (Lin and Walker, 2011; Walker et al., 2011), except that we focused on TV series because they offered more dialogue.

Despite overlaps, our work differs in that we: 1) extract linguistic stylistic features based on personality studies from psychology; 2) focus on features that can be generated given our current system; 3) find significant features and use them as building blocks to 4) create models using techniques such as standard scores and classification; and 5) apply the models to applications such as natural language generation.

Natural Language Generation Engine

PYPER (Bowden et al., 2016) is a spin-off implementation of PERSONAGE (Mairesse and Walker, 2007) in Python that provides new controls for expressive NLG. It is currently part of the M2D Monolog-to-Dialogue generation (Bowden et al., 2016) framework, which we briefly describe the architecture below (Figure 1).

The EST framework (Rishes et al., 2013) produces a story annotated by SCHEHERAZADE (Elson and McKeown, 2009) as a list of sentences represented as Deep Syntactic Structures (DSyntS). DSyntS is therefore a dependency-tree structure with nodes containing lexical information about words. This is the input format for the surface realizer RealPro (). M2D converts the story (list of DSyntS) into two-speaker dialogue by accepting input parameters that control the allocation of content, pragmatic markers, etc.

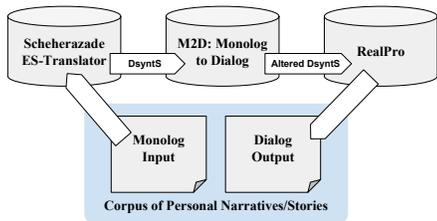


Figure 1: M2D Monolog-to-Dialogue Generation (Bowden et al., 2016)

3 Corpus

We parsed fan-transcribed BBT scripts, seasons 1-4 and partial season 5, to obtain scenes, speakers, and utterances. The series centers around 5 characters, 4 of them (all male) are scientists/engineers working at Caltech, and 1 (Penny) is a waitress. The comedy’s theme focuses on the contrast between the geekiness of the male characters and Penny’s social skills. Two additional female characters, both scientists, were introduced as love interests to two main male characters, and have since become main characters themselves.

4 Stylistic Features Extraction

After extracting dialogic utterances from transcripts, we extract features reflecting particular linguistic behaviors for each character. Table 1 describes major feature sets, which include sentiment polarity, dialogue act, passive voice, word categories from LIWC (Pennebaker et al., 2001), tag questions, etc.

Character Stylistic Models

We calculate a standard score (z-value) for each feature to measure the differences between main characters: Leonard, Sheldon, Penny, Howard, Raj, Bernadette, and Amy. A better measurement could be used due to the small population and normal assumption, however we reviewed the results and they seem to capture enough relative differences among characters. Character models are composed of significant features with $|z| \geq 1$. While using features with $|z| \geq 2$ might be a better choice, our NLG engine can manipulate many features under $|z| \geq 1$.

The number and examples of significant features for each character are shown in Table 2. We see that for $|z| \geq 1$, Sheldon, Penny, Bernadette and Amy have over 200 significant features. Sheldon, more specifically, has close to 400 significant features. When we narrow them down to $z \geq 2$, significant features for Bernadette and Amy decreased by over 85%, Leonard, Penny, Howard, and Raj decreased by 70%, and Sheldon decreased by 54%.

5 Generating Expressive Utterances

The workflow for generation is to 1) annotate stories using SCHEHERAZADE; 2) use EST to automatically translate annotated stories to deep syntactic structures (DSyntS); 3) PYPER reads

Table 1: Automatically Annotated Linguistic Features for TV Dialogue

Feature Set and Description
1. Basic. Tokens per sentence, tokens per utterance, etc., plus words from different types of emotion and other psychological categories from the Nodebox English Linguistics library.
2. Sentiment Polarity. Overall polarity, polarity of sentences, etc., using SentiWordNet ¹ to calculate positive, negative, and neutral score.
3. Dialogue Act. Train Naive Bayes classifier with NPS Chat Corpus’ 15 dialogue act types using simple features. We also determine “First Dialogue Act”, where we look at the dialogue act of the first sentence of each turn.
4. Merge Ratio. Use regular expression to detect the merging of subject and verb of two propositions.
5. Passive Voice. Using a third party software (see text) to detect passive sentences.
6. Concession Polarity. Look for concession cues, then calculate polarity of concession portion.
7. LIWC Categories. Word categories from the Linguistic Inquiry and Word Count (LIWC) text analysis software.
8. Markers - PERSONAGE. collect words used in PERSONAGE for generation, which were selected based on psychological studies to identify pragmatic markers of personality that affect the utterance.
9. Tag Questions. Use regular expression to capture tag questions.
10. Verb Strength. Averaged sentiment values of verbs.
11. Content Words Length. Find the average length of content words.
12. Markers - Others. Inspired by PERSONAGE words. Extended set.
13. Hedges. Collect words from a list of pre-defined hedges and their categories. LACKOFF hedges.
14. Repeating Verbs. Find verbs that are repeated used in a turn.
15. BIGRAMs. Top 10 bigrams.
16. Part-of-Speech BIGRAMs. Top 10 POS bigrams.

Table 2: Number and Examples of Significant Features for *The Big Bang Theory* Characters

Speaker	$ z \geq 1$	$ z \geq 2$	Example Features for $z \geq 1$ (i.e., positive z-values only)
Leonard	172	54	words:[<i>even if, nevertheless, whereas, even though</i>], Dialogue Act-{Greet, Bye}, LIWC-{Causation, Impersonal Pronouns}, hedges per sentence, connect words, concept words
Sheldon	394	180	words: [<i>all the same, although, despite, however, nevertheless, on the other hand, whereas, more or less, though, all, yet</i>], passive-ratio, important words per utt/sent, LIWC-{Inhibition, Prepositions, Number, Quantifiers}
Penny	232	68	words:[<i>nevertheless, even if, while, even though, on the other hand, yet</i>], connect words, emotional words, Dialogue Act-{Greet, Bye}, swear/near swear words, LIWC-{Adverbs, Present Tense, Dictionary Words}
Howard	133	41	words:[<i>although, even if, whereas</i>], LIWC-{Hear, See, Third Person Singular}, concept words, in-group words, hedges-per-sent
Raj	179	51	words:[<i>on the other hand, however, despite, though, also, even though, but</i>], in-group words, LIWC-{Conjunctions, Third Person Plural, See}, hedges per sentence
Bernadette	283	43	persuasive words, emotional words, conceptual words, words:[<i>even though, yet, while</i>], Dialogue Act-emphasis, LIWC-{Personal Pronouns, Second person, Auxiliary Verbs, Function Words, Past Tense}
Amy	246	43	LIWC-{Quantifiers, Unique Words, Future Tense, Causation}, RID Emotion words, Dialogue Act - Continuer, opinion words, words:[<i>though, but</i>]

and manipulates DSyntS to add expressive elements, and 4) send “expressive” DSyntS to Real-Pro (Lavoie and Rambow, 1997) (a sentence realizer) for generation. We focus on operation 3 where we use our learned character stylistic models to add expressive elements to generic sentences.

5.1 Mapping Stylistic Features to NLG Decisions

The re-written and better-controlled PYPER allows for more useful mapping of character models for NLG. For example, hedge insertion patterns are kept in a library where new additions can be

easily added. As an example, a partial mapping for LIWC categories are shown in Table 3. For multiple features mapped to the same PYPER parameter, we calculate a weighted average of the features.

5.2 Narrative Content

Our narrative content comes from fables and stories: 1 fable (*The Fox and the Crow*) and 6 blog stories about garden, protest, squirrel, bug, employer, and storm (Gordon et al., 2007). We use *The Fox and the Crow* fable as an example to describe our process shown in Figure 2.

Some phrases are highlighted to show how they

Table 3: Partial Mapping of LIWC Categories to Expressive NLG Parameters

PYPER Parameter	LIWC Category	PYPER Parameter	LIWC Category
near-expletives	swear, anger	low-expletives	swear, anger
emph-actually	certain	emph-exclamation	excl
emph-really	certain	emph-great	assent
emph-you-know	filler	emph-particularly	certain
emph-technically	certain	emph-literally	certain
emph-quintessential	certain	emph-essentially	certain, i
emph-somewhat	tentat	emph-very	certain
emph-especially	certain	emph-roughly	tentat
in-group-marker	family, friends, we, incl	init-reject	tentat

were annotated and translated. Many complicated sentences have been broken down into shorter ones. Note that some additional descriptions (adjectives) were added in order to provide enough search space for PYPER to exercise enough expressive parameters, so that characters’ personalities will come through in different variations of the story.

The final, expressive version of the story shows different stylistic features such as converting a statement to a question and adding character dialogue inspired expressions such as *Typical*.

6 Evaluation with User Perceptual Experiment

We used Mechanical Turk to get user feedback on the generated dialogue. The PYPER generated output dialogue were post-processed to get rid of typos and minor grammatical issues. Referring to the MTurk survey (one HIT) in Figure 3, we first show some information about the character in interest (Sheldon, in this case), followed by two sets of dialogue: one by Sheldon and the other by a different random character. The worker does not know which one was modeled by Sheldon. S/he was asked to pick the dialogue that sounded most similar to Sheldon, along with providing reasons.

Referring to results in Table 5, we used three participants per pair of characters comparison per story. The character on the left-most column indicates the modeled character, and the compared-to character was the “other random character” in the survey. Each circle (empty or filled) indicates a worker’s choice. A filled circle (●) means the worker picked the “matched” generated dialogue to the intended character, otherwise an empty circle (○) is shown.

The probability that at least two out of three participants agree on the right character is > 50% (Table 4), while all three participants agree on the right character is 31.3%, which is higher than

chance (12.5%).

Table 4: Participants Agreement

Choose the right character	# HITs (out of 294)	%
●●● (3 out of 3)	92	31.3
●●○ (2 out of 3)	57	20.4
●○○ (1 out of 3)	122	41.5
○○○ (0 out of 3)	23	7.8

Overall the 7 characters over all 7 stories were recognized about 65.5% of the time (out of 882 ratings). Per character-wise over all 7 stories, Penny was recognized the most with 82.5% of the time, followed by Leonard (78.6%), Bernadette (66.7%), Amy and Sheldon (both 61.9%), Howard (57.9%), and finally Raj, who was recognized the least with 49.2% of the time.

Certain character pairs were easier to distinguish than others. For example, Leonard-Penny and vice-versa (95.2%), Sheldon-Penny and vice-versa (85.7%, 90.5%), and Amy-Bernadette and vice-versa (85.7%). On the other hand, these were among the pairs harder to distinguish: Amy-Leonard and vice-versa (47.6%, 57.1%), Bernadette-Penny and vice-versa (33.3%, 57.1%), and Sheldon-Howard and vice-versa (47.6%, 57.1%).

7 Character Analysis

In this section we provide some insights to the results, including with participants’ comments about their perception of characters through the generated dialogue.

We found the following pairs of characters most distinguishable from each other (Table 6): Amy vs. Bernadette, Bernadette vs. Howard, Howard, Leonard, Sheldon vs. Penny, Penny vs. Leonard, and Raj vs. Sheldon. On the other hand, we found the most similar pairs of characters to be Amy-Leonard, Bernadette-Penny, Howard-Leonard, Leonard-amy, Penny-Bernadette, Raj-

The Fox and the Crow Partial (Original)	
A Crow was sitting on a branch of a tree with a piece of cheese in her beak when a Fox observed her and set his wits to work to discover some way of getting the cheese.	
SCHEHERAZADE & EST	
The crow sat on the tall tree's branch. The crow has elegant talons. The crow has a good beak. The crow has ugly wings. The crow has bad eyes. The delicious cheese was in the crow's beak. The hungry fox observed the crow. The fox has good claws. The fox has bad teeth. The fox tried to discover for the fox to get the cheese.	
PYPER: Monologue to Dialogue	
Speaker 1: The crow sat on the tall tree's branch. The crow has elegant talons, and a good beak. The crow has ugly wings, and bad eyes.	Speaker 2: The delicious cheese was in the crow's beak. The hungry fox observed the crow. The fox has good claws, but bad teeth.
Speaker 1: The fox tried to discover for the fox to get the cheese. The fox came and stood under the tall tree. The fox looked toward the crow.	Speaker 2: The fox said the fox saw the crow. The fox said the crow's beauty was incomparable. The fox said the crow's plumage's hue was exquisite.
PYPER: M2D + Stylistic Parameters	
Speaker 1: She sat on the tall tree's branch. She has elegant talons, and a good beak. She has ugly wings, and bad eyes.	Speaker 2: He has good claws, but bad teeth.
Speaker 1: What was in her beak?	Speaker 2: The delicious cheese was, i see, in her beak. Typical. Really ok, I do not remember what happened next? Do you want to know more about that?

Figure 2: The *Fox and the Crow* Story Transformation

Leonard, and Sheldon-Howard/Leonard. Note that the comparison is not symmetrical because in the survey we gave a “known” (reference) character, which is the first column in the table.

It is not surprising to see Penny being different from most of the male characters, as it is the premise of BBT. Raj is an exception, mainly due to his lack of (expressive) dialogue, though he is definitely different from Sheldon. It is also believable that Leonard is similar to many characters, as he is the most “normal” character out of the group.

We further explore a few characters below.

7.1 Penny

7.1.1 Perception of Penny in comparison to Leonard (most distinguishable)

Penny is one of the best expressed character in the experiment, missing only by one selection in comparison to Leonard (95.2%), and missing by two in comparison to Sheldon (90.5%). Here we take a look at the comparison with Leonard, where 20 (out of 21) Penny-modeled generated dialogue were rated more similar to Penny, and only 1 (out of 21) Leonard-modeled generated dialogue were rated more similar to Penny.

Overall, participants’ perception of **Penny-modeled generated dialogue** seem to agree with Penny’s personality, capturing her “bubbly, cheerfulness”, as mentioned by one worker. Some no-

table descriptions include:

- talkative, randomness, random pauses, better wording, more personality
- seek feedback from others, lots of questions, not always sure of what she’s saying, hesitation
- good mix of colloquialisms and Penny-like filler, some brief, fairly simple statements
- stand-out word choices: *magic, huh?, mmmm, let’s see, that..., the crow needed what?, oh gosh, I mean, damn yeah*

Participants perceived **Leonard-modeled generated dialogue** as not suitable for Penny, mostly because of his bland language. Here are some notable descriptions:

- too simple, monotone, boring, direct, bare, straightforward, matter-of-fact, boxy, bland, not enough questioning for Penny
- too much adverb usage on precision or intellect for Penny
- not like Penny to use complex words and phrases
- not like Penny to use: *technically, darn*
- too rude for her to use, since she wants people to like her: *everybody knows that, obviously*

The MTurk worker of the one missed selection cited Penny being a very simple speaker, implying that her dialogue would contain brief and simple statements. While this is true, she also uses quite a bit of fillers and questions around her “simple” dialogue to sound chatty.

7.1.2 Perception of Penny in comparison to Bernadette (least distinguishable)

It is not surprising to see Penny being the least distinguishable with Bernadette (57.1%). Bernadette was introduced in the series as Penny’s friend and coworker working as a waitress. Her role on the show seemed to be more similar to Penny (friendly

Instructions

Sheldon is a character from the TV show **"The Big Bang Theory"**. Here are some samples of Sheldon's dialogue showing his character style. It is a bit long, but it's the same dialogue for all 6 HITs in this group. Please take your time and think about what kind of character the dialogue projects in terms of speech patterns.

Scene: A corridor at a sperm bank.

Sheldon: So if a photon is directed through a plane with two slits in it and either slit is observed it will not go through both slits. If it's unobserved it will, however, if it's observed after it's left the plane but before it hits its target, it will not have gone through both slits.

Leonard: Agreed, what's your point?

Sheldon: There's no point, I just think it's a good idea for a tee-shirt.
(Arriving at a location.)

Leonard: Excuse me?

Receptionist: Hang on. (Working on crossword puzzle.)

Leonard: One across is Aegean, eight down is Nabakov, twenty-six across is MCM, fourteen down is... move your finger... phylum, which makes fourteen across Port-au-Prince. See, Papa Doc's capital idea, that's Port-au-Prince. Haiti.

Receptionist: Can I help you?

Leonard: Yes. Um, is this the High IQ sperm bank?

Receptionist: If you have to ask, maybe you shouldn't be here.

Sheldon: I think this is the place.

Receptionist: Fill these out.

Leonard: Thank-you. We'll be right back.

Receptionist: Oh, take your time. I'll just finish my crossword puzzle. Oh wait.

Below shows **computer generated dialogue** between two speakers telling a story about a **garden**. Each speaker's utterances have been generated using a particular character model from the show. Because it's generated automatically by computer, the dialogue have typos and odd phrasings. Please try to look beyond that and see if you can still perceive some personalities.

Read the two versions of the dialogue and then indicate whether Speaker 2 (**S2**) in **Dialogue 1** or **Dialogue 2** is more similar to **Sheldon**. You must answer all questions or your work will be rejected. Note that Dialogue 1 is fixed for all 6 HITs in this group.

Dialogue 1	Dialogue 2
<p>S1: The quite droopy radishes charmed the butterflies. The communal garden was weedy. The communal garden was swampy because rained.</p> <p>S2: Rained, technically. It seems to me that Winston planted the chards the lettuces and the spinach. Winston did not expect for the chards the lettuces and the spinach to grow. Winston planted what, as it were.</p> <p>S1: I suppose the quintessential Winston planted the standing plants. The chards the lettuces and the spinach sprouted. Winston mistakenly dug the chards the lettuces and the spinach.</p> <p>S2: Quite unfortunately, the somewhat quintessential Winston saw that the chards the lettuces and the spinach sprouted. Quite unfortunately, the particularly communal garden was not quite weedy. What happened next?</p> <p>S1: The communal garden was not swampy. The communal garden was productive. The Winston was proud because the communal garden was productive.</p> <p>S2: Actually, Winston wanted to reap the lettuces. I might be wrong but, the quintessential Winston particularly planned to remove the largely quite droopy radishes because the radishes were droopy.</p> <p>S1: Winston thought the flowers charmed the cool butterflies.</p>	<p>S1: The droopy radishes charmed the butterflies. The communal garden was weedy. The communal garden was swampy because rained.</p> <p>S2: Actually, rained. Winston planted the very standing plants , pretty Wait ... Then, you might be somewhat wonderful in knowing that did it not! Winston planted the chards what and the spinach, literally actually. Finally, Really mmhm ... the Winston did not expect for the chards the lettuces and the spinach to grow.</p> <p>S1: The chards the lettuces and the spinach sprouted. Winston mistakenly dug the chards the lettuces and the spinach.</p> <p>S2: I see, Winston saw you might be wonderful in knowing that that the chards the lettuces and the spinach sprouted. Literally.</p> <p>S1: The communal garden was not swampy. Was what not weedy? The communal garden was productive. Winston was proud because the communal garden was productive.</p> <p>S2: Winston wanted to reap the like well, lettuces. Ok, Winston planned to remove the droopy radishes because the radishes were droopy.</p> <p>S1: Winston thought the flowers charmed the cool butterflies.</p>
<input type="radio"/> Pick S2 from Dialogue 1	<input type="radio"/> Pick S2 from Dialogue 2

Why did you pick dialogue 1 or 2? What are your general thoughts on these dialogue?

Figure 3: Amazon Mechanical Turk Survey (One HIT) Example

and sociable) than everyone else (nerdy and socially awkward), despite that she became a scientist eventually.

While the Bernadette-model contain chatty word choices (similar to Penny's), it also contains "intellect" word choices. However due to the randomness of the generated dialogue, where not all features are expressed/activated, some dialogue/story might not show enough of her nerdy side. For example, precise adverbs such as *essentially*, *particularly* are more likely to be used by a

scientist/engineer (Bernadette) but not by Penny.

In terms of stories, Bug and Garden did the best at distinguished the character pair, while Employer and Storm did the worst (none of the Penny-modeled dialogue sounded like Penny).

7.2 Sheldon

7.2.1 Perception of Sheldon in comparison to Penny (most distinguishable)

With Sheldon differs the most with Penny (85.7%), we focus on comments by participants

Table 5: Characters and Stories MTurk Results by HITs

Each HIT had 3 participants, each indicated by a circle (o).
 A solid circle (●) indicates the worker picked the "matched" generated dialogue to the original character.
 Characters are listed in alphabetical order; circles are sorted by ● then o
 blue: best result; red: worst result

Character compared to	Story							#/% similar (out of 21)	
	Bug	Employer	FoxCrow	Garden	Protest	Squirrel	Storm		
Amy	Bernadette	●●●	●●●	●●○	●●●	●●●	●●●	●○○	18 / 85.7
	Howard	○○○	●●○	●●○	●●○	●●○	●●○	●○○	12 / 57.1
	Leonard	○○○	●●○	○○○	●●●	●●○	●●○	●○○	10 / 47.6
	Penny	○○○	●●○	●●○	●●○	●●○	●●○	●○○	13 / 61.9
	Raj	○○○	●●●	○○○	●●○	●●○	●●○	●○○	12 / 57.1
	Sheldon	●●○	●●○	●●●	●●●	●○○	●●○	○○○	13 / 61.9
	# / % similar (out of 18)	8 / 44.4	14 / 77.8	9 / 50.0	15 / 83.3	12 / 66.7	13 / 72.2	7 / 38.9	78 / 61.9 (out of 126)
Bernadette	Amy	●●●	●●○	●●○	●●○	●●○	●●○	●●●	18 / 85.7
	Howard	●●●	●●●	●●○	●●○	●●○	●●○	●●○	18 / 85.7
	Leonard	●●○	●●○	○○○	●●●	○○○	●●○	●●○	13 / 61.9
	Penny	●●○	○○○	○○○	○○○	●●○	○○○	○○○	7 / 33.3
	Raj	●●○	●●●	●●●	○○○	●●○	○○○	●●○	14 / 66.7
	Sheldon	●●○	●●○	●○○	●○○	●○○	●○○	●●○	14 / 66.7
	# / % similar	14 / 77.8	14 / 77.8	10 / 55.6	11 / 61.1	12 / 66.7	9 / 50.0	14 / 77.8	84 / 66.7
Howard	Amy	●●○	●●○	●○○	●●○	●○○	●●○	●○○	12 / 57.1
	Bernadette	●●○	●●○	○○○	●●○	○○○	●●○	●○○	11 / 52.4
	Leonard	●●○	●●○	○○○	○○○	○○○	●●○	●○○	10 / 47.6
	Penny	●●○	○○○	●○○	○○○	●●○	●●○	●○○	15 / 71.4
	Raj	●●○	●●●	●○○	○○○	●●○	●●○	●○○	13 / 61.9
	Sheldon	●●○	●○○	●●○	●○○	●○○	●○○	●○○	12 / 57.1
	# / % similar	13 / 72.2	12 / 66.7	8 / 44.4	7 / 38.9	9 / 50.0	13 / 72.2	11 / 61.1	73 / 57.9
Leonard	Amy	●●○	●●○	●●○	●●○	●○○	○○○	●●○	12 / 57.1
	Bernadette	●○○	●●○	●●○	●●○	○○○	●●○	●●○	15 / 71.4
	Howard	●○○	●●○	●●○	●●○	●●○	●○○	●●○	19 / 90.5
	Penny	●○○	●●○	●●○	●●○	●●○	●●○	●●○	20 / 95.2
	Raj	○○○	●●○	●○○	●●○	●○○	●●○	●○○	16 / 76.2
	Sheldon	●●○	●●○	●○○	●○○	●○○	●○○	●○○	17 / 81.0
	# / % similar	11 / 61.1	16 / 88.9	15 / 83.3	16 / 88.9	12 / 66.7	12 / 66.7	17 / 94.4	99 / 78.6
Penny	Amy	●●○	●●○	●●○	●●○	●●○	●●○	○○○	18 / 85.7
	Bernadette	●●○	○○○	●○○	●●○	●●○	●○○	○○○	12 / 57.1
	Howard	●●○	●●○	●○○	●●○	●●○	○○○	●○○	17 / 81.0
	Leonard	●●○	●●○	●○○	●●○	●●○	●●○	●●○	20 / 95.2
	Raj	●●○	●●○	●○○	●●○	●●○	●○○	●○○	18 / 85.7
	Sheldon	●●○	●●○	●○○	●●○	●●○	●●○	○○○	19 / 90.5
	# / % similar	18 / 100	15 / 83.3	13 / 72.2	18 / 100	17 / 94.4	14 / 77.8	9 / 50.0	104 / 82.5
Raj	Amy	●○○	○○○	●○○	○○○	○○○	○○○	○○○	9 / 42.9
	Bernadette	●○○	○○○	●○○	○○○	○○○	○○○	●○○	10 / 47.6
	Howard	○○○	○○○	●○○	●●○	●●○	○○○	○○○	13 / 61.9
	Leonard	○○○	○○○	●○○	●●○	○○○	○○○	○○○	6 / 28.6
	Penny	○○○	●●○	○○○	○○○	●●○	○○○	●●○	10 / 47.6
	Sheldon	●●○	●●○	●○○	●●○	○○○	●○○	○○○	14 / 66.7
	# / % similar	9 / 50.0	7 / 38.9	11 / 61.1	10 / 55.6	9 / 50.0	6 / 33.3	10 / 55.6	62 / 49.2
Sheldon	Amy	●●○	○○○	○○○	●●○	○○○	●●○	●●○	13 / 61.9
	Bernadette	●○○	○○○	●○○	●●○	●●○	●○○	○○○	16 / 76.2
	Howard	●○○	○○○	●○○	●●○	○○○	○○○	○○○	10 / 47.6
	Leonard	●○○	○○○	●○○	●●○	●●○	○○○	○○○	10 / 47.6
	Penny	●○○	○○○	●○○	●●○	●●○	○○○	●●○	18 / 85.7
	Raj	●●○	●●○	●○○	●○○	○○○	●○○	○○○	11 / 52.4
	# / % similar	14 / 77.8	9 / 50.0	10 / 55.6	17 / 94.4	10 / 55.6	10 / 55.6	8 / 44.4	78 / 61.9
# / % similar (out of 126)	87 / 69.0	87 / 69.0	76 / 60.3	94 / 74.6	81 / 64.3	77 / 61.1	76 / 60.3	578 / 65.5 (out of 882)	

Table 6: Most/Least Distinguishable Characters

Ref. Char	Most distinguishable with	Least distinguishable with
Amy	Bernadette (85.7%)	Leonard (47.6%)
Bernadette	Amy, Howard (85.7%)	Penny (33.3%)
Howard	Penny (71.4%)	Leonard (47.6%)
Leonard	Penny (95.2%)	Amy (57.1%)
Penny	Leonard (95.2%)	Bernadette (57.1%)
Raj	Sheldon (66.7%)	Leonard (28.6%)
Sheldon	Penny (85.7%)	Howard, Leonard (47.6%)

who confused the two characters. It turns out that certain phrases intended for Penny were perceived as "arrogant" when spoken by Sheldon. Here are the actual comments by participants:

- "mhm..." I can picture coming from Sheldon in an irritated manner.
 "...you are kidding, right" would be said by Sheldon in an arrogant and condescending manner.
 - "You might be interested in knowing..." sounds like an arrogant Sheldon line, followed by the "Oh God..." I can actually picture Sheldon saying this line.
 - "You might be interested in knowing..." is used twice in Dialogue 2, and would be something Sheldon might say to make another person feel inferior.

7.2.2 Perception of Sheldon in comparison to Leonard (least distinguishable)

As roommates and colleague at work, their similarity is understandable. Here are summarized comments by participants describing the dialogue:

- matter-of-fact, straightforward
- clear, unhesitant
- shorter, more direct sentences; to the point
- use "technically"
- do not use a long string of adjectives

7.3 Leonard

For Leonard, Penny is considered the most distinguishable. Even though Leonard is considered less nerdy than other male characters, his language is still very different from Penny's.

Amy being the least distinguishable for Leonard is also believable. Amy, despite her language closely resembles Sheldon, is also interested in relationships and friendship (e.g., with Penny and Bernadette).

Here are some participants' comments on perceiving Amy's dialogue as Leonard's dialogue:

- intelligently spoken but also have a natural tone to them
- quick and to the point without over complicating things
- "I mean..." sounds like Leonard in his somewhat whiny manner
- Leonard sometimes smooths things over for Sheldon so he doesn't get upset. I think he would soften some things he says when he uses "I think" or "I mean"
- intelligent yet normal way of speaking
- both dialogue work okay really

7.4 Other Observations

Leonard and Penny represent the opposite-attracts couple. The biggest differentiating factor is that Penny's dialogue are perceived as being more emotional than Leonard's.

A general theme for Leonard's dialogue is that his speech pattern is "normal", implying that everyone else has a more stylized dialogue. This is an interesting observation because Leonard is not "normal" relative to the general population; he is being characterized as a typical nerd. Yet he is "normal" relative to his friends and therefore easier to identify on many cases.

According to (Brooks and Hébert, 2006), individual's social identities are largely shaped by the popular media: what it means to be white, black, male, female, heterosexual, homosexual, etc. Since characters are expressed through language and therefore connected to characters' identity as an individual and as part of a community (Hurst, 2011), the media such as television often provides the first (and sometimes the only) impression of certain groups of people.

In the context of BBT and the significant features we used to represent characters, it seems that

Penny's language represents the typical female as identified by Lackoff (Lakoff, 1973): hedging, emotional emphasis, adjectives, etc. This is in contrast with the male characters as scientists, who tend to be more matter-of-fact.

Do scientists talk differently from the general population? Our results answer with a "yes" in that Penny's language is mostly in contrast with male scientists' language. Such contrast is also reflected in the real world (e.g., % of scientists versus U.S. population believe in climate change).

What makes the show interesting is the "in-between" characters: female scientists Amy and Bernadette. The perception of the dialogue showed that the Penny-Bernadette, and Leonard-Amy pairs shared some similar language. With the right intention and scripts, the media can help narrow the perception and narrative gap between scientists and the general public.

8 Conclusion and Future Work

We explored character voice from the TV show BBT by building stylistic models relating character dialogue's linguistic features to natural language generation decisions. These models are then used to manipulate an expressive NLG to transform regular sentences into an expressive version. The generated, expressive dialogue are then used in a perceptual experiment to see how users perceive expressed personalities. Our results were encouraging in that people were able to perceive differences among characters, though some better than others. For the ones that were hard to distinguish, participants' comments provided great insight into how to better express the extracted features through NLG.

One possible future work is to use people's blogs as a source to create speaker-specific models. Another possible future work is to use character models to drive the monologue-to-dialogue process that created the stories used in our experiment. For example, if the character sounds mostly negative, the process can try to allocate all negative sentences to a story character's dialogue.

We believe our work can be applied to storytelling applications, such as video games, interactive narrative, chatbots, or education systems where dialogue with personalities may improve user experience, in a more controllable way (than using a neural network for generation, for example).

References

- Monika Bednarek. 2011a. The language of fictional television: a case study of the dramedy gilmore girls. *English Text Construction* 4(1):54–83.
- Monika Bednarek. 2011b. The stability of the television character: A corpus stylistic case study. *R. Pizzi, M.*
- Monika Bednarek. 2012. Constructing "nerdiness": Characterisation in the big bang theory. *Multilingua* 31(2-3):199–229.
- Kevin K Bowden, Grace I Lin, Lena I Reed, Jean E Fox Tree, and Marilyn A Walker. 2016. M2d: Monolog to dialog generation for conversational story telling. In *Interactive Storytelling: 9th International Conference on Interactive Digital Storytelling, ICIDS 2016, Los Angeles, CA, USA, November 15–18, 2016, Proceedings 9*. Springer, pages 12–24.
- Dwight E Brooks and Lisa P Hébert. 2006. Gender, race, and media representation. *Handbook of gender and communication* 16:297–317.
- Claudia Bubel. 2005. *The linguistic construction of character relations in TV drama: Doing friendship in Sex and the City*. Ph.D. thesis, Universität des Saarlandes.
- D.K. Elson and K.R. McKeown. 2009. A tool for deep semantic encoding of narrative texts. In *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*. Association for Computational Linguistics, pages 9–12.
- Andrew S Gordon, Qun Cao, and Reid Swanson. 2007. Automated story capture from internet weblogs. In *Proceedings of the 4th international conference on Knowledge capture*. ACM, pages 167–168.
- M Hurst. 2011. *Language, Gender, and Community in Late Twentieth-century Fiction: American Voices and American Identities*. Springer.
- Robin Lakoff. 1973. Language and woman's place. *Language in society* 2(1):45–79.
- Benoit Lavoie and Owen Rambow. 1997. A fast and portable realizer for text generation systems. In *Proceedings of the fifth conference on Applied natural language processing*. Association for Computational Linguistics, pages 265–268.
- Boyang Li, Mohini Thakkar, Yijie Wang, and Mark O Riedl. 2014. Data-driven alibi story telling for social believability. In *Proceedings of the FDG 2014 Social Believability in Games Workshop*. Citeseer.
- 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*. volume 11.
- François Mairesse and Marilyn Walker. 2007. Personage: Personality generation for dialogue. In *Annual Meeting-Association For Computational Linguistics*. volume 45, page 496.
- James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates* 71:2001.
- Paulo Quaglio. 2009. *Television Dialogue: The sitcom Friends vs. natural conversation*. John Benjamins Publishing Company.
- Elena Rishes, Stephanie M Lukin, David K Elson, and Marilyn A Walker. 2013. Generating different story tellings from semantic representations of narrative. In *Interactive Storytelling*, Springer, pages 192–204.
- Susan Schreibman, Ray Siemens, and John Unsworth. 2008. *A companion to digital humanities*. John Wiley & Sons.
- Efstathios Stamatatos. 2009. A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology* 60(3):538–556.
- 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. *International Conference on Interactive Digital Storytelling*.