

RESEARCH ARTICLE

Linguistic spatial classifications of event domains in narratives of crime

Blake Stephen Howald

Department of Linguistics, Georgetown University, Washington, DC, 20057-1051, USA

Received: December 9, 2009; returned: March 17, 2010; revised: April 7, 2010; accepted: May 7, 2010.

Abstract: Structurally, formal definitions of the linguistic narrative minimally require two temporally linked past-time events. The role of space in this definition, based on spatial language indicating *where* events occur, is considered optional and non-structural. However, based on narratives with a high frequency of spatial language, recent research has questioned this perspective, suggesting that space is more critical than may be readily apparent. Through an analysis of spatially rich serial criminal narratives, it will be demonstrated that spatial information qualitatively varies relative to narrative events. In particular, statistical classifiers in a supervised machine learning task achieve a 90% accuracy in predicting Pre-Crime, Crime, and Post-Crime events based on spatial (and temporal) information. Overall, these results suggest a deeper spatial organization of discourse, which not only provides practical event resolution possibilities, but also challenges traditional formal linguistic definitions of narrative.

Keywords: spatial information, discourse structure, machine learning, event classification, crime narratives, granularity, narrative event domains

1 Introduction

The essential elements of the formal structure of the linguistic narrative are events and time; for example, defined by Labov as “a sequence of two clauses which are temporally ordered [such that] a change in their order will result in a change in the temporal sequence of the original semantic interpretation” [18]. The clauses in (A) conform to this definition:

- (A) (1) Grimsby ate the treats
(2) and then sniffed the bowl

If (A2) preceded (A1), the series of narrative events would unfold differently. Therefore, (A) is a narrative. Now consider (B):

- (B) (1) Grimsby ate the treats in his kennel
(2) and then sniffed the bowl at the front desk

In (B), the inclusion of referential locations *in the kennel* (B1) and *at the front desk* (B2) adds information, such that the eating and sniffing events happen in, presumably, different locations. However, if this information is absent, as in (A), the ability to recognize (B) as a narrative is not disrupted. Now consider (C):

- (C) (1) Grimsby got up from the couch
(2) and then took a long walk

In (C), the spatial information is integrated into the event information and cannot be deleted in the same way as (B). However, this does not mean that space is critical for narrative structure. It simply happens to be the case that certain events are spatial. Event and time perspectives of narrative do not prevent this state of affairs. As long as the form of “Event 1 + temporal juncture + Event 2” is satisfied, then a narrative is both constructed and recognizable as such.

The purpose of the present investigation is to determine whether or not an expansion of the role of space in the formal structural definition of narrative is tenable, despite the diagnostics in (A–C) suggesting otherwise. Overall, there is very little research that focuses on the role of space in the formal definition of narrative. However, based on ghost story narratives, which include a large amount of spatial information, research by Herman has suggested that space plays an integral role in defining narrative structure [11]. This article furthers Herman’s perspective by focusing on serial crime narratives, which also contain a large amount of spatial information. First, “spatial information” is operationalized as a coding scheme, based on Talmy [44] which focuses on *figure* and *ground* relationships. Variation is accounted for within each of these elements based on granularity of spatial description [29] and mereotopological (part-whole) verb classifications [33]. The spatial information, in addition to temporal information (via discourse sequencing), is then accounted for across each narrative and used in a supervised machine learning task [48] to classify Pre-Crime, Crime, and Post-Crime events. Statistical classifiers are able to achieve 90% accuracy of prediction. Consequently, because this result suggests that event types have a unique spatial structure, it will be argued that space does indeed provide a contribution to narrative structure. This ultimately strengthens the argument for the formal definition of narrative being defined by event, time, and space.

This article is arranged as follows. Section 2 presents the theoretical assumptions about space and narrative; the development of a spatial coding scheme; a discussion of the analyzed data, including a counterpoint of non-crime narratives selected from the American National Corpus Charlotte Narrative and Conversation Collection (ANC) [14]; and an assessment of the distribution of the coded spatial information in two example narratives (one crime and one non-crime). Section 3 presents the supervised machine learning task. The results are analyzed and integrated into a larger discussion about the role of space in the formal structure of narrative. Limitations of the study are also presented. Section 4 concludes the article.



2 Narrative and space

Before moving forward, it is necessary to further define two key concepts, *narrative* and *space*, as well as present critical insights from recent research. First, the discussed Labovian definition of narrative stems from the sociolinguistic research tradition, which focused on narratives of personal experience (i.e., rather than vicarious experience or co-constructed narratives). However, there are a number of competing perspectives to Labov that consider numerous other factors (e.g., ethnomethodological [15], contextual [31], and identity-based [27]). Nonetheless, for purposes of this article and the generalizability of any results, almost every linguistic perspective on narrative adopts a similar event and time minimal text-type definition [1, 47]. There is an associated discourse-type, which includes abstract elements such as a particular narrator-based “point” to the narrative (e.g., *I almost died that day*) and evaluations of the narrative events (e.g., *I couldn't believe what was happening*) [19]. While the discourse-type of narrative is ultimately not a focus of this article, all of the narratives selected for analysis fit within Labovian text- and discourse-type definitions. Narrative text- and discourse-types are important for purposes of contrasting other texts. For example, discourse types such as past-tense descriptions of habitual actions, argumentation, or expository speech, may contain narrative text-type elements [47]. Conversely, the narrative discourse-type can be conveyed by alternative narrative text-types such as list structures [37], not just temporally linked events.

Second, use of the term *space* in this article refers to explicit linguistic constructions that correspond to physical relationships in the actual environment of the narrative events. While this information is considered non-critical for formally defining narrative, the information has been accounted for in secondary discourse-type structures. For example, space is typically relegated to providing background or setting information and returning to the present time via a deictic shift (e.g., orientation and coda respectively in Labov's model [18]). Further, some analyses focus on the resolution of actual events based on verb and thematic role semantics, which are applicable if the verb is subcategorized for a spatial prepositional phrase or spatial referent [20]. Some other non-Labovian frameworks and analyses include, e.g., deictic [8, 51]; cognitive [42]; cross-disciplinary (literary and cognitive) [3, 4, 35, 36, 50]; and computational approaches, namely Shapiro and Rapaport's SNePS system which included spatial information in knowledge representation and inferencing [38–40] and Yuhan and Shapiro's treatment of frame of reference resolution [49]. However, while these approaches seek to leverage spatial information for a particular purpose, the information is still secondary in terms of defining the linguistic structure of narrative.

Third, as indicated above, research by Herman, which has sought to re-evaluate the role of space in the formal structure of narratives, relies on data that is spatially very rich; in particular, ghost stories. Consider (D) (from [11]):

- (D) (1) When my grandmother died it sounded like somebody was standing in the window.
 (2) I was sitting on the bed.
 (3) And ... a door goes through to the other room
 (4) And then there was a window that high
 (5) I could look, sit on my bed and look straight out that window
 (6) And I was sitting in there [on the bed] sewing a baby's dress.

- (7) And sound like somebody jumped down out from out of that window on the floor.
- (8) And I stopped and looked.
- (9) I said “What in the world was that?”
- (10) And I got up and went to the window and looked out
- (11) I didn’t see nobody.
- (12) And a little while after that my mother done come and told me my grandmother had died.

While it is possible to come to the same conclusion about the formal structure of (D), as was possible with (A–C), at the discourse (text) level, Herman argues that reliance on spatial constructions to locate and track supernatural entities gives rise to the view that narratives are a collection of temporally linked *narrative domains* [11]. Narrative domains are defined as “mental construct[s] that encompass ... the history of spatial relationships between storyworld objects” [11]. More explicitly, narrative domains are spatial localizations consisting of “sets of verbal or visual cues anchored in mental models” [11]. Herman’s perspective is interdisciplinary as narrative domains are defined relative to the cognitive map: our internal representation of the external environment [10, 45]. Linguistically, narrative domains are defined by spatial discourse cues such as: deictic shift; figure and ground relationships; region, landmark, and path relationships; topological/projective locations; motion verbs; and what/where systems [21] (spatial language will be discussed further below in Section 2.1). For example, as applied to (D), areas are set up which include the narrator (*the bed*: D2, 5, 6, 8, 9) and the supernatural entity (*window*: D1, 4, 7, 10, 11). The narrative events are located and integrated into these areas and a broader, albeit binary (e.g., *the bed* or *the window*), spatial architecture of the narrative discourse emerges.

Narratives of serial crime, from institutionalized legal settings (e.g., guilty pleas, police statements, and confessions), are an additional source of spatially rich data to test and further Herman’s insights. While the amount of space included is, arguably, influenced by the legal setting (e.g., the inclusion of spatial information for purposes of determining jurisdiction), the offenders in these narratives are relatively unrestricted in their ability to respond. The narrators often speak for long stretches of time with minimal to no interruption, similar to the types of narratives elicited in sociolinguistic interviews. In analyzing these narratives to gain a better understanding of the contributions that spatial information may be making to narrative structure, this article has three goals:

1. Develop a coding scheme based on a comprehensive understanding of the linguistics of space, which, minimally, consolidates Herman’s perspective. The coding scheme is designed to capture multiple linguistic spatial phenomena;
2. As space is considered optional on the linguistic surface by many theories of narrative¹, the ability to account for space is potentially limited no matter how robust a coding scheme is. To garner a better understanding of how often space emerges in narrative, the distribution of the spatial codings in both crime and non-crime (ANC) narratives is compared; and

¹Of course, although the deictic center is omnipresent—something is being said by someone, at some time, at some place [5, 24]—surface realizations, either through deictic words (*here, there, come, go*) or otherwise, are not always present, nor are they crucial for narrative structure.

3. Consistent with the observation that space may or may not emerge in narrative, it follows that the type of space that does emerge could exhibit a number of different patterns (randomly, relative to structural elements (event and time), or relative to non-structural elements). Therefore, the distribution of the different coding elements across narrative will also be considered.

Each goal will be addressed in turn (Sections 2.1–2.3) and lead to the presentation of the supervised machine learning task in Section 3.

2.1 Coding of spatial information in narrative

Perspectives on spatial language tend to focus on static and dynamic relationships from both broad (e.g., what/where systems [21], vector grammar [32, 52]) and narrow points of view (e.g., spatial verb classification, cf. Levin [22], as well as spatial preposition classifications—95 of the 334 prepositions in The Preposition Project [26] have a spatial sense). However, all extant perspectives adopt some combination of syntactic, semantic, and spatial cognitive insight. The coding scheme used in this article adopts Talmy's figure and ground dichotomy with the inclusion of spatial verb information [44]. This approach to coding linguistic spatial information facilitates an evaluation of both static and dynamic spatial relationships relative to both simple and complex linguistic structures.

Figure and ground relationships are triggered by the existence of a deictic verb or adverb (e.g., *went*, *here*) (E1); a spatial preposition (e.g., *in*, *on*, *at*, *across*) (E2); a particle spatial verb (e.g., *put on*, *got out*) (E3); or spatial verb (e.g., *drive*, *sit*, *follow*, *detach* [30, 33]) (E4) (see [17, 41, 43]).

- (E) (1) [*Grimsby*]_{Fig.} *is* [*here*]_{Grd.}
 (2) [*Grimsby*]_{Fig.} *is in* [*the park*]_{Grd.}
 (3) [*Grimsby*]_{Fig.} [*rolled over* [∅]_{Grd.}]_{S.Verb}
 (4) [*Grimsby*]_{Fig.} [*ran (up) to* [*the park*]_{Grd.}]_{S.Verb}
 (5) [*Grimsby*]_{Fig.} [*ran 1.63 miles Northwest to* [*the park*]_{Grd.}]_{S.Verb}

As demonstrated in (E), several different syntactic combinations (and the inclusion of degree and measurement information (*1.63 miles Northwest*, E5) all give rise to figure and ground relationships.

Each narrative was first broken into independent clauses and then evaluated for the existence of spatial language triggering a figure and ground relationship. As the immediate concern is with the linguistic conveyance of spatial relationships of the narrative events as they correspond to the physical environment, a number of expressions were not coded. These include spatial relationships within quotatives (*she said*, "*drive me to Florida*") and metaphorical or idiomatic expressions (*in the back of my mind*, *I thought I was off scot-free*; *she went down for a nap*). The following coding were based on both the noted literature in spatial language and observations of the serial crime narrative data. Figure included grammatical person (first person singular, third person plural) and non-person objects. Although 2 and 5 (e.g., *you*, *your*) are possible, they were not observed in the data.

- Figure - Person / Object:
 - 1 - (*I*, *my*)

- 3 - (*he, she, his, her*)
- 4 - (*we, our*)
- 6 - (*they, them*)
- D - (*the tree, a hammer*)

Coding of spatial verbs is based on Pustejovsky and Moszkowicz's [33] re-organization of Muller's [30] mereotopological classifications of motion verbs (based on region connection calculus 8). This organization accounts for Move, Outside, and Hit type verbs. State type verbs have also been included to capture static spatial relationships:

- Spatial verb - mereotopologically (part-whole relationship) based classes:
 - **Move** - (*drive, fly, walk*)
 - **Outside** - (*pass, leave, deviate*)
 - **State** - (*stay, sit, is*)
 - **Hit** - (*strike, attach, detach*)

Classification of the ground follows Montello [29], who draws relationships between psychological perspectives and the linguistic realizations thereof, providing four levels of spatial knowledge. While a number of different insights are helpful, some are much too broad to be functional, such as Tversky's observation that "[at] the highest level [of organized spatial knowledge], indoor scenes are distinguished from outdoor scenes" [46].

- Ground - granularity:
 - **Geographic** - larger than the body and "learned via symbolic representations" [29]
 - **Environmental** - larger than the body with multiple (scanning) point(s) of view
 - **Vista** - larger than the body from a single point of view
 - **Figural** - smaller than the body

These four levels of spatial knowledge are classified as granularity of spatial information based on the human body (Figural) and increasingly larger spaces. (F) clarifies the ground codings.

- (F) (1) [*The football*]_{Fig.} *is in*[*New York*]_{Grd.}
 (2) [*The football*]_{Fig.} *is being thrown around* [*the field*]_{Grd.}
 (3) [*The football*]_{Fig.} *is on*[*the couch*]_{Grd.}
 (4) [*The football*]_{Fig.} *is in*[*my lap*]_{Grd.}

Geographic grounds are characterized by an object noun (of the preposition or verb) that is a toponym or proper noun named location (F1). Figural grounds (F4) are indicated by the use of a personal pronoun (*her, him*), names (*Catherine, Richard*), body parts (*hands, feet*), and anything smaller than the human body. For both Environmental (F2) and Vista (F3) grounds, the object nouns are entities larger than the human body. These grounds are distinguished by their co-occurrence with certain verb and preposition types. For example, Environmental grounds co-occur with prepositions and verbs indicating movement (*ran out of the house, went up the right hand side of the wood yard, running around the park*) and Vista grounds co-occur with stative verbs or simple prepositions (*put in the trunk, sat on the bench, staying at the bus stop*).

Putting all of the coding information together, consider (G). (G1) is coded with a 1 figure, Move spatial verb, and a Geographic ground. (G2) is coded with a 1 figure, State spatial verb, and an Environmental ground. (G3) is coded with a 3 figure, State spatial verb, and a Vista ground. An additional coding could have been included where *I* is a *figure*₂ and the area where *her standing in the intersection* occurs would be the *ground*₂. However, only the primary *figure* and *ground* relationships indexed by the spatial verbs and prepositions were coded (this applies to G5 as well). (G4) expresses no spatial relationships and is not coded. (G5) is coded with a D figure, Hit spatial verb, and a Figural ground.

- (G) (1) [I] [drove] to [Leeds]
 1 Move Geographic
- (2) [I] [was looking] around [the car] for my coffee
 1 State Environmental
- (3) and I saw [her] [standing] in [the intersection]
 3 State Vista
- (4) I thought, “she is a possibility.”
- (5) I [pulled] [my sunglasses] from [the back of the driver’s seat]
 Hit D Figural

The second goal—determining the occurrence of these codings to better understand how often space occurs in narrative—will be addressed in the next section.

2.2 Distribution of spatial information in narrative

Ten serial crime narratives (Crime), each describing one crime, were randomly selected from a corpus of serial offender narratives and coded in conformity with the parameters set forth in Section 2.1. The Crime narratives represent three different offenders who committed interpersonal violence crimes. To determine how often space occurs, it was initially noted only which clauses contained a figure and ground relationship, and which did not (the distribution of the coding elements will be discussed in Section 2.3). To compare, ten ANC narratives were randomly selected as well (the ANC narratives (except for ANC-8 below) were not fully coded).

Consider Table 1. Looking first at the ANC narratives (ANC-1–ANC-10), there is variability in the length of a narrative and, proportionally, how much space is included. The ANC narratives range from 6 (ANC-3) to 56 (ANC-8) total independent clauses. Proportions range from 4.35% (ANC-1) to 66% (ANC-3) with a 35.19% average. For those ANC narratives with a seemingly high proportion of space (ANC-3, 7, and 10), the total number of clauses is below the average. In contrast, the crime narratives exhibit some variability in the inclusion of space but are largely more consistent, ranging from 42 (Crime-5) to 162 (Crime-4) total independent clauses, 21 (Crime-7) to 89 (Crime-4) of which include spatial information. Proportions range from 33.33% (Crime-7) to 76.56% (Crime-10) with a 51.59% average. In terms of standard deviations, while the crime narratives exhibit a broader distribution in length (33.48 independent clauses) and the inclusion of spatial

Narrative ANC	Spatial / Independent Clause	Proportion	Narrative Crime	Spatial / Independent Clause	Proportion
ANC-1	1 / 23	4.35%	Crime-1	32 / 61	52.46%
ANC-2	11 / 48	22.92%	Crime-2	53 / 100	53.00%
ANC-3	4 / 6	66.67%	Crime-3	33 / 74	44.59%
ANC-4	4 / 31	12.90%	Crime-4	89 / 162	54.94%
ANC-5	20 / 55	36.36%	Crime-5	26 / 42	61.90%
ANC-6	6 / 22	27.27%	Crime-6	27 / 66	40.90%
ANC-7	5 / 10	50.00%	Crime-7	21 / 63	33.33%
ANC-8	22 / 56	39.29%	Crime-8	28 / 58	48.27%
ANC-9	4 / 14	28.57%	Crime-9	39 / 78	50.00%
ANC-10	7 / 11	63.64%	Crime-10	49 / 64	76.56%
Total	84 / 276		Total	397 / 768	
Average	8.40 / 27.60	35.19%	Average	39.70 / 76.80	51.59%
StDev	7.14 / 19.07	20.39	StDev	20.09 / 33.48	11.80

Table 1: Percentage of independent spatial clauses in Crime and ANC narratives

information (20.09), the average distribution for the crime narratives is smaller (11.80) for a higher average percentage (51.59%) as compared to the ANC narratives. The ANC narratives, which have a 19.07 average length, 7.14 including spatial information, have a higher average distribution for a lower average percentage of inclusion (35.19%).

Whether or not this greater and more consistent inclusion of spatial information is the result of a genre effect, the Crime narratives seem to be spatially different in comparison to the ANC narratives and based on what traditional perspectives on narrative would suggest. The higher variability in the ANC narratives is consistent with an expected optionality. Overall, because the spatial makeup of the Crime narratives are different from the ANC narratives, the extension of any results based on the Crime narratives must be guarded. The third goal—determining if there are patterned correspondences in the spatial information from the Crime narratives in relation to narrative events and temporal considerations—will be addressed in the next section.

2.3 Analysis of spatial information

In order to garner an understanding of what contributions the coded spatial information may be making to narrative structure, the distribution of the elements across a given narrative is informative. There are two elements to be considered, what coding elements emerge (consistent with events) and where they emerge (consistent with time/ temporal sequencing). Each coded element was thus calculated as a percentage distribution per independent clause. For example, in (G1), the 1 figure, Move spatial verb, and Geographic ground each have a value of 1.0, and the remaining elements have a value of 0. In (G2), the 1 figure again has a value of 1.0; the Move and State spatial verbs each have a value of 0.5, as do the Environmental and Geographic grounds; and the remaining elements a value of 0. In (G3), the 1 and 3 figures have respective values of 0.66 and 0.33, the Move and State spatial verbs have respective values of 0.33 and 0.66, and the Environmental, Geographic, and Vista grounds each have a value of 0.33, and so on. The coding distributions for two narratives (ANC-8 and Crime-8) are found in Tables 2 and 3.

In Table 2, ANC-8 exhibits 1 (40%) and D (48%) figures in roughly equal proportions; State spatial verbs (60%) with some Move (28%) and Hit (12%) types; and only two grounds, Environmental (25%) and Vista (75%). In Table 3, Crime-8 has a very different distribution. Crime-8 exhibits 1 (47%) and 3 (34%) figures, Move (74%) spatial verbs, and a comparatively homogenous distribution of grounds. As mentioned, if space is truly optional in narrative, any possible permutations of how much space is included, and the distributions within the included space, would be expected. The emergence of the spatial information in the narratives does not seem to follow any noticeable pattern with the exception of grounds in Crime-8.

Clauses	Figure					S. Verb				Ground			
	1	3	4	6	D	M	O	S	H	G	E	V	F
1	1.0	0	0	0	0	0	0	1.0	0	0	0	1.0	0
2	1.0	0	0	0	0	0	0	1.0	0	0	0	1.0	0
...
9	0.44	0.11	0	0	0.44	0	0	0.88	0.11	0	0.33	0.67	0
10	0.40	0.10	0	0	0.50	0	0	0.90	0.10	0	0.30	0.70	0
11	0.45	0.09	0	0	0.45	0	0	0.91	0.09	0	0.27	0.73	0
...
24	0.38	0.04	0.08	0	0.50	0.25	0	0.62	0.13	0	0.25	0.75	0
25	0.40	0.04	0.08	0	0.48	0.28	0	0.60	0.12	0	0.24	0.76	0

Table 2: Spatial coding distribution of ANC-8

As shown in Table 3, grounds in Crime-8 approach equilibrium and move uniformly from larger (Geographic, Environmental) to smaller (Vista, Figural) grounds. Interestingly, the remaining nine crime narratives follow a similar pattern (and the contributions of figure and spatial verb become more significant as well). There is some variation in ground ordering, which is consistent with what the actual narrated (physical) environments were. However, one point of regularity within this variation emerged. In particular, Figural grounds are common in narrating the commission of the crime and whenever the actual criminal act (e.g., stabbing, strangling) or extensions (e.g., throwing the murder weapon away later) are referenced. 3 figures and Hit spatial verbs also displayed a close association with reference to the victim and criminal activity.

Following these observations, the narratives were broken up into Pre-Crime (PRE), Crime (CRI), and Post-Crime (POS) groupings of narrative events. Considering PRE, CRI, and POS distinct narrative event domains is consistent with investigative and environmental criminology literature, which considers these relevant categories for the analysis of behavioral patterns [7, 34] and routine activity spaces [2, 9]. These behavioral patterns

Clauses	Figure					S. Verb				Ground			
	1	3	4	6	D	M	O	S	H	G	E	V	F
1	1.0	0	0	0	0	1.0	0	0	0	1.0	0	0	0
2	1.0	0	0	0	0	1.0	0	0	0	0.5	0	0.5	0
3	0.6	0.3	0	0	0	1.0	0	0	0	0.33	0.33	0.33	0
4	0.5	0.5	0	0	0	1.0	0	0	0	0.5	0.25	0.25	0
...
14	0.5	0.43	0	0	0.07	0.93	0.07	0	0	0.43	0.14	0.36	0.07
15	0.53	0.4	0	0	0.06	0.87	0.06	0.06	0	0.40	0.13	0.33	0.13
...
21	0.48	0.48	0	0	0.05	0.86	0.05	0.1	0.05	0.29	0.09	0.33	0.29
22	0.46	0.46	0	0	0.09	0.82	0.05	0.1	0.05	0.27	0.14	0.32	0.27
...
38	0.47	0.34	0.05	0	0.13	0.74	0.03	0.1	0.13	0.29	0.15	0.29	0.27

Table 3: Spatial coding distribution of Crime-8

have a given spatial architecture relative to an offender’s cognitive map. CRI domains started with the first encounter with the victim. POS domains started at the completion of the central crime. The boundaries between PRE and CRI event groups occur close to, or at the emergence of the Figural ground (14–15 in Table 3). And, more subtly, the boundary for CRI and POS groups corresponds to the beginning of a steady increase in Environmental or Geographic grounds through to the end of the narrative (21–22 in Table 3). Consequently, these observations provided the genesis for a supervised machine learning task that exploits all of the spatial information for purposes of classifying the PRE, CRI, and POS events. If this task is successful, then the events under inspection here can be said to have a spatial structure based on the location of self versus other, movement, and state verb types and scale-based spatial granularities. This strengthens the link between spatial information and the temporal event structure of crime narratives.

3 Classification of narrative event domains

Machine learning is particularly well suited to finding patterns in large amounts of data that may not be readily apparent by human observation. Depending on the type of algorithm used, a model (set of decisions or rules) is developed to formalize the patterns.

These formalisms provide a level of predictive accuracy, which should hold for unseen data. The formalisms can be applied to the new data and greater refinement can be accomplished (see [6,16,28,48]). *Supervised* machine learning is analogous to a classification task. A given string of data is used to classify a category. For our purposes, a task where algorithms are asked to determine a PRE, CRI, or POS event classification given a string of spatial information, is constructed. For example, consider Table 4, summarizing (G) above:

Clause	Figure	S. Verb	Ground	Event
(1)	1	Move	Geographic	Pre-Crime
(2)	1	State	Environmental	Pre-Crime
(3)	3	State	Vista	Crime
(4)	?	?	?	Crime
(5)	D	Hit	Figural	Post-Crime

Table 4: Coding of example (G)

In Table 4, each clause is given a PRE, CRI, or POS classification. Uncoded clauses were initially not included (4). The task, then, seeks to answer the question, "Can a given algorithm develop a model to predict event categories based on the spatial information of the events?"

3.1 Results

In the 10 Crime narratives, there were a total of 418 instances (each event-based string of coding is an instance, e.g., in Table 4, 1–5, are individual instances). Using the Waikato Environment for Knowledge Analysis (v3.6.0) [48], unigram and bigram instantiations were tested. For example, unigrams for (1, 3, 5) in Table 4:

Pre-Crime = 1, M, G
 Crime = 3, S, V
 Post-Crime = D, H, F

and bigrams for (1, 3, 5) in Table 4:

Pre-Crime = 11, MS, GE
 Crime = 3?, S?, V?
 Post-Crime = D_, H_, F_

An average of several different algorithms, specifically naïve bayes (NB), support vector machine (SVM), Quinlan decision tree (C4.5), and K^* , run at 10-fold cross validation, is presented in Table 5.

N=418	MAP	Figure	S. Verb	Ground	Total
Unigrams	42	52.15	44.50	48.80	60.13
Bigrams	42	55.26	51.67	52.63	65.43

Table 5: Unigram and bigram classifications of narrative event groups

The first column of Table 5 represents the *maximum a posteriori* (MAP), which is the majority classification of a given data set. In particular, for the Crime narratives, 42% of

the 418 instances were CRI events (36% of the instances were PRE events and 22% were POS events). If an individual was presented with 418 strings of spatial information, and guessed CRI each time, then the individual's predictive performance would be 42%. The MAP measure is used to determine how well a given statistical classifier is performing.

Overall, based on Table 5, bigrams outperform unigrams, and use of all spatial information (Total) outperforms any individual coding (Figure, Spatial Verb, or Ground columns). While figure performs the best compared to spatial verbs and ground (13 points above MAP for bigrams), the total spatial information does much better (unigrams are 18 points above MAP and bigrams are 23 points above MAP). However, despite good performance for certain types of information (Figure and Total) and its form (bigrams), the overall accuracy of prediction is low—65% in the best scenario (total bigrams). Consequently, a logical next step is to find ways to improve the performance of the classifier. This is not to say that 65% accuracy is unacceptable. It is simply the case that an attempt should be made to achieve the highest possible accuracy within the theoretical confines of the task.

The results in Table 5 suggest avenues for boosting the performance of the spatial information. In particular, because bigrams perform better than unigrams, some form of temporal information could improve performance. To this end, more explicit temporal information in the form of rankings was included—illustrated in Table 6.

Clause	Figure	S. Verb	Ground	Ordinal	Proportional	Centered	Event
(1)	1	Move	Geographic	1	0.2	-4	Pre-Crime
(2)	1	State	Environmental	2	0.4	-3	Pre-Crime
(3)	3	State	Vista	3	0.6	-2	Crime
(4)	(3)	(State)	(Vista)	4	0.8	-1	Crime
(5)	D	Hit	Figural	5	1.0	0	Post-Crime

Table 6: Coding and ranking information of example (G)

Three different ranks were considered to boost the performance of the classifier: (1) an ordinal ranking (1 through n , where n is the total number of independent clauses); (2) a proportional ranking, represented as a percentage of the total narrative (0.0 through 1); and (3) a centered rank where a 0 was assigned to the first occurrence of a Figural ground. Clauses from the Figural ground to the beginning of the narrative received a negative ordinal rank and a positive ordinal rank to the end of the narrative. In addition to the inclusion of temporal information, those utterances lacking a spatial coding (sparse data) were coded with the previous utterances spatial coding (clause 4 in Tables 5 and 6). The reason for this is that the codings overall represent a change in the spatial information in the narrative. For those clauses without spatial information, the spatial information is, arguably, the same as the previous clause's spatial information. This increased the number of coded instances to 792. Table 7 summarizes the results of testing bigrams with filled-in coding for those instances which previously did not have spatial information (Space) and temporal ranking information.

Compared to the results in Table 5, the K^* classifier has a better overall accuracy, 75% (33 above the MAP and 10 above total bigrams in Table 5). With the inclusion of the temporal ranking information (ordinal, proportional, and centered columns), the overall accuracy is increased even further. However, using MAP to evaluate the performance of classifiers with ranking information, while technically accurate, does not tell the whole story. It

N=792	Space	Ordinal	BR	Proportional	BR	Centered	BR
K*	75.15	88.38	75	89.27	77	90.78	75
C4.5	74.22	81.31	74	84.22	75	81.94	75
SVM	66.67	79.67	68	86.11	75	81.57	70
NB	63.00	80.30	68	81.44	78	80.43	70

Table 7: Boosted bigram classifications of narrative event groups

is important to understand how well the temporal information performs independent of the spatial information. This measure is given in the baseline rank (BR) columns. The information given to the classifier is simply the ordinal, proportional, and centered ranks and the classifying category (PRE, CRI, POS). The baseline rank is different for each classifier and form of information. Table 7 indicates that the best baseline rank is 78 (proportional NB classifier). This accuracy outperforms the best space only accuracy (75 for the K* classifier). This indicates that temporal considerations (limited to discourse sequence) are very strong relative to space. With this measure of performance, including the spatial information, the best classifier is the K* with 90% accuracy (15 points above BR and 48 points above MAP).

3.2 Summary of results

A brief summary is appropriate at this point, as a lot of material has been presented in the previous sections. First of all, with an eye toward evaluating the role of space in structural definitions of narrative, a coding scheme, designed to capture a range of spatial information, was developed and applied to narratives of serial crime, Crime narratives. This data is spatially rich as, compared to ANC narratives, the Crime narratives exhibit consistently high proportions of spatial information. Second, the distribution of the coded information was considered and, intuitively, patterns in ground information (granularity), relative to structural elements of time and event, emerged. Specifically, each of four granularity types emerged in the Crime narratives at certain points, leading not only to a uniform distribution, but also correspondences to the narrative event domains of PRE, CRI, and POS. This intuition was then tested in a supervised machine learning task to determine if each of these narrative domains possessed a discriminatory pattern of spatial information. Processing both spatial and temporal information led to a high accuracy of prediction (90%). This indicates that spatiotemporal patterns are recoverable from the Crime narratives and that space plays some role in structural elements of narrative. The patterning was not random or in opposition to time and event structural elements. The implication of these results for the formal definition of narrative will now be briefly discussed.

3.3 Discussion and limitations

There are several points of discussion on the relationship between the presented results and the formal linguistic definition of narrative. In particular, the role of granularity in the analysis, associated notions of spatial complexity across narrative discourse, and general limitations and directions for future research are discussed.

First, despite Table 5's indication that figure provided the highest overall individual accuracy compared to spatial verb and ground, it is actually ground which is most discriminating for the CRI narrative domain. For example, 80% of the CRI narrative domains included at least a Figural ground (e.g., Figural-Vista, Environmental-Figural, Geographic-Figural) and all Figural-Figural bigrams corresponded to the CRI narrative domains. This is arguably linked to the fact that granularity has a more uniform distribution (compared to figure and spatial verb) which parallels different event types in the discourse. Unigram and bigram granularities were also the highest node in the C4.5 decision tree. The distribution of figure and spatial verb is more dependent on the content of the events (3 figures typically indicate victims and Hit spatial verbs typically indicate criminal activity). Linguistically, granularity is a very different measure than figure and spatial verb. For figure and spatial verb, if a given word is encountered in a given syntactic position, then a certain coding is used. For granularity, however, information from the entire clause (with the exception of Geographic granularities) is needed to disambiguate granularity types. Relevant information is drawn from the object of the verb or preposition, from the verb or preposition itself, and general intuitions about sizes. Cohen's kappa coefficient indicates substantial agreement (0.694) for granularity codings. The largest disagreements were between Vista and Environmental granularities.

There is nothing inherent in the actual narrative events that would seem to dictate one granularity over another; at least, not in the same way as figure and spatial verb (i.e., there either was, or was not motion, and there either was, or was not somebody other than the narrator). Further, there is nothing which seemingly dictates why these crime narratives should move from larger (Geographic) to increasingly smaller (Figural) back to increasingly larger spaces as the discourse unfolds (cf. Table 3). This suggests that, for granularities, there is little contingency or dependency on the actual events of the narrative. Nonetheless, something about discourse-level spatial organization seems to be emerging. This could be an effect of certain thresholds of space, or an effect of the narratives dealing with crimes of interpersonal violence. In order to rule out these scenarios, it will be necessary to analyze, in future research, non-crime narratives with similar proportions of space and non-interpersonal violent crime narratives (e.g., burglaries, vandalism).

As granularities "zoom" from larger to smaller spaces, there is a sense of increasing spatial complexity. In relation to the crime narrative domains used in this article, research has accounted for other types of complexity of description increasing in the CRI narrative event domain. For example, Howald [12] indicates that coordinated frames of reference [23] (*in front of, to the side of, East of*) cluster in the CRI narrative event domain, while non-coordinated frames of reference (*in, at, here*) cluster in the PRE and POS narrative event domains. Howald [12] also indicates that more complex vertical "map" perspectives are invoked in PRE and POS narrative event domains, whereas less complex horizontal "tour" perspectives are found in CRI narrative event domains (following Linde and Labov [25]).

The fact that spatial information exhibits patterns relative to event and time, as demonstrated here, strengthens the argument that the formal structure of narrative is composed of event, time, and space (when available). However, in terms of events, the presented analysis is top-down; i.e., the narrative event domains were imposed on the data. Further, there is nothing inherently linguistic about PRE, CRI, and POS narrative event domains beyond their maintaining some cognitive status within environmental criminology literature, which leverages similar theoretical constructs as those perspectives which argue for a critical role of space in narrative (e.g., the cognitive map). In future research,

from a machine learning standpoint, it will be relevant to construct an unsupervised (i.e., clustering rather than classifying) learning task to determine if there are “natural” divisions in the spatial coding which are detectable and if these divisions are consistent in each narrative.

4 Conclusion

This article presents the results of a supervised machine learning task that classifies, to a 90% accuracy, event domains based on spatial (and temporal) information in narratives with a high proportion of explicit space. The success of this classification indicates that spatial information, when available, patterns relative to both time and event structures; with granularity of spatial description playing a key discriminatory role. The variation in granularity is a variation in complexity, which may ultimately be shown to be restricted to the types of narratives under analysis,² or to be representative of something deeper. Given the nature of the coding and theoretical linguistic and environmental criminological background employed, the structures revealed suggest links between linguistic form and a spatial cognitive semantic organization of discourse. This observation is consistent with Herman and the suggestion that traditional event and time perspectives on narratives of personal experience be reassessed.

Acknowledgments

I would like to thank the anonymous reviewers for helpful comments on this article; and my advisor, E. Graham Katz, for tremendous mentoring and support. In addition, I would like to thank Lissa Krawczyk, Vitaly Nickolaev, the audiences of the COSIT '09 workshop on granularity, and the University of Minnesota Linguistics Colloquium for productive discussions.

References

- [1] BAYNHAM, M. Narratives in space and time: Beyond “backdrop” accounts of narrative orientation. *Narrative Inquiry* 13 (2003), 347–366. doi:10.1075/ni.13.2.07bay.
- [2] BRANTINGHAM, P., AND BRANTINGHAM, P. *Patterns in Crime*. Macmillan, New York, NY, 1984.
- [3] BRIDGEMAN, T. Time and space. In *The Cambridge Companion to Narrative*, D. Herman, Ed. Cambridge University Press, Cambridge, UK, 2007, pp. 52–65. doi:10.1017/CCOL0521856965.

²It is important to remind the reader that, despite the scientific tenor of this article, the data analyzed, in particular the Crime narratives, depict actual crimes against actual people. These crimes are quite grim and detail a dark and unfortunate reality of human nature and existence. Consequently, in addition to expanding our perspective on linguistically defined spatial information in natural language discourse, this research has been used to improve our understanding of the spatial aspects of crime and augment the tools at the disposal of those who combat crime [13].

- [4] BUCHHOLZ, S., AND JAHN, M. Space in narrative. In *The Routledge Encyclopedia of Narrative Theory*, D. Herman, M. Jahn, and M. Ryan, Eds. Routledge, London, UK, 2005, pp. 551–555.
- [5] BUHLER, K. The deictic field of language and deictic words. In *Speech, Place and Action: Studies in Deixis and Related Topics*, R. Jarvella and K. Wolfgang, Eds. John Wiley and Sons, Chichester, UK, 1982, pp. 9–30. doi:10.1017/S0142716400004999.
- [6] CLARKE, B., FOKOUE, E., AND ZHANG, H. *Principles and Theory for Data Mining and Machine Learning*. Springer, New York, NY, 2009.
- [7] DOUGLAS, J., BURGESS, A., BURGESS, A., AND RESSLER, R. *Crime Classification Manual: A Standard System for Investigating and Classifying Violent Crimes*. Jossey-Bass Publishers, San Francisco, CA, 1992.
- [8] DUCHAN, J., BRUDER, G., AND HEWITT, L., Eds. *Deixis in Narrative: A Cognitive Science Perspective*. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1995.
- [9] FELSON, M. *Crime and Nature*. Sage Publications, Thousand Oaks, CA, 2006.
- [10] GOLLEDGE, R. Environmental cognition. In *Handbook of Environmental Psychology*, D. Stokols and I. Altman, Eds., vol. 1. John Wiley and Sons, New York, NY, 1987, pp. 131–174.
- [11] HERMAN, D. Spatial reference in narrative domains. *Text* 21 4 (2001), 515–541.
- [12] HOWALD, B. Discursive constraints on space in narrative: Evidence from guilty plea discourse. *Georgetown Working Papers in Language, Discourse & Society* 3 (2009).
- [13] HOWALD, B. The spatial aspects of victims in the narratives of serial offenders: Linguistic insights on environmental criminology and geographical profiling. *Crime Mapping: A Journal of Research and Practice* 1 (2009), 40–64.
- [14] IDE, N., AND SUDERMAN, K. The Open American National Corpus (OANC). 2007.
- [15] JEFFERSON, G. Sequential aspects of story telling in conversation. In *Studies in the Organization of Conversation*, J. Schenkein, Ed. Academic Press, New York, NY, 1978, pp. 219–248. doi:10.1016/0304-422X(86)90041-0.
- [16] KONONENKO, I., AND KUKAR, M. *Machine Learning and Data Mining: Introduction to Principles and Algorithms*. Horwood Publishing Limited, Chichester, UK, 2007.
- [17] KRACHT, M. The fine structure of spatial expressions. In *Syntax and Semantics of Spatial P*, A. Asbury, J. Dotlacil, B. Gehrke, and R. Nouwen, Eds. John Benjamins Publishing Company, Amsterdam, Netherlands, 2007, pp. 35–62.
- [18] LABOV, W. The transformation of experience in narrative syntax. In *Language in the Inner City*, W. Labov, Ed. University of Pennsylvania Press, Philadelphia, PA, 1972, pp. 354–396.
- [19] LABOV, W. Some further steps in narrative analysis. *Journal of Narrative and Life History* 7 (1997), 395–415.



- [20] LABOV, W. Uncovering the event structure of narrative. In *Georgetown University Round Table on Language and Linguistics 2001, Linguistics, Language and the Real World: Discourse and Beyond*, D. Tannen and J. Alatis, Eds. Georgetown University Press, Washington, DC, 2001, pp. 63–83.
- [21] LANDAU, B., AND JACKENDOFF, R. “What” and “where” in spatial language and cognition. *Behavioral and Brain Sciences* 16 (1993), 217–265. doi:10.1017/S0140525X00029733.
- [22] LEVIN, B. *English Verb Classes and Alternation: A Preliminary Investigation*. University of Chicago Press, Chicago, IL, 1993.
- [23] LEVINSON, S. *Pragmatics*. Cambridge University Press, Cambridge, 1983.
- [24] LEVINSON, S. Language and space. *Annual Review of Anthropology* 25 (1996), 353–382. doi:10.1146/annurev.anthro.25.1.353.
- [25] LINDE, C., AND LABOV, W. Spatial networks as a site for the study of language and thought. *Language* 51 (1975), 924–939.
- [26] LITKOWSKI, K. Diagraph analysis of dictionary preposition definitions. In *Proc. SIGLEX/SENSEVAL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions* (Philadelphia, PA, 2002), Association for Computational Linguistics, pp. 9–16. doi:10.3115/1118675.1118677.
- [27] MISHLER, E. Models of narrative analysis: A typology. *Journal of Narrative and Life History* 5 (1995), 87–123.
- [28] MITCHELL, T. *Machine Learning*. McGraw Hill Higher Education, New York, NY, 1997.
- [29] MONTELLO, D. Scale and multiple psychologies of space. In *Spatial Information Theory: A Theoretical Basis for GIS*, A. Frank and I. Campari, Eds., vol. 716 of *Lecture Notes in Computer Science*. Springer, Berlin, Germany, 1993, pp. 312–321. doi:10.1007/3-540-57207-4_21.
- [30] MULLER, P. A qualitative theory of motion based on spatio-temporal primitives. In *KR-98: Principles of Knowledge Representation and Reasoning* (1998), A. Cohn, L. Schubert, and S. Shapiro, Eds., Morgan Kaufmann, pp. 131–141.
- [31] OCHS, E., AND CAPPS, L. *Living Narrative*. Harvard University Press, Cambridge, MA, 2001.
- [32] O’KEEFE, J. The spatial prepositions in english, vector grammar and the cognitive map theory. In *Language and Space*, P. Bloom, M. Peterson, L. Nadel, and M. Garrett, Eds. MIT Press, Cambridge, MA, 1996, pp. 277–316.
- [33] PUSTEJOVSKY, J., AND MOSZKOWICZ, J. Integrating motion predicate classes with spatial and temporal annotations. *Coling* (2008), 95–98.
- [34] RESSLER, R., BURGESS, A., AND DOUGLAS, J. *Sexual Homicide: Patterns and Motives*. The Free Press, New York, NY, 1988.

- [35] RYAN, M. *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Indiana University Press, Bloomington, IL, 1991.
- [36] RYAN, M. Cognitive maps and the construction of narrative space. In *Narrative Theory and the Cognitive Sciences*, D. Herman, Ed. CSLI Publications, Stanford, CA, 2003, pp. 214–242.
- [37] SCHIFFRIN, D. Making a list. *Discourse Processes* 17 (1994), 377–405. doi:10.1080/01638539409544875.
- [38] SHAPIRO, S., AND RAPAPORT, W. SNePS considered as a fully intensional propositional semantic network. In *The Knowledge Frontier: Essays in The Representation of Knowledge*, N. Cercone and G. McCalla, Eds. Springer-Verlag, New York, NY, 1987, pp. 262–315. doi:10.1002/bimj.4710300807.
- [39] SHAPIRO, S., AND RAPAPORT, W. The SNePS family. *Computers and Mathematics with Applications* 23 (1992), 243–275. doi:10.1016/0898-1221(92)90143-6.
- [40] SHAPIRO, S., AND RAPAPORT, W. An introduction to a computational reader of narratives. In *Deixis in Narrative: A Cognitive Science Perspective*, J. Duchan, G. Bruder, and L. Hewitt, Eds. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1995, pp. 79–106.
- [41] SVENONIUS, P. Projections of P. In *Syntax and Semantics of Spatial P*, A. Asbury, J. Dotlacil, B. Gehrke, and R. Nouwen, Eds. John Benjamins Publishing Company, Amsterdam, Netherlands, 2007, pp. 63–84.
- [42] TALMY, L. Narrative structure in a cognitive framework. In *Deixis in Narrative: A Cognitive Science Perspective*, J. Duchan, G. Bruder, and L. Hewitt, Eds. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1995, pp. 421–460.
- [43] TALMY, L. *Toward a Cognitive Semantics, Vol. 1, Typology and Process in Concept Structuring*. The MIT Press, Cambridge, MA, 2000.
- [44] TALMY, L. *Toward a Cognitive Semantics, Vol. 2, Typology and Process in Concept Structuring*. The MIT Press, Cambridge, MA, 2000.
- [45] TOLMAN, E. Cognitive maps in rats and men. *Psychological Review* 55 (1948), 189–208. doi:10.1037/h0061626.
- [46] TVERSKY, B. Levels and structure of spatial knowledge. In *Cognitive Mapping: Past, Present and Future*, R. Kitchin and S. Freundschuh, Eds. Routledge, London, UK, 2000, pp. 24–43.
- [47] VIRTANEN, T. Issues of text typology: Narrative—a “basic” type of text? *Text* 12 (1992), 293–310. doi:10.1515/text.1.1992.12.2.293.
- [48] WITTEN, I., AND FRANK, E. *Data Mining—Practical Machine Learning Tools and Techniques with Java Implementation*. Morgan Kaufmann, 2002. doi:10.1145/507338.507355.
- [49] YUHAN, A., AND SHAPIRO, S. Computational representation of space. In *Deixis in Narrative: A Cognitive Science Perspective*, J. Duchan, G. Bruder, and L. Hewitt, Eds. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1995, pp. 191–226.



- [50] ZORAN, G. Towards a theory of space in narrative. *Poetics Today* 5 (1984), 309–335.
- [51] ZUBIN, D., AND HEWITT, L. The deictic center: A theory of deixis in narrative. In *Deixis in Narrative: A Cognitive Science Perspective*, J. Duchan, G. Bruder, and L. Hewitt, Eds. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1995, pp. 129–155.
- [52] ZWARTS, J. Vectors as relative positions: A compositional semantics of modified PPs. *Journal of Semantics* 14 (1997), 57–86. doi:10.1093/jos/14.1.57.

