Grounding Characters and Places in Narrative Texts

Sandeep Soni and **Amanpreet Sihra**
University of California, Berkeley
sandeepsoni,amisihra@berkeley.edu

Elizabeth F. Evans
Wayne State University
e.f.evans@wayne.edu

Matthew Wilkens
Cornell University
wilkens@cornell.edu

David Bamman
University of California, Berkeley
dbamman@berkeley.edu

Abstract

Tracking characters and locations throughout a story can help improve the understanding of its plot structure. Prior research has analyzed characters and locations from text independently without grounding characters to their locations in narrative time. Here, we address this gap by proposing a new spatial relationship categorization task. The objective of the task is to assign a spatial relationship category for every character and location co-mention within a window of text, taking into consideration linguistic context, narrative tense, and temporal scope. To this end, we annotate spatial relationships in approximately 2500 book excerpts and train a model using contextual embeddings as features to predict these relationships. When applied to a set of books, this model allows us to test several hypotheses on mobility and domestic space, revealing that protagonists are more mobile than non-central characters and that women as characters tend to occupy more interior space than men. Overall, our work is the first step towards joint modeling and analysis of characters and places in narrative text.

1 Introduction

The association between characters and the places they navigate is central to a wide range of literary phenomena: *Bildungsromane* depict a character's journey across geographic space as a component of their psychological coming of age (Bakhtin, 1987; Jeffers, 2016); the *flâneur*, who walks and observes throughout a city, epitomizes the power that follows from peripatetic access to public spaces (Benjamin, 2002; Wolff, 1985; Wilson, 1992); class, gender, and racial associations can render surprising or scandalous a character's mere presence in an otherwise innocuous location.

While much work in literary history and theory has explored this interaction, it has remained out of reach for empirical observation and large-scale comparisons with social constructs such as gender,

Figure 1: In *Around the World in 80 Days*, Phileas Fogg is mentioned in conjunction with a wide range of places, but is only physically grounded in some. Disentangling the places he is *in* from the places he is *not* is crucial for tracking his movement throughout this work.

.

social status (Cresswell, 2012) and agency (Sen, 1993). In order to explore these questions empirically, we not only need to know the characters and places that exist within a narrative (Piper et al., 2021), but specifically how they interact: when is a character depicted as being *in* a place?

NLP research has made substantial progress in the individual components of this endeavor: under entity tagging, spans of text that correspond to entities are identified and categorized by their entity types (e.g., Bamman et al., 2019; Hamdi et al., 2021); methods can ground textual spans referencing a location to their real-world coordinates (e.g., Roller et al., 2012); and extensive schemas to precisely describe the relationship between locations have been proposed (e.g., Mani et al., 2010; Pustejovsky et al., 2011). Yet, despite this progress an important gap still remains: the technology supports the identification of characters and locations in text but falls short when it comes to placing the character with respect to a given location at any time in the story.
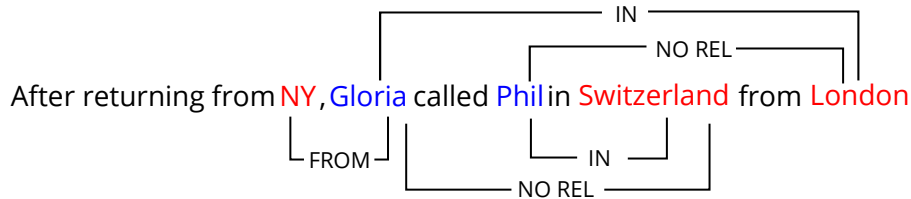
11723

Figure 2: Example sentence to illustrate the different relationships characters and locations can hold. Characters are marked in red whereas the locations are marked in blue. The objective is to label each pair of character and location co-mention. The definition of the spatial relationships can be found in § 2

.

In this work, we address this gap by proposing a classification task whose objective is to determine the spatial relationship between a candidate character and candidate location. The classification task helps make a judgment about the nature of the spatial relationship between the character and the location, allowing us to differentiate between scenarios in which the character is, for instance, at the location, approaching a location, has left a location, or has no relationship with the location in question. To illustrate the point even further, consider the toy example in Figure 2, in which multiple characters and locations are mentioned and their spatial relationships are annotated under the task we propose.

To make progress on this task, we annotate excerpts from books and use this annotated dataset to construct a predictive model. In applying this predictive model to a larger set of books, we test two hypotheses on mobility and domestic space, finding that protagonists are often depicted as being more mobile than other characters, and finding a strong gender effect on the kinds of spaces that are accessible: women as characters are more likely to occupy indoor or domestic spaces compared to men.

Overall, our contributions in this paper can be summarized below.

- We propose a new task to ground characters to locations in the story. The proposed task is an instance of a multi-class classification task with classes denoting the spatial relationship between the candidate character and location.

- We provide an in-depth annotation scheme for the different classes in the task and approximately 2500 annotated examples, which we openly release for public use.[1]

---

[1]The code and data for this paper can be found at `https://github.com/sandeepsoni/mobility-books`

- We operationalize the construct of mobility and spatial positioning with the help of a trained model. We use this to test macro-level hypotheses about the mobility and centrality of the characters, in the process corroborating known claims and providing quantitative evidence for previously unverified claims.

2 Task

Given a selection of narrative text containing a mention of a character and a mention of place, our overall task is to determine the nature of the relationship between that character and place at that instant in the story. Formally, consider a piece of text as a sequence of tokens $w_1, w_2, \ldots, w_n$. An entity tagger identifies C as the character mention spanning tokens w_c to w_{c+k}, where $1 \leq c \leq c + k \leq n$. Similarly, the entity tagger also identifies L to be the place mention spanning w_l to w_{l+m}, where $1 \leq l \leq l + m \leq n$. Both C and L are within 10 tokens of each other, *i.e.*, if $c + k > l$, then $c + k - l \leq 10$; otherwise, $l + m - c \leq 10$.

To operationalize the task further, we use entity definitions from Bamman et al. (2019). Characters are defined as entities of the type PER, which include instances that refer to a single person or a group. Places or locations are defined as entities that are natural locations (LOC) such as *the forest* or *Mars*, human-built structures (FAC) such as *the kitchen* or *the church*, and geo-political entities (GPE) such as *London* or *the village*. Places may be entities that exist in the real world (with attendant latitude/longitude coordinates), common noun phrases that lack such geolocation, and places that exist only within imagined worlds (e.g., *Hogwarts*).

We decompose the overall task into four staged sub-tasks described next with examples. In each example, the characters marked by an entity tagger are underlined with a straight line, whereas places are marked with a wavy underline.

11724

Category	Short description	Annotations
IN	C is at, in, or on the place L	868
NEAR	C is in proximity of L but not at L	184
THRU	C is passing through L	41
TO	C is moving towards and is certain to reach L	171
FROM	C was at L before but has moved	98
NO REL	C and L have no relationship	622

Table 1: Short description for each individual category in the spatial relation identification subtask.

2.1 Identifying groundable characters and places

Not all people and place mentions represent entities that can be grounded with respect to each other; one important category that cannot are generic mentions (Reiter and Frank, 2010), such as *a private establishment* below:

> He insists that it must be from a private establishment.
>
> (Reeve, *The Soul Scar*)

Here, *a private establishment* refers to a class of entity, rather than an specific establishment that might exist in the narrative world. As a preprocessing step, we filter out examples whose target character or location is not able to be grounded, whether through being a generic mention or through an error in entity tagging. We formalize this as a binary classification task. An example is considered *valid* if it is correctly tagged by the entity tagger and both the character and the place entity are groundable; if either of the condition fails then the example is considered *invalid*. In the sample we annotated (described in § 3), we found 20.8% (522 out of 2506) examples to have the invalid label.

2.2 Spatial relationship categorization

Our core task seeks to categorize the relationship between the character and the place. In total there are 6 categories whose definitions are given in Table 1, with examples in Table 2. We formalized these categories to be sufficiently expressive about the different scenarios after initial rounds of grounded coding on small samples taken from books.

IN. This category is a direct judgment of whether a character is contained by the spatial boundaries delimiting the place. This relation may be evoked explicitly by prepositions such as *at, in* or *on*; more

commonly, it must be inferred, as in the example provided in Table 2.

NEAR. This label denotes whether a character is close to a location, but not contained within it. This judgment is highly contextual and relative; much like discourse can compress and decompress two entities to be more or less similar for the purpose of coreference (Recasens et al., 2010), so too can the discourse shorten and lengthen the apparent proximity of a character to a place.

THRU. Unlike previous categories in which characters are more likely to be stationary, this category implies motion through a place. THRU generally implies that an IN relation holds as well, but provides a more specific view on the nature of that relation. Crucially, this category entails that the origin and the final destination of the character are different from the place they are marked to be passing through.

TO. This category describes a character in motion towards a destination, where we are meant to draw the inference that the destination has certainly been reached. Like THRU, this category generally entails an IN relation with the destination, but provides more specificity in the nature of that movement.

FROM. This category captures movement, where a character was IN a place and has moved away from it.

NO REL. At the core of our work is a goal of differentiating character/place pairs that co-occur in the text but that do not assert that a spatial relation holds between them. NO REL describes this lack of a relation, including the cases where there is not sufficient information to deduce the relationship between C and L. NO REL can apply when a character is moving towards a destination but it is uncertain if the destination has been reached or if

Category	Example
IN	Mr. Warner stood there, his enormous bulk seeming to fill the corridor.
NEAR	But the swollen waters of the river bar our progress. I would pay its weight in gold for a raft that would transport us to the other side!
THRU	If you will not dance with me again, will you walk through the rooms? "Many admiring glances followed them–a handsomer pair was seldom seen.
TO	On his return to his room, one day, he found a glass dish on the table.
FROM	Mrs Buzzby intimated her wish, pretty strongly, that the neighbours should vacate the premises; which they did, laughingly.
NO REL	I know where Mr. Peregrine's house is.

Table 2: Examples for each individual category in the spatial relation identification subtask.

there is epistemic narrative uncertainty expressed within the text — where the narrator or characters do not know the relation between C and L.

As the examples in Table 2 illustrate, the relations between a character and place are often very obliquely expressed, relying very strongly on a reader's inference rather than explicit spatial signals within the text (a point we take up again in § 3 below).

The next two subtasks consider the time for which the spatial relationship exists (§ 2.3) and the current status of the relationship with respect to the narrative time (§ 2.4).

2.3 Temporal span classification

Characters may have a short-term or long-term spatial relationship with a place: when deciding whether a character is IN their primary home or city, for example, we can differentiate whether they are physically present there at a given moment or whether they have a habitual relationship with that place (but not necessary a punctual one at that instant).

To capture this, we mark this distinction by indicating the temporal span of the spatial relationship. The temporal span measures the amount of time that a character experiences with the place. If the relationship is short-lived or instantaneous then it is termed *punctual*, as in the following example:

> He was on the lawn, among the shrubbery.
>
> (Altsheler, *The Rulers of the Lakes*)

On the other hand, if the relationship is long-term, seasonal, or recurring, then it is considered as *habitual*, as in the following example.

> Emil Correlli flew to the nearest telegraph office and dashed off a message to a New York policeman, with whom he had had some dealings while living in that city.
>
> (Sheldon, *The Masked Bridal*)

In this example, while the text does not state whether the specific *New York policeman* is IN *New York* at the moment of utterance, we can draw the inference that they have a habitual relation to it.

2.4 Narrative tense classification

We also want to differentiate interactions between the characters and the location as taking place in the narrative past or present. On face value, this might appear as just marking the tense of the sentence but more precisely this requires marking the tense relative to the time of narration. For example, consider the following:

> He would dream night after night of his village home, and long to be back there. He sat in the parlor remembering the glorious meadow where he used to fly his kite all day long; the broad river-banks where he would wander about the livelong day singing and shouting for joy.
>
> (Tagore, *The Hungry Stones and Other Stories*)

In this case, the overall narration is happening in the past tense. Within this narrative time frame, "He sat in the parlor" is contemporaneous with this frame and hence is considered an *ongoing* relationship. However, the narration involves a reminiscence of an event that took place at some time

previous to the current narrative time ("wandering the river-banks"); the relation between *he* and *the broad river-banks* is hence considered an IN relationship that has *already happened.*

3 Annotation

We apply this framework to a sample of English-language books from Project Gutenberg. These titles span multiple centuries, index different genres, and contain a mix of fiction and narrative non-fiction. All works are in the public domain in the United States and are able to be openly published along with our annotations.

From this collection, we apply the LitBank entity tagger to identify all person and place mentions, and sample passages containing at least one character and one location separated by 10 or fewer tokens. 2506 of these samples were annotated in total by 3 annotators. The annotation process started by first carrying out pilot annotations. After an initial round of annotations, a codebook was created which was further refined in each subsequent pilot annotation round. Next, the codebook was formalized into an annotation guideline document. The document described the tasks, defined the categories per task, and gave intuitive and real examples for each category. The annotation guidelines were iteratively refined throughout the annotation process.

Training. Every annotator, other than the lead author, underwent training by reading the annotation guidelines and getting familiarized with the task. A small random sample of 50 examples were annotated and the annotations were discussed with the lead author. At this stage, any disagreements were discussed, discrepancies in the guideline were corrected, and additional clarification, if any, was added to the guideline. During training, the annotators were specifically asked to manage each annotation in under 2 minutes.

Interannotator agreement. After training, a common randomly picked sample of 261 examples was annotated independently by 2 annotators, yielding a Cohen's κ of 0.53 on identifying the validity of entities, 0.58 on the spatial relationship categorization task, 0.48 on the temporal span classification task, and 0.53 on the narrative tense classification task. After this phase, every annotator separately carried out the annotations. During this separate annotation, difficult and ambiguous ex-

amples were marked and discussed by all three annotators. The annotation guidelines were refined if necessary. Any remaining disagreements were resolved by the lead author. A distribution of the labels in the annotated data is given in Table 1.

3.1 Annotation challenges

As the examples in Table 2 make clear, along with our moderate agreement rate, spatial relationship classification is a challenging task that extensively draws on inference rather than overt lexical cues. The SpatialML task, which in many respects is conceptually similar to our task, also has low inter annotator agreement, pointing, in general, to the difficulty in relating locations with other entities (Mani et al., 2008).

To make a judgment on an example, an annotator has to make several inferences. They have to draw upon world knowledge to avoid false positives in recognizing characters and locations; for instance, the flower "Lily of the Valley" should not be mistaken for a reference to a character (or containing a reference to a specific place). In some cases, an annotator has to perform common sense reasoning; for instance, "looking out of the kitchen window" has a common sense implication of being inside the kitchen. Furthermore, narrative texts are replete with dialogues that implicitly need to be disentangled in order to attribute locations to characters. Similarly, to correctly link characters to locations, entity coreference—naturally found in narrative texts—needs to be tracked carefully.

4 Model

We build classifiers for different tasks using the annotated data. Every annotated example consists of a span of tokens denoting a character and a span of tokens denoting a location. We use the token representations from a BERT language model (Devlin et al., 2019), which we then aggregate into span representations by averaging the token representations. Assuming that c and l are vector representations for the character and location respectively, the classification model is simply:

$$\mathrm{Softmax}(\mathrm{Feedforward}(c \oplus l)),$$

where $\oplus$ is the concatenation operator between vectors. The feedforward network stacks linear layers with sigmoid activation between the layers.

Setup. We train a binary classification model to identify groundable character/place pairs, temporal span classification and narrative tense classification, and a multi-class classification model for spatial relationship classification. To assess accuracy, we divide all the annotations into 70% for training, 10% for tuning the hyperparameters, and the remaining 20% for testing. Features of the classification model are concatenated span representations from the final hidden layer of the `bert-base-cased` language model (Wolf et al., 2020). We train our models to minimize the cross entropy loss, and maintain a learning rate of $1e^{-5}$. We find the optimum value of the hyperparameters by testing the accuracy of the model against the development set. Specifically, we tune the number of epochs to train the model for early stopping from the set $\{1\dots 15\}$, the number of hidden layers from the set $\{0, 1\}$, and length of each excerpt passed to the model, operationalized as the number of tokens from the set $\{10, 50, 100\}$ before the first and last mention of the character and place in the excerpt.

Results. We compare the performance of a BERT-based classifier with several alternatives. First, we create a simple **majority class** baseline by assigning the most frequently occurring label in the training set. This baseline assigns the same label to each example. We also compare the performance with two large language models, **ChatGPT** and **GPT-4** (OpenAI, 2023), in order to explore the degree to which prompting alone—as distinct from optimizing directly on the task—is able to recover information about this complex phenomenon; such LLMs may offer promise for research in cultural analytics and computational social science by reducing the necessity for large amounts of training data (Underwood, 2023; Ziems et al., 2023), though recent work has urged caution in using these models with literary texts (Chang et al., 2023). For each of the two, we created prompts that included the description of the task, a few examples, and a rationale for each of the labels.

The overall results are presented in Table 3, which illustrate the comparative advantages of optimizing directly on the task. BERT improves over the simple majority classifier on all tasks, and also demonstrates an improvement over GPT-4 by close to 6 points on the spatial relation classification. This performance gap can be attributed to the fact that ChatGPT and GPT-4 is a limited in-context learning setup whereas BERT is trained on many

Model	Valid	Rel.	Temp.	Tense
Majority class	79.8	43.5	56.5	71.0
ChatGPT	47.7	30.2	31.4	70.7
GPT-4	83.8	51.0	59.3	69.2
BERT classifier	87.4	56.8	73.3	79.0

Table 3: Accuracy of the model on each task (whether a character/place pair is valid; classifying spatial relationship, temporal span and narrative tense) shown in percentage; The majority class model for every task predicts the most common label in the training set. The performance is reported on the test set.
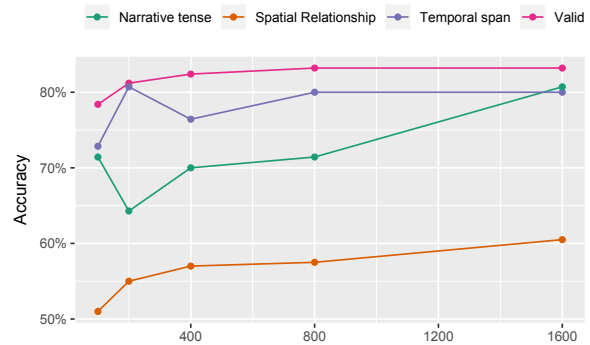

Figure 3: BERT classifier's performance as a function of the size of the training set.

examples. To elaborate this further, we show the accuracy of the BERT classifier as a function of the size of the training set in Figure 3. We find that expanding the training set with more examples benefits the classifier on all the tasks.

Finally, we break down the performance statistics by each category for this task in Table 4. As expected, the model struggles to make predictions about rare classes suggesting the need to annotate more data. The classifier's F1 is highest for the two most prominent classes (`IN` and `NO REL`) and we rely on these for the analysis that follows.

5 Analysis

A predictive model of the spatial relationships can be used to draw inferences for individual character and location co-mentions but it also opens up the possibility of macro-analysis by aggregating these individual predictions across books. In this section, we show how the model can be applied to

Category	Precision	Recall	F1
IN	0.68	0.61	0.65
NO REL	0.54	0.68	0.60
TO	0.40	0.53	0.45
FROM	0.43	0.45	0.44
NEAR	0.50	0.24	0.32
THRU	0.20	0.33	0.25

Table 4: The performance of the BERT-based classifier on the spatial prediction task. The table shows the performance broken down per class.

test known hypotheses and discover new findings about the mobility of characters in books.

Dataset. To measure the association between characters and place, we draw on several textual sources, each approximately 100 novels: the collection of English-language books from Project Gutenberg that form LitBank ([Bamman et al., 2019](#)), spanning 1719–1922; Pulitzer prize nominees from 1923–2020 (one per year); bestsellers from the *NY Times* and *Publishers Weekly* from 1923–2020 (one per year); novels written by Black authors, either from the Black Book Interactive Project[2] or Black Caucus American Library Association award winners from 1928–2018; works of Global Anglophone fiction (outside the U.S. and U.K.) from 1935–2020; and genre fiction, containing science fiction/fantasy, horror, mystery/crime, romance and action/spy novels from 1928–2017.

5.1 Protagonist Mobility

A long-held understanding in narrative studies is that stories of development, epitomized by the *Bildungsroman*, nearly always involve movement through both time and space ([Bakhtin, 1981](#)). The main characters who provide the centralized focus of such narratives are generally more mobile in comparison to other characters. We quantitatively test this hypothesis by applying our model to ground characters in places, then using the model's predictions to measure the mobility of central characters in a story, which we then compare to the mobility of non-central characters.

We operationalize the distinction between protagonists and non-protagonists by the frequency of their mentions in text (selecting the single most frequent character as the protagonist and all others as non-protagonists) and the mobility of a character

as the number of *distinct* locations among a fixed set of location mentions at which the character is grounded (*i.e.*, has prediction from the model as IN); this is analogous to a type-token ratio over grounded place. We calculate mobility over the same number of fixed location mentions for all characters to remove frequency effects from our estimation of mobility.

To test our hypothesis, we pair the most central character in a book with a randomly picked character from the next 5 central characters. We then sample 50 IN predictions for each of the characters in the pair and calculate the mobility from this sample. Averaging the mobility across all books, we can compare the mobility of protagonists to that of non-protagonists. To control for randomness due to sampling, we repeat the process 100 times.

We find that the protagonist is approximately 22% ($\pm 10\%$) more mobile, on average, than the next 5 non-protagonists. Thus, we provide positive empirical evidence for the claim that lead characters are, in general, more mobile.

We also test the hypothesis with respect to referential gender obtained using gender inference in BookNLP,[3] separating the books where protagonists align with {*she, her*} pronouns from the books where the protagonists align with {*he, him, his*} pronouns. By repeating the same procedure on this stratified set, we find slight but statistically insignificant variation in mobility across gender: lead characters who are women are 28% ($\pm 13\%$) more mobile compared to non-central characters in those books; in contrast, lead characters who are men are 19% ($\pm 12\%$) more mobile than their non-central characters. This gender parity suggests that mobility is intricately linked to the leading role of the characters, independent of their gender.

5.2 Interior space and gender

Are there gender differences in characters' position in space? Prior work in literary studies has pointed to the alignment between feminized characters and domestic spaces, especially (though not exclusively) in novels that predate the Second World War ([Armstrong, 1987](#)). When we are able to ground characters in the specific places they occupy, do we see this association empirically?

To test this hypothesis, we mark a total of 500 most frequently occurring locations as either "indoor" (e.g., *his chamber*) or "outdoor" (e.g. *the*

---

[2]http://bbip.ku.edu/novel-collections

[3]https://github.com/booknlp/booknlp

gender	indoor probability
he/him/his	0.54 ± 0.002
she/her	0.64 ± 0.002

Table 5: The proclivity of characters based on their referential gender to occupy indoor or domestic spaces. Both men and women as characters tend to occupy indoor spaces, possibly suggesting that the *de facto* settings in a novel are indoor spaces; women tend to be more indoors than men. The 95% confidence intervals are calculated using a Wald test.

coast). Next, we query the model's assignment of the spatial category to each character and location co-mention and filter out every spatial category except IN. We use BookNLP's gender inference to obtain the referential gender of each character, focusing on characters aligned with {*he, him, his*} and {*she, her*} pronouns.

We calculate the proclivity towards occupying indoor spaces by gender as $P(L = \text{``indoor''}|g(C))$, where $g(C)$ gives the referential gender of the character; $g(C) = \{he, she\}$. We find that among this set of indoor/outdoor places, women appear indoors 64% of the time, while men appear indoors only 54% of the time, a relative disparity of 18.5% (Table 5).

5.3 Interior space and time

After establishing variation in spaces occupied by characters based on the gender, we also test if this variation exists over time. To do this, we repeat the analysis in § 5.2, but on temporal slices of the data: we place books into four temporal buckets (<1873; 1873-1923; 1923-1973; 1973-2020) and calculate the association with indoor spaces for books in each temporal slice. The results are shown in Figure 4.

We see that the tendency to be depicted indoors for characters of both genders drops over time since the end of nineteenth century; in fact, characters in books from the twentieth century who are aligned with the {*he, him, his*} pronouns are equally inclined to occupy exterior spaces. The proclivity of the characters to stay indoors also shows sign of converging over time towards 0.5. An exception to this trend is the period between 1923 to 1973 which saw a surprising rise in the proclivity of characters aligned with {*he, him, his*} pronouns to stay indoors. We leave the deeper investigation of this surprising empirical fact to future work.
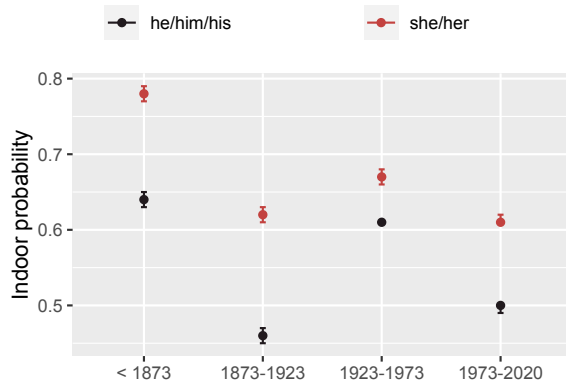

Figure 4: Proclivity of characters towards occupying indoor spaces over time. We find that characters are generally less inclined to reside in indoor spaces over time. The proclivity for both the genders to stay indoors is slowly converging towards 0.5.

6 Related Work

Our two primary contributions are a new task on classifying the spatial relationship between characters and locations and subsequent analysis about the movement of characters in narrative texts. We briefly describe the relevant work along these aspects.

In the analysis of characters and locations, named entity recognition (NER) has attracted considerable attention for its use in narrative texts such as books (e.g., Brooke et al., 2016; Bamman et al., 2019); a more focused task is the identification of characters from text (e.g., He et al., 2013; Bamman et al., 2014; Vala et al., 2015). Progress has also been made at coreference resolution, crucial for correct identification of character references in text (Bamman et al., 2020). In this work, we use a named entity tagger to mark spans of text as characters and locations, but extend the technology to categorize the relationship between a character and a location.

An analytical lens that previous research has used is to study character networks in novels (e.g., Elson et al., 2010; Agarwal et al., 2013; Dekker et al., 2019) and films (e.g., Krishnan and Eisenstein, 2015). Similarly, recent research has developed models of inter-character relationships in literary text (e.g., Iyyer et al., 2016; Srivastava et al., 2016; Chaturvedi et al., 2017). Unlike prior research, our work does not focus only on analyzing characters but instead seeks to understand their spatial relationship with places.

With respect to locations, there has been some work on toponym resolution in historial texts — a task that computationally links the text to geographic coordinatess (e.g., Smith and Crane, 2001; Speriosu and Baldridge, 2013; DeLozier et al., 2016). Others have proposed rich annotation schemes to relate multiple placenames (e.g., Mani et al., 2008; Pustejovsky et al., 2011). Our proposed task and its associated annotations differs from the previous work because we relate locations to characters.

Finally, our analytical work exemplifies application of computational methods to historical, literary text. Previous work has analyzed single attributes of a character such as gender in English fiction books (Underwood et al., 2018). In contrast, our analysis considers multiple attributes such as gender and spatial location. Prior work has also analyzed the dynamics of spatial locations, including mobility (e.g., Evans and Wilkens, 2018), and related it to measurements of emotions (e.g., Heuser et al., 2016; Semyan et al., 2022) and race (Burgers, 2020). The unit of such analyses have been books, in contrast to our work where we zoom in to analyze the mobility of characters within books.

7   Conclusion

We propose a new, challenging task of grounding characters to places in narrative texts. Unlike other domains that rely more heavily on surface lexical features to denote spatial relationships between entities, narrative texts often rely on indirect inference, exploiting a reader's commonsense knowledge and mental models of the geography within the imagined world to establish relationships between characters and the places they inhabit and navigate. This complexity presents challenges for both annotation and modeling, but we find that predictive models are able to ground characters in places for relations that are well attested in our data (e.g., IN and NO REL in particular), which offers promise for increasing the size of training data for less represented categories.

In focusing on the core category measuring when a character is IN a place, we illustrate the affordances of this work: grounding characters in place allows us not only to measure the formal elements of narrative (the structural mobility of protagonists) but also capture the degree to which mobility and the experience of space in narrative is intimately bound with gender.

8   Limitations

Our conceptualization of the core task has some important limitations. We highlight three main limitations here. First, in order to tie a character to a place, we require that both the character and the place are explicitly mentioned in the text. This simplyfying approach helps annotation and modeling but is inadequate against the general setting of grounding any character at any time in the story.

Another limitation with our current approach is the assumption that the location of a character is independent at every instance in the story. It is because of this assumption that we can label every character and location co-mention without considering any other labels. In reality, however, location of a character at some time is highly dependent on the location of the character at a previous time.

Finally, the spatial relationship categories are designed to be coarse. This is helpful in setting up the task as a classification task but collapses information that can be useful. For example, if a character is described to be *standing outside the southern gate of a building*, our current approach will assign the NEAR label retaining only the aspect of distance and not the spatial orientation.

9   Ethics Statement

While our analysis covers a wide range of English-language novels (including global Anglophone fiction, bestsellers, Pulitzer nominees, works by Black authors, genre fiction and largely canonical texts written before the 20th century), our annotated data is drawn exclusively from works in the public domain on Project Gutenberg. Our choice of Project Gutenberg as our sole source of annotated data carries a potential risk of bias in our modeling and analysis. This is because Project Gutenberg consists of data imbalances by favoring books written in English and predominantly by authors from the U.S. and the U.K. The exclusion of authors from other demographics continues the long-standing issue of *underexposure* because of which our tools and analyses are rooted in the same language and cater to a small, highly privileged demographic (Hovy and Spruit, 2016).

Acknowledgements

References

Apoorv Agarwal, Anup Kotalwar, and Owen Rambow. 2013. Automatic extraction of social networks from literary text: A case study on alice in wonderland. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1202–1208, Nagoya, Japan. Asian Federation of Natural Language Processing.

Nancy Armstrong. 1987. *Desire and Domestic Fiction: A Political History of the Novel*. Oxford University Press.

Mikhail Bakhtin. 1981. Forms of time and of the chronotope in the novel: Notes toward a historical poetics. *The Dialogic Imagination: Four Essays*, pages 84–258.

Mikhail Bakhtin. 1987. *Speech Genres and Other Late Essays*, 2nd edition edition. University of Texas Press, Austin.

David Bamman, Olivia Lewke, and Anya Mansoor. 2020. An annotated dataset of coreference in English literature. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 44–54, Marseille, France. European Language Resources Association.

David Bamman, Sejal Popat, and Sheng Shen. 2019. An annotated dataset of literary entities. 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 2138–2144, Minneapolis, Minnesota. Association for Computational Linguistics.

David Bamman, Ted Underwood, and Noah A. Smith. 2014. A Bayesian mixed effects model of literary character. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 370–379, Baltimore, Maryland. Association for Computational Linguistics.

Walter Benjamin. 2002. *The Arcades Project*, 3rd edition edition. Belknap Press, Cambridge.

Julian Brooke, Adam Hammond, and Timothy Baldwin. 2016. Bootstrapped text-level named entity recognition for literature. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 344–350, Berlin, Germany. Association for Computational Linguistics.

Johannes Burgers. 2020. Familial places in Jim Crow spaces: Kinship, demography, and the color line in william faulkner's yoknapatawpha county. *Journal of Cultural Analytics*, 5(2):14112.

Kent K Chang, Mackenzie Cramer, Sandeep Soni, and David Bamman. 2023. Speak, memory: An archaeology of books known to chatgpt/gpt-4. *arXiv preprint arXiv:2305.00118*.

Snigdha Chaturvedi, Mohit Iyyer, and Hal Daume III. 2017. Unsupervised learning of evolving relationships between literary characters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Timothy Cresswell. 2012. The production of mobilities: An interpretive framework. In *On the move*, pages 13–36. Routledge.

Niels Dekker, Tobias Kuhn, and Marieke van Erp. 2019. Evaluating named entity recognition tools for extracting social networks from novels. *PeerJ Computer Science*, 5:e189.

Grant DeLozier, Ben Wing, Jason Baldridge, and Scott Nesbit. 2016. Creating a novel geolocation corpus from historical texts. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, pages 188–198, Berlin, Germany. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. 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 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

David Elson, Nicholas Dames, and Kathleen McKeown. 2010. Extracting social networks from literary fiction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 138–147, Uppsala, Sweden. Association for Computational Linguistics.

Elizabeth Evans and Matthew Wilkens. 2018. Nation, ethnicity, and the geography of british fiction, 1880-1940. *Journal of Cultural Analytics*, 3(2):11037.

Ahmed Hamdi, Elvys Linhares Pontes, Emanuela Boros, Thi Tuyet Hai Nguyen, Günter Hackl, Jose G Moreno, and Antoine Doucet. 2021. A multilingual dataset for named entity recognition, entity linking and stance detection in historical newspapers. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2328–2334.

Hua He, Denilson Barbosa, and Grzegorz Kondrak. 2013. Identification of speakers in novels. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long*

Papers), pages 1312–1320, Sofia, Bulgaria. Association for Computational Linguistics.

Ryan Heuser, Franco Moretti, and Erik Steiner. 2016. The Emotions of London. Technical Report 13, Stanford University, Stanford, CA.

Dirk Hovy and Shannon L. Spruit. 2016. The social impact of natural language processing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 591–598, Berlin, Germany. Association for Computational Linguistics.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former Friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, San Diego, California. Association for Computational Linguistics.

Thomas Jeffers. 2016. *Apprenticeships: The Bildungsroman from Goethe to Santayana*. Palgrave MacMillan, New York.

Vinodh Krishnan and Jacob Eisenstein. 2015. "you're mr. lebowski, I'm the dude": Inducing address term formality in signed social networks. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1616–1626, Denver, Colorado. Association for Computational Linguistics.

Inderjeet Mani, Christy Doran, Dave Harris, Janet Hitzeman, Rob Quimby, Justin Richer, Ben Wellner, Scott Mardis, and Seamus Clancy. 2010. SpatialML: annotation scheme, resources, and evaluation. *Language Resources and Evaluation*, 44(3):263–280.

Inderjeet Mani, Janet Hitzeman, Justin Richer, Dave Harris, Rob Quimby, and Ben Wellner. 2008. SpatialML: Annotation scheme, corpora, and tools. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

OpenAI. 2023. Gpt-4 technical report. https://arxiv.org/abs/2303.08774.

Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative theory for computational narrative understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 298–311, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

James Pustejovsky, Jessica L Moszkowicz, and Marc Verhagen. 2011. Iso-space: The annotation of spatial information in language. In *Proceedings of the Sixth Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, volume 6, pages 1–9. Citeseer.

Marta Recasens, Eduard Hovy, and M. Antònia Martí. 2010. A typology of near-identity relations for coreference (NIDENT). In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Nils Reiter and Anette Frank. 2010. Identifying generic noun phrases. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 40–49, Uppsala, Sweden. Association for Computational Linguistics.

Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing, and Jason Baldridge. 2012. Supervised text-based geolocation using language models on an adaptive grid. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1500–1510, Jeju Island, Korea. Association for Computational Linguistics.

Tatyana Semyan, Eugene A Smyshlyaev, Olga I Babina, and Svetlana O Sheremetyeva. 2022. A map of the Urals emotional perception (based on modern regional poetry). *Digital Scholarship in the Humanities*, 37(4):1223–1239.

Amartya Sen. 1993. Capability and well-being. *The quality of life*, 30:1–445.

David A Smith and Gregory Crane. 2001. Disambiguating geographic names in a historical digital library. In *International Conference on Theory and Practice of Digital Libraries*, pages 127–136. Springer.

Michael Speriosu and Jason Baldridge. 2013. Text-driven toponym resolution using indirect supervision. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1466–1476, Sofia, Bulgaria. Association for Computational Linguistics.

Shashank Srivastava, Snigdha Chaturvedi, and Tom Mitchell. 2016. Inferring interpersonal relations in narrative summaries. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Ted Underwood. 2023. Using GPT-4 to measure the passage of time in fiction. https://tedunderwood.com/2023/03/19/using-gpt-4-to-measure-the-passage-of-time-in-fiction/.

Ted Underwood, David Bamman, and Sabrina Lee. 2018. The transformation of gender in English-language fiction. *Journal of Cultural Analytics*, 3(2):11035.

Hardik Vala, David Jurgens, Andrew Piper, and Derek Ruths. 2015. Mr. bennet, his coachman, and the archbishop walk into a bar but only one of them gets recognized: On the difficulty of detecting characters in literary texts. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 769–774, Lisbon, Portugal. Association for Computational Linguistics.

Elizabeth Wilson. 1992. The invisible flâneur. *New Left Review*, 1(191):90–110.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Janet Wolff. 1985. The invisible flâneuse. women and the literature of modernity. *Theory, Culture & Society*, 2(3):37–46.

Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2023. Can large language models transform computational social science? *arXiv submission 4840038.*

ACL 2023 Responsible NLP Checklist

A For every submission:

☑ A1. Did you describe the limitations of your work?
8

☑ A2. Did you discuss any potential risks of your work?
9

☑ A3. Do the abstract and introduction summarize the paper's main claims?
1

☒ A4. Have you used AI writing assistants when working on this paper?
Left blank.

B ☑ Did you use or create scientific artifacts?
3

☐ B1. Did you cite the creators of artifacts you used?
Not applicable. Left blank.

☐ B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
Not applicable. Left blank.

☐ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
Not applicable. Left blank.

☐ B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
Not applicable. Left blank.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
5

☐ B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
No response.

C ☑ Did you run computational experiments?
4 and 5

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
4

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

11735

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
4

☑ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
4 and 5

☐ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
Not applicable. Left blank.

D ☑ Did you use human annotators (e.g., crowdworkers) or research with human participants?

3

☐ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
Not applicable. Left blank.

☒ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
Left blank.

☐ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
Not applicable. Left blank.

☐ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
Not applicable. Left blank.

☐ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
Not applicable. Left blank.