

Using Referring Expression Generation to Model Literary Style

| | | | |
|---|---|--|--|
| Nick Montfort Massachusetts Institute of Technology Cambridge, MA, USA nickm@nickm.com | Ardalan SadeghiKivi Massachusetts Institute of Technology Cambridge, MA, USA ardalan@mit.edu | Joanne Yuan Massachusetts Institute of Technology Cambridge, MA, USA joanneyu@mit.edu | Alan Y. Zhu Massachusetts Institute of Technology Cambridge, MA, USA alanyzhu@mit.edu |
|---|---|--|--|

Abstract

Novels and short stories are not just remarkable because of what events they represent. The narrative style they employ is significant. To understand the specific contributions of different aspects of this style, it is possible to create limited symbolic models of narrating that hold almost all of the narrative discourse constant while varying a single aspect. In this paper we use a new implementation of a system for narrative discourse generation, *Curveship*, to change how existents at the story level are named. This by itself allows for the telling of the same underlying story in ways that evoke, for instance, a fabular or parable-like mode, the style of narrator Patrick Bateman in Brett Easton Ellis’s *American Psycho*, and the unusual dialect of Anthony Burgess’s *A Clockwork Orange*.

1 Introduction

It is well-known in narrative theory (narratology) that narratives achieve their power not only because of *what* they represent but also because of *how* they represent it. Narrative theorists generally agree that there is utility in at least conceptualizing a story level (the “content”) separate from the discourse level (the “expression”).

With awareness of this separate narrative discourse, we can ask how the expression particularly contributes to the effect of a novel or short story. For instance, Herman Melville’s *Moby-Dick* is a narrative of personal experience told by Ishmael. Imagine that all the events of the novel remained the same, it still focalized Ishmael, but the narration was like that of Hemingway’s *The Old Man and the Sea*. Ishmael would no longer be the “I” of the story and the book would no longer be told in his characteristic voice or enunciation. The narrator would be rather covert and unmarked. How would the effect be different? Or, what if Ishmael remained the “I” and spoke as he does in the novel,

but was not on the Pequod — another character who had all the same experiences took his place on the whaling voyage — and so the narrator had his distinctive voice, but without being a character in the diegesis?

Researchers have investigated individual novels using statistical methods that have more typically been applied to large corpora of literature (Clement, 2008; Kelleher and Keane, 2017; Wang and Iyyer, 2019). Hypothetical and speculative questions of the sort we pose above, about literary style and the style of individual novels, can be addressed in a different way, by modeling narrative style computationally. Generating variations of a large classic novel is not a reasonable goal. By creating small-scale narratives, however, more akin to folktales or conversational stories but in different literary styles and in fine-grained genres, we can try out different ways of telling the same underlying story to see what literary effects can be produced using a compact model.

Building on our earlier work generating different narrative styles, in this study we focus on one type of variation, corresponding to a classic area of natural language generation research, referring expression generation (REG). That is, we look at how the narrator names existents (those entities that exist) in the story, corresponding to characters, places, and other things or objects. Our question is, when we focus on the naming of existents, allowing few other variations, how much influence can we exert over literary style? Which literary styles can we computationally model in this way?

Unlike almost all work on literary study done with computational linguistics approaches, ours does not involve automatic feature extraction, stylistics, or the annotation of a corpus. We devise our symbolic models of narrative style “manually” by specifying the possible vocabulary that can be used and high-level narrative parameters.

2 A Style Generator, Not a Story Generator

A story or plot generator is a computational procedure, a set of instructions that, when applied to given inputs, produces an output considered a story. These generators work at the level of content (the story level). Curveship is focused on the narrative discourse and is *not* a story or story-level generator.

An important early story generation project is TALE-SPIN (Meehan, 1997), operating on story data similar to that in (Klein et al., 1973) to produce animal stories. It simulates reasoning and behavior in a virtual world, using planning to find how the characters can accomplish (or fail at) their goals. Along these lines, TV melodrama system UNIVERSE (Lebowitz, 1985) has a character creation cycle and simulates each personal life. MINTREL (Turner, 1994) tells stories about King Arthur's court, operating at two different levels of author and actors goals, using a planning stage and a problem-solving stage. Many earlier projects along with FABULIST (Riedl and Young, 2010) refer to an explicit model of narrative or, in the case of MEXICA (Pérez y Pérez, 1999), one of the creative writing process.

These related projects do not focus on the generation of the *narrative discourse*, as Curveship does. Because both levels are important to narrative, two collaborations have already been done to integrate the system with MEXICA (Montfort, 2009; Montfort et al., 2013).

A system operating at the expression level is PAULINE (Hovy, 1988), generating appropriate texts from a single representation and accounting for the hearer's sensitivity and contexts of communication. Along similar lines, but working in creative and narrative contexts, are systems for dialog generation, including (Hämäläinen and Alnajjar, 2019). Curveship differs from both in that it is specialized to deal with narration, not pragmatics or dialog. Of recent work, that on style transfer, e.g. (Reif et al., 2021), is related, although in this case we see a diametrically opposite approach of using large, opaque language models and not even having a "source text" in a particular style. While a provocative direction and one we hope could inform narrative studies, our approach is to use a highly parsimonious model in which all narrative parameters are explicit. Our goal is the study and human understanding of style, rather than the efficient and general automated transfer of it.

3 Stylistic Variation in Literature and Narratology

Narrative and other stylistic variation has been explored for centuries without computing, long before narratology. Erasmus's 1512 *Copia*, in Latin, systematically shows how to embellish, expand, and vary speech and writing, explaining that expression as well as subject matter should be "abundant" (Erasmus, 2007). His variations include syntactical ones along with the use of synonyms, heterosis, enallage, and other figures of speech.

A creative work more specific to narrative is *Exercises in Style*, first published in French in 1947 (Queneau, 1981). It narrates the same uninteresting events 99 times. Only the style engages the reader, who finds a quotidian story told as a book blurb, an entire three-act play, a haiku, etc. Each section has a practical and frank title, e.g., "Notation," "Official Letter," "Comedy." Some variations are specific to narrative, as with "Retrograde," in which events are represented in reverse chronological order. There are others that are not, including lexical variations in which the story is narrated in anagrams and permutations along with variations in which the text is dismantled and listed by the different parts of speech.

Queneau's book inspired many others, including the graphic novel *99 Ways to Tell a Story* (Madden, 2006). Although the particular banal sequence of events is different, the framework is the same, exploring the possibilities of the comic medium rather than text. The visual narratives include a map, a how-to-draw process, a public service announcement, an advertisement, and an image in the style of the Bayeux tapestry. Some are in the style of famous comic artists.

More theoretical attempts to systematically understand how the narrative discourse can vary, independent of the story or content, have been expressed in the *narratologies* or specific narrative theories of different scholars since the 1960s, e.g., (Bal, 1985; Chatman, 1980; Rimmon-Kenan, 1983) and new editions thereof. The system we have developed and use in this study is based mainly on (Genette, 1979) as revised in (Genette, 1988), with awareness of other aspects of narrative pointed out in (Prince, 1982) and other sources.

Narratology began with the the novel and drew from linguistic ideas but is now clearly transmedial (Ryan et al., 2004) and, in dealing with all sorts of representations of events, is not limited to literary

or linguistic sorts of study. Nevertheless, if it is considered in a linguistic framework it must be acknowledged to operate at the discourse level, where for instance intersentential reference is a concern.

4 Relevant Work on Referring Expression Generation

In a narrative framework, and focusing on text, we consider a referring expression to be any noun phrase or substitute for an NP that designates an existent. Our definition, limited to narrative, is consistent with standard ones, presented when the task is to refer to objects in a blocks world (Winoograd, 1972) or in photographs of the environment (Zarriß and Schlangen, 2016). Our research does not involve such representations, but our system nevertheless needs to convert non-linguistic information (the story-level “content”) to natural language, and in this regard (Krahmer and van Deemter, 2012) it undertakes this REG task.

In the context of our work so far on very short stories, we are not mainly concerned with REG that selects from available attributes to produce NPs with appropriate modifiers, a common goal of REG research. Instead, we are concerned with adherence to systematic naming conventions of different sorts.

We use REG in a contained and tractable framework in which existents and events are symbolically modeled and the narrative style is, similarly, completely and formally defined. Our approach is not corpus-based and our system, rather than being trained on any data, uses human-authored expressions, selected and combined computationally. Our project makes use of very established REG techniques to investigate literary style.

5 Our Platform for Narrative Modeling

We use a system developed originally in Python to implement parser-based interactive fiction (Montfort, 2007, 2009), now in a new JavaScript formulation. The current Curveship-js allows for teaching and creative work as well as research. Students have used it to learn about narrative theory concepts. Although finished creative works have not yet been published using the system, it has seen some use by writers and artists.

Curveship-js allows for a wide variety of changes to be automatically made at the level of narrative discourse, including changes to the order of events in the telling, ellipsis, focalization, time of narrating, and the “I” and “you” of the narrative

(the clearest signs of narrator and narratee). Using a simple underlying representation, Curveship-js generates text and determines grammatical specifics according to high-level narrative parameters.

To pursue this current project, we implemented a clean split between story (the “content”) and narrator, with each level modeled in a different files and with the ability to associate arbitrary numbers of narrators with the same story. As much linguistic information as possible has been eradicated from the story-level representation, so that now narrators not only have different global or general parameters but also each have their own names for existents and their own verb phrases associated with events.

6 Referring Expressions and Literary Style

Of many compelling literary styles and methods of naming existents, we have identified a few for initial modeling. We test our idea that these styles are distinctive because of how referring expressions are used. We have selected literary styles which present different sorts of challenges. Some have only subtle differences. Others use very marked and elaborate styles of naming.

References to characters using an NP are in bold. Similar references to other existents are in *Italic*. All pronominal references remain in Roman.

6.1 Parable or fable

Uses very generic names, identifying individuals by at most profession or with a single descriptive adjective, sometimes even less.

Again, the *kingdom of heaven* is like unto *treasure* hid in a *field*; the which when a **man** hath found, he hideth, and for joy thereof goeth and selleth all that he hath, and buyeth *that field*. (Matthew 13:44, KJV)

The fox who longed for *grapes*, beholds with pain / The tempting *clusters* were too high to gain; (Dudley, 1970)

Go and visit **grandmother**, who has been sick. Take her *the oatcakes* I’ve baked for her on *the hearthstone* and a little *pot of butter*. **The good child** does as **her mother** bids ... (Carter, 1979)

Note that no characters at all are named in Carter’s “fable.” It is based on “Little Red Rid-

ing Hood,” which fits, in terms of genre, in 6.3, but is in the *style* described here in 6.1.

6.2 Literary simplicity

Evoking the parable or fable, some literary writing has extremely straightforward syntax and uses simple, generic expressions for characters, mentioning proper names only rarely.

He was **an old man** who fished alone in a *skiff* in the *Gulf Stream* and he had gone eighty-four days now without taking a fish. In the first forty days **a boy** had been with him. (Hemingway, 2003)

A syntactically simple run-on sentence here is characteristic of Hemingway’s style. While the old man’s name, Santiago, is mentioned in this novel, almost all references to the character are “the old man” or use a pronoun.

The boy leaned on *the cart* and adjusted *the wheel*. What do you see? **the man** said. Nothing. He lowered *the glasses*. It’s raining. Yes, **the man** said. I know. (McCarthy:2006)

Unmarked direct discourse (speech that is not surrounded by quotation marks) is seen in this extract, and is characteristic of McCarthy’s style. Again, of the very syntactically simple sentences, one is a run-on with two main clauses joined by “and.” The duo encounter a named character, Ely, but neither of the main characters’ names are mentioned. The differences in style from 6.1 Parable or fable here are subtle. That style actually often includes more complex syntax, however, and completely eschews proper names.

6.3 Folktale or fairy tale

Oral versions of these were significantly different, but we are considering literary style. The main character is often identified by a short sort of proper name or nickname, not including a family name: Cinderella, Snow White, Rapunzel, Little Red Riding Hood. This main character’s name is of course also the name by which the story is known. Subsidiary characters are often named more generically according to their roles: The prince, the Big Bad Wolf. Literary works that draw on folklore sometimes use similar naming conventions alongside metafictional (self-conscious and self-referential) writing techniques.

All on *one summer’s day*, **the King of Hearts** calls for *the tarts* baked for him by **his Queen**, only to find they have been stolen. All fingers point to his former page, the **Knave of Hearts** ... (Coover, 2005)

6.4 Initials only for some existents

At various points in literary history authors have chosen to indicate existents (characters, villages, etc.) using only their first initials, sometimes with a long dash, period, or other special typographical intervention after them. It is possible this allows a narrative to be read in a way that is more general and universal; it may also suggest that the narrative is about real-world existents the narrator is choosing to anonymize.

In my return from Italy I brought him with me to the country in whose language he had learn’d his notes—and telling the story of him to **Lord A—Lord A** begg’d *the bird* of me—in a week **Lord A** gave him to **Lord B—Lord B** made a present of him to **Lord C**—and **Lord C**’s gentleman sold him to **Lord D** for a shilling—**Lord D** gave him to **Lord E**—and so on—half round the alphabet... (Sterne, 1986)

In this, I am bound to say, **Mr. A.** was but sustaining the tradition conceived originally by his predecessor, **Mr. P.**, a Harvard man, who until his departure from *Vingt-et-Un* succeeded in making life absolutely miserable for **B.** and myself. (Cummings, 1949)

I am a well-known folklorist, an authority on the *A—s*, a tribe I have no intention of disgracing by my interest. (Cohen, 1993)

6.5 The novel of manners

While literary writing is often concerned with characters’ place in society, this is magnified in the novel of manners. Proper names are used with courtesy titles, which are repeated, and referring expressions otherwise relate to social status.

Not all that **Mrs. Bennet**, however, with the assistance of her **five daughters**, could ask on the subject, was sufficient to draw from **her husband** any

satisfactory description of **Mr. Bingley**. They attacked him in various ways; with barefaced questions, ingenious suppositions, and distant surmises; but he eluded the skill of them all, and they were at last obliged to accept the second-hand intelligence of **their neighbour Lady Lucas**. Her report was highly favourable. **Sir William** had been delighted with him. (Austen, 2008)

6.6 Brand names for objects

An obsessive use of brand names and mention of the places where items were purchased can indicate a consumer fixation or even psychopathology.

He continues talking as he opens *his new Tumi calfskin attaché case he bought at D. F. Sanders*. He places *the Walkman* in the case alongside *a Panasonic wallet-size cordless portable folding Easa-phone* (he used to own *the NEC 9000 Porta portable*) and pulls out *today's newspaper*. (Ellis, 1991)

6.7 Dialect and idiolect

Narration is sometimes done using idiosyncratic lexical or vocabulary choices. These can include misspellings and generally infelicitous word usage. An invented dialect can also be employed, perhaps to project science-fictional worlds with an uncanny relationship to our own.

I took *the large moloko plus* to one of *the little cubies* that were all around *this mesto*, there being like *curtains* to shut them off from *the main mesto* ... and then there were colours like no one had ever viddied before, and then I could viddy like *a group of statues* a long long way off ... (Burgess, 1986)

7 The Example Story and Corresponding Narrative Specifications

A very short story, “First Class,” was modeled at the content level in Curveship-js. While “First Class” is an original story, it was inspired by part of the 15th track of the album *Black on Both Sides* by Yasiin Bay (previously known as Mos Def), which describes a racial microaggression not present in our story. We encourage those consulting this research to listen to this song.

The underlying story model has no linguistic information attached to it that is used in realization. The existents and events do have tags, for internal use: A Curveship-js author or researcher uses these to connect the story-level existents and events to discourse-level names and verb phrases. In “First Class,” the existents are two Places (gate and first-Class), four Actors (the class for modeling characters; these are celebrity, gateOfficial, flightAttendant, and passenger), and eleven Things including boardingPass, scanner, seat1A, seat1B, notepad, pen, and some items of clothing.

There are also sixteen (16) events. The story file that includes these existent and event representations, excerpted in figure 1, is concise (less than 3KB).

To produce different specific narratives, a narrator file is used for each. Although Curveship-js has many capabilities for varying the narrative, we have limited our use of the system so that we strongly emphasize only one: the facility for different naming or generation of referring expressions. To generate a few of the final results, we also vary the verb phrases used to produce representations of events, and we also make other minor changes in time of telling, ellipsis, focalization, and the like.

One does not need to define all names for existents in a particular narrator file. If an entity is a character (Actor) who is female and a child, for instance, Curveship-js will automatically generate “a girl” as a name for the character on first reference, and “the girl” thereafter, by instantiating the GenericName class. Or, using the Names class, one can define a simple, short name with an initial NP to be used and (optionally) an NP to be used on subsequent reference. There are extensions of this to allow for Actors with first and last names and courtesy titles (ProperNames) and for Things that are indicated by trade name and their place of purchase (BrandNames).

The system is a free/libre/open-source software system, so others may use it to reproduce our results or as the basis for their own research or even creative projects. (Montfort and contributors, 2021)

8 Generated Results

The following are all outputs from the system, each a different version of “First Class” that attempts to imitate a narrative style in section 6.

```

actor.celebrity = new Actor(place.gate, "male");
actor.gateOfficial = new Actor(place.gate, "female"); ...
thing.boardingPass = new Thing(actor.celebrity);
thing.boardingPass.owner = actor.celebrity;
thing.scanner = new Thing(place.gate);
thing.jacket = new Thing(actor.celebrity); ...
// EVENTS
ev.scan = new Event(actor.celebrity, thing.boardingPass, temporal.against, thing.scanner);
ev.beep = new Event(thing.scanner);
ev.gasp = new Event(actor.gateOfficial);
ev.pocket = new Event(actor.celebrity, thing.boardingPass, temporal.in, thing.jacket);
ev.pocket.alternates(thing.boardingPass, "location", actor.celebrity, thing.jacket);

```

Figure 1: An excerpt from the story file for “First Class.”

8.1 Parable or fable

A man walked from the gate. He went to the first class cabin. He sat in a wide seat. A woman was surprised. A flight attendant noticed the man’s sneakers. He said “sir, I believe you’re in the wrong cabin” to the man. The man took his boarding pass from his jacket. He showed his boarding pass to the flight attendant. The flight attendant apologized. He said “I shouldn’t ask, but ... my daughter would really love to have your autograph” to the man.

8.2 Literary simplicity

The scanner beeped. The airline employee gasped. Harry had held his boarding pass against the scanner. He departed the gate. He walked to the first class seating area. He sat in a seat in the first row. A woman expressed a bit of surprise. A flight attendant noticed the celebrity’s sneakers. He said “sir, I believe you’re in the wrong cabin.” The celebrity took his boarding pass from his jacket. He showed his boarding pass to the flight attendant. The flight attendant muttered “oh, I’m sorry.” He procured a pen and a notepad. He said “I shouldn’t ask, but ... my daughter would really love to have your autograph.”

8.3 Folktale or fairy tale

One day... The Frog Prince held his boarding pass against a panel on the airport turnstile. It beeped. A woman gasped. The prince placed his boarding pass in his jacket. He walked from the gate. He entered to the first class cabin. He sat in a comfortable seat. A woman puckered up involuntarily. A flight attendant noticed the prince’s sneakers. He said “sir, I believe you’re in the wrong cabin.” The prince took his boarding pass from his jacket. He

showed his boarding pass to the flight attendant. The flight attendant apologized by saying “oh, I’m sorry.” He obtained a pen and a notepad. He said “I shouldn’t ask, but ... my daughter would really love to have your autograph.” The end.

8.4 Initials only for some existents

Mr. S——— holds his boarding pass against a scanner. The scanner beeps. A lady working for the airline gasps. Mr. S——— places his boarding pass in his sportscoat. He departs the gate. He walks to first class. He sits in seat 1B. A young female executive reacts. A flight attendant notices Mr. S———’s kicks. He sneers “sir, I believe you’re in the wrong cabin” to Mr. S———. Mr. S——— gets his boarding pass from his sportscoat. He shows his boarding pass to the flight attendant. The flight attendant mutters “oh, I’m sorry” to him. He grabs a pen and a notepad. He says “I shouldn’t ask, but ... my daughter would really love to have your autograph” to Mr. S———.

8.5 The novel of manners

An airline employee gasped. Sir Harry Styles had held his boarding pass against a scanner. He placed his boarding pass in his sportscoat. He departed the turnstile. He walked to the first class cabin. He sat in seat 1B. Ms. Carly Fiorina reacted. A flight attendant noticed Sir Styles’s casual shoes. He sneered “sir, I believe you’re in the wrong cabin” to Sir Styles. Sir Styles got his boarding pass from his sportscoat. He showed his boarding pass to the attendant. The attendant muttered “oh, I’m sorry” to him. He grabbed a pen and a notepad. He said “I shouldn’t ask, but ... my daughter would really love to have your autograph” to Sir Styles.

```

names.gate = new Names("the gate");
names.firstClass = new Names("the first class cabin"); ...
names.seat1A = new Names("a wide seat"); ...
vp.depart = new VerbPh("walk from");
vp.board = new VerbPh("go");
vp.sit = new VerbPh("sit");
vp.beSurprised = new VerbPh("is surprised");

```

Figure 2: An excerpt from the “parable teller” narrator file for “First Class.”

8.6 Brand names for objects

The type of guy who can get a reservation at Le Bernardin walks to first class. He sits in seat 1B. I notice coolly. A male flight attendant glances at the famous guy’s Air Jordan 4 Retro Kaws purchased from Flight Club. He sneers “sir, I believe you’re in the wrong cabin” to the famous guy. The famous guy gets his boarding pass from his bespoke Michael Andrews sportscoat. He shows his boarding pass to the male flight attendant. The male flight attendant mumbles “oh, I’m sorry” to him. He pulls out a BIC pen from K-Mart on Astor Place and a Mead memo pad bought at Key Foods. He says “I shouldn’t ask, but ... my daughter would really love to have your autograph.”

8.7 Dialect and idiolect

This uniformed devotchka gasped. Sir Harry Styles had held his boarding pass against a scanner. He placed his boarding pass in his carman. He walked to the first class cabin. He sat in seat 1B. This forella reacted. A veck viddied Sir Styles’s sabogs. He sneered “sir, I believe you’re in the wrong cabin” to Sir Styles. Sir Styles got his boarding pass from his carman. He showed his boarding pass to the veck. The veck muttered “oh, I’m sorry” to him. He grabbed a pen and a notepad. He skazated “I shouldn’t ask, but ... my daughter would really love to have your autograph” to Sir Styles.

9 Discussion

We refer to styles presented in section 6 and modeled in section 8 as 1–7. While results varied, we attempted to model these seven types of narrative style seriously and the process has provided us with some insights into what aspects of linguistic and narrative representation are most important to each sort of style. In other words, the system has helped us think about style in ways that complement other inquiries.

This method has allowed us to see how some

styles are more easily imitated, or at least signaled, by varying the way referring expressions are generated, while others (if they are in fact best considered distinct styles) will require different sorts of intervention. Perhaps we have been able to discern that some styles are simply more straightforward while others — forgive us — are more hairy.

Rather than discuss these in seven sections, we offer a synthetic and comparative discussion of the ways narrative style has been modeled and the text that resulted.

There are obviously many aspects of style we have not yet computationally modeled. We tried to avoid translated texts, but many canonical examples of 1 Parable or fable are in translation. The first quotation is of very archaic English (translated from Koine Greek) and the second in verse — this time in translation from Ancient Greek. Carter’s fable, on the other hand, is in prose and in contemporary English. To present work in the style of parables from the King James Bible or from 17th Century translations of Aesop’s Fables, finer distinctions must be made.

Although our focus has been on referring expressions, we found it necessary to use ellipsis, change the order of events in the telling, change the time of speaking, and in one case (8.6) focalize a particular character in order to do a reasonable job of modeling style. We also changed the way that events were represented by having our narrators use different verb phrases, not just their own names for Existents. This was particularly important in 8.7, where the dialect includes unusual verbs.

We found that we were able to do reasonable work matching only three styles when strictly chronological narration was used, even in the limited context of this simple underlying story. The ones that seemed apt when a chronological ordering was used were 1 Parable or fable, 3 Folktale or fairy tale, and 6 Brand names for objects. The first two styles are associated with orature and with

```

names.firstClass = new Names("the first class cabin"); ...
names.gateOfficial = new Names("this uniformed devotchka");
names.celebrity = new ProperNames("Harry", "Styles", pronoun.masculine, "a celebrity", "Sir"); ...
names.boardingPass = new Names("a boarding pass");
names.scanner = new Names("a scanner"); ...
vp.scan = new VerbPh("hold");
vp.gasp = new VerbPh("gasp");
vp.pocket = new VerbPh("place");
vp.board = new VerbPh("walk");

```

Figure 3: An excerpt from the “Alex Delarge” (narrator of *A Clockwork Orange*) narrator file for “First Class.”

fairly simple and short stories, so it is not surprising that they can be told in chronologically straightforward ways and have their style remain identifiable. The last is perhaps most interesting, as it involves complex naming and concern with not only branding but also purchase history. That the narration is flat and direct in other aspects may help to highlight the narrator’s fascination, or indeed obsession, with consumerism.

6 Brand names for objects was also the only style in which it seemed important to strictly focalize a single character, the woman passenger on the plane whose seat is next to the celebrity’s. Although **7** Dialect and idiolect is also associated with narratives of personal experience, the unusual enunciation of the diegetic narrator is so distinctive that it seems admissible to have the narration include events that this narrator may not have seen.

While **2** presents a style similar to that of Hemingway in *The Old Man and the Sea* or McCarthy in *The Road*, there are some noticeable differences. Simply in terms of referring expressions, it would be an improvement for the main character’s name to be mentioned later in the discourse. Typographically, McCarthy employs unmarked direct discourse, foregoing any quotation marks. Hemingway uses quotation marks to indicate speech, but presents the main character’s thoughts in free indirect discourse. While this does not end up confusing readers, it can mean that when beginning to read a sentence, it is not immediately clear whether the words are being narrated or are spoken by a character, adding some interest and complexity to the reading process. Adding this capability to our system would of course be useful.

In both of those “simple” books, even restricting our examination to representations of action rather than description or exposition, we find long sentences that are beyond what our system can

currently generate, for instance: “He settled comfortably against the wood and took his suffering as it came and the fish swam steadily and the boat moved slowly through the dark water.” (Hemingway, 2003) This is a run-on sentence with four parts, three of which would be straightforward to model using Curveship-js. It would be an improvement to be able to produce run-on sentences of this sort to create a flow or press of several actions that are not punctuated.

In **3** Folktale or fairy tale, a story “preface” and “postface” of one fixed sentence each was significant to signaling the style or genre. The tone that was generated is even more outrageous than in many of Robert Coover’s stories in this style, because the Frog Prince immediately is involved in the quotidian, contemporary bureaucracy of having his boarding pass scanned. The moral of this seems to be that this “style” cannot always be generated at the level of narrative discourse, as it sometimes relates to the underlying content.

In **4** Initials only for some existents, the effect of this one stylistic change seemed most obscure or oblique. This may be because of the very wide range of contexts in which this sort of naming has been used. Our generated text reads like the production of a contemporary writer who might be trying to imitate a style from centuries ago, but without really understanding the relevant aspects of any particular style. Perhaps this is no surprise, as our “initials only” examples in 6.4 were from the 18th Century, the early 20th Century, and the late 20th Century. It seems that it is simply not enough, or may end up appearing to be an affectation, to alter only this way of making reference to existents.

There are certainly some limitations to the way the style of the novel of manners **5** is modeled by the system. The use of courtesy titles and proper names for supposedly important characters, while

others of lower class are referred to by role, does make this text consonant with *Pride and Prejudice* and similar works of fiction. We miss, however, certain aspects of framing and the sort of abstract declarations that Austen makes to inform the reader about the social world. These elements are not as overt in **1**, **2**, and **3**, for instance.

In **7** Dialect and idiolect, the incorporation of distinctive nouns and verbs certainly signals the style we are hoping to imitate — anyone familiar with *A Clockwork Orange* will be unable to avoid noticing that this narrative is in the style of the novel. While it uses these distinctive words to make a connection, there are nevertheless noticeable failings in the way text is generated. Alex interjects “like” as well as phrases such as “O my brothers and only friends” throughout the novel, interjections which seem to us important to the style and conspicuously missing from the generated text. He narrates what he is feeling and thinking; It is important to the style that the narrative is internally focalized. He negatively evaluates certain events and positively appraises others, for instance as “real horrorshow.”

Our system does not currently have the ability to paraphrase utterances in direct discourse. Although parables and fables sometimes include direct discourse, it may be better paraphrased in 8.1 and probably in 8.3. Paraphrase would give the opportunity for additional dialectical narration in 8.8. Generally, the extent to which speech is directly quoted or is paraphrased seems important to literary styles such as these.

We believe 8.6 and 8.7 are most remarkable and easily identifiable as examples of particular literary styles. Rather than imagining that we were particularly good at modeling styles **6** and **7**, we take this as evidence that the original literary styles are so distinctive that fairly simple gestures toward them can indicate them.

We have omitted some uses of referring expressions that we have observed in literary work. For instance, we know of one case in a very short literary story in which only pronouns are used to refer to characters (Eason, 1992). Generally, better control over pronominalization (improving our current, primitive algorithm) would improve our ability to model styles, because some styles, as seen in **2**, are extremely spare, even to the extent of leaving some initial ambiguity in reference. For instance, to aid in modeling styles similar to **2** we would like to be able to introduce a character using a generic

name or even a pronoun and, via cataphora, give the character’s proper name later.

10 Future Work

Rather than try to draw more specific conclusions, we have chosen to identify next steps that would be productive for us and others seeking to computationally model narrative style.

Part of our project of modeling narrative style involves testing to what extent it is sensible, following narrative theory, to consider the underlying content or storyworld separately from the expression or narrative discourse. Future work should involve models of each level being further ramified. If a unified model could be developed by others that is simpler and more powerful than our two-level model, this would argue against the fundamental model narrative theory posits.

Some questions to answer, then, are whether a very wide variety of styles can be parsimoniously modeled using a single underlying story representation. Of course, expanding the number of styles modeled will be one direction for future work. A very broad investigation could involve taking on the “Queneau challenge” and producing 99 narratives that parallel Queneau’s in *Exercises in Style*. As with our tentative research here, we would learn from this study which of the styles can be produced distinctively even with a limited model and which require more elaborate modeling.

An in-depth comparative analysis of a few important literary styles would also be an important direction for further research. Just as an attempt to address the “Queneau challenge” would encourage work on broad coverage, this effort would compel detailed study as the research digs deeply into a few important styles. We would need to generate lengthier stories in which reference is made to existents in different contexts.

Finally, other good evidence for the ability to usefully separate the content and expression levels would come from multi-lingual generation. We aim to generate narratives in different natural languages, initially, working with collaborators with appropriate expertise and adding the ability to generate in one other language. This will help us further develop a story/content representation that is not dependent on the particularities of expression. If we accomplish this, we should also be able to generate styles characteristic of different world literatures and gain insight into style across languages.

References

- Jane Austen. 2008. *Pride and prejudice*. Oxford University Press, Oxford, England.
- Mieke Bal. 1985. *Narratology: Introduction to the theory of narrative*. University of Toronto Press, Toronto.
- Anthony Burgess. 1986. *A clockwork orange*. Norton, New York.
- Angela Carter. 1979. The werewolf. In *The bloody chamber*. Harper & Row, New York.
- Seymour Chatman. 1980. *Story and discourse: Narrative structure in fiction and film*. Cornell University Press, Ithaca, New York.
- Tanya E Clement. 2008. ‘a thing not beginning and not ending’: using digital tools to distant-read gertrude stein’s the making of americans. *Literary and Linguistic Computing*, 23(3):361–381.
- Leonard Cohen. 1993. *Beautiful losers*. Vintage Books, New York.
- Robert Coover. 2005. Heart suit. In *A child again*. McSweeney’s, San Francisco, California.
- E. E. Cummings. 1949. *The enormous room*. The Modern Library, New York.
- Aesop; Aphra Behn; Francis Barlow; Thomas Philipot; Robert Codrington; Thomas Dudley. 1970. *Æ*. H. Hills for Francis Barlow, London.
- Bruce Eason. 1992. The appalachian trail. In *Flash fiction: 72 very short stories*. Turnstone Press, Winnipeg, Canada.
- Bret Easton Ellis. 1991. *American psycho*. Vintage Books, New York.
- Desiderius Erasmus. 2007. *On copia of words and ideas*. Marquette University Press, Milwaukee, Wisconsin.
- Ge rard Genette. 1979. *Narrative discourse*. Blackwell, Oxford.
- Ge rard Genette. 1988. *Narrative discourse revisited*. Cornell University Press, Ithaca, New York.
- Mika H m l inen and Khalid Alnajjar. 2019. Creative contextual dialog adaptation in an open world rpg. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, pages 1–7.
- Ernest Hemingway. 2003. *The old man and the sea*. Scribner, New York.
- Eduard H. Hovy. 1988. *Generating Natural Language under Pragmatic Constraints*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Conor Kelleher and Mark Keane. 2017. [Plotting Markson’s “mistress”](#). In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 33–39, Vancouver, Canada. Association for Computational Linguistics.
- Sheldon Klein, J. F. Aeschlimann, D. Balsiger, Steven L. Converse, Claudine Court, Mark Foster, Robin Lao, J. Oakley, and Joel Smith. 1973. *Automatic Novel Writing: A status report*. Computer Science Department, The University of Wisconsin, Madison, Wisconsin.
- Emiel Krahmer and Kees van Deemter. 2012. [Computational generation of referring expressions: A survey](#). *Computational Linguistics*, 38(1):173–218.
- Michael Lebowitz. 1985. *Story-telling as planning and learning*. Poetics, Volume 14.
- Matt Madden. 2006. *99 ways to tell a story*. Jonathan Cape, London.
- James R. Meehan. 1997. *Tale-Spin, an interactive program that writes stories*. The Fifth International Joint Conference on Artificial Intelligence at MIT, Cambridge, Massachusetts.
- Nick Montfort. 2007. *Generating Narrative Variation in Interactive Fiction*. Ph.D. thesis, University of Pennsylvania.
- Nick Montfort. 2009. [Curveship: An interactive fiction system for interactive narrating](#). In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, pages 55–62, Boulder, Colorado. Association for Computational Linguistics.
- Nick Montfort and contributors. 2021. [Curveship-js github repository](#). [Online; accessed 5-December-2021].
- Nick Montfort, Rafael P rez y P rez, D. Fox Harrell, and Andrew Campana. 2013. Slant: A blackboard system to generate plot, figuration, and narrative discourse aspects of stories. In *Proceedings of the International Conference on Computational Creativity (ICCC) 2013*, pages 168–175.
- Gerald Prince. 1982. *Narratology: The form and functioning of narrative*. Walter de Gruyter, Berlin.
- Rafael P rez y P rez. 1999. *MEXICA: A Computer Model of Creativity in Writing*. Ph.D. thesis, The University of Sussex.
- Raymond Queneau. 1981. *Exercises in style*. New Directions, New York.
- Emily Reif, Daphne Ippolito, Ann Yuan, Andy Coenen, Chris Callison-Burch, and Jason Wei. 2021. A recipe for arbitrary text style transfer with large language models. *arXiv preprint arXiv:2109.03910*.

- Mark O. Riedl and Michael R. Young. 2010. *Narrative Planning: Balancing Plot and Character*. *Journal of Artificial Intelligence Research*, Volume 39.
- Shlomith Rimmon-Kenan. 1983. *Narrative fiction: Contemporary poetics*. Methuen, London.
- Marie-Laure Ryan, James Ruppert, and John W. Bernet. 2004. *Narrative across media: The languages of storytelling*. University of Nebraska Press, Lincoln, Nebraska.
- Laurence Sterne. 1986. *A sentimental journey through France and Italy*. Harmondsworth; Penguin Books, Middlesex, England; New York.
- Scott R. Turner. 1994. *The Creative Process: A computer model of storytelling and creativity*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Shufan Wang and Mohit Iyyer. 2019. [Casting Light on Invisible Cities: Computationally Engaging with Literary Criticism](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1291–1297, Minneapolis, Minnesota. Association for Computational Linguistics.
- Terry Winograd. 1972. Understanding natural language. *Cognitive Psychology*, 3.
- Sina Zarrieß and David Schlangen. 2016. [Easy things first: Installments improve referring expression generation for objects in photographs](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 610–620, Berlin, Germany. Association for Computational Linguistics.