And then I saw it: Testing Hypotheses on Turning Points in a Corpus of UFO Sighting Reports

Jan Langenhorst[1,*,†], Robert C. Schuppe[1,‡] and Yannick Frommherz[1]

[1]TUD Dresden University of Technology

Abstract

As part of developing a *Computational Narrative Understanding*, modeling events within stories has recently received significant attention within the digital humanities community. Most of the current research aims at good performance when predicting events. By contrast, we explore a focused approach based on qualitative observations. We attempt to trace the role of structural elements – more specifically, temporal function words – that may be characteristic of a narrative's turning point. We draw on a corpus of UFO sighting reports in which authors employ a prototypical narrative structure that relies on a turning point at which the extraordinary intrudes the ordinary. Using binary logistic regression, we can identify structural properties which are indicative of turning points in our data, showcasing that a focus on detail can fruitfully complement NLP models in gaining a quantitatively informed understanding of narratives.

Keywords

turning points, events, computational literary studies, corpus linguistics, logistic regression

1. Introduction

(1) *I was in my room of a paying guest flat, 5th floor, and was about to go for my bath and* ***then*** *I suddenly noticed from my window an object glowing/flashing over a jungle area more than 1 km away from my apartment.* (Report 76519)

(2) ***As*** *we drove north, 2 out of four of us saw a big bright blue ball of fire that looked as if it got brighter the closer it got to the ground.* (Report 65963)

(3) *I was getting in my car,* ***when*** *all four of us – my grandson, my grandson's tutor, my granddaughter, and myself – noticed a low, slow-moving, sideways teardrop-shaped object moving from north to south through the San Gabriel Mountains.* (Report 4061)

When people tell the story of something extraordinary which has happened to them, they use a particular kind of language. This is especially true for recounting moments when the

extraordinary intrudes the ordinary. The excerpts above are sentences stemming from texts about alleged UFO sightings which were collected online. Within these narratives, the first appearance of something out of the ordinary marks an important *turning point*. When looking at these sentences, a pattern emerges: While they might differ from other parts of the text contentwise, they also tend to stand out structurally. More precisely, they typically include an adverbial which temporally grounds the event in relation to other parts of the narrative.

Concepts related to what we just introduced as *turning points* are, among others, Labov's *most reportable event* [6], the *disruptive event* in the narrative theory of Todorov [19] or Field's *plot point* [3]. Hühn [5] distinguishes between *type-I-events* and *type-II-events*: Whereas every change of state in a story marks a type-I-event, a type-II-event is characterized by further differentiating traits, such as its unpredictability and its deviation from the norm. We see a turning point as a type-II-event which has a particular function and prototypical position in narratives. Hühn argues that type-II-events can only be identified hermeneutically [5]. However, he also notes following Schmid [17, 16] that there are criteria which hint at the presence of a type-II-event in a sentence such as, e.g., the *non-iterativity* of an action. In our context, this should entail a higher frequency of temporal function words such as the highlighted ones in (1) – (3) in sentences recounting a turning point compared to other sentences, as these words typically hint at a singular event. In this short paper, we aim at testing whether there is a systematic association between sentences containing a turning point and the use of certain context-independent markers of temporality.

Following observations such as (1) – (3), we opted to focus on *then_ADV*, *as_SCONJ*, and *when_ADV* as function words which frequently introduce temporal adverbials and seem to characterize turning point sentences. We test our hypothesis by specifying a model that predicts whether a sentence is a turning point or not, assessing whether the selected words are associated with a higher probability. By using a limited number of linguistic factors as predictors in our model, we aim to contribute to a better understanding of turning points as simpler models help to keep the impact of individual variables more transparent.

2. Related Work

The computational modeling of narratives – both their constituent elements (e.g., characters or events) and their overall structure – is a vibrant field of research [1, 2, 14] and can be seen as part of a project that strives to develop what Piper calls a *Computational Narrative Understanding* [10]. Literary event detection is a key element in this enterprise [18, 10, 11]. NLP research on how to best predict events in a narrative has yielded models that have been tested on various datasets with recent approaches reporting good performance [20, 7, 8]. Not all of these studies try and measure the same theoretical construct, since *event* and related concepts are not defined consistently [4]. Also, events can be measured on different levels, e.g., sentence vs. word. Nevertheless, all approaches have in common that they aim to extract those parts of a narrative that are distinguished from other parts in the way they contribute to the development of the story.

While these studies broadly investigate the same phenomenon, they differ from our approach in that they are predominantly concerned with predicting events while we aim at identifying

certain characteristics of those events we consider turning points. While, e.g., approaches like the one by Ouyang/McKeown [9] make extensive use of prior findings from linguistics and literary studies when selecting features for their models, they are still aimed at good predictive performance. This is typically achieved by including a myriad of different factors which, on the downside, hampers disentangling variables that contribute to what constitutes a turning point. In contrast, to be able to better interpret results, we aim to keep our model as simple as possible when estimating turning point probabilities.

3. Data

The data stem from a larger corpus of approximately 110,000 reports of UFO sightings submitted via the online platform *UFO Stalker* (https://ufostalker.com) and scraped by one of the authors.[1] Texts are mostly written in English, presumably by people from the U.S., even though these metadata cannot be verified. The reports' narrative shape is typically as follows: In a short exposition or staging phase, authors describe the – usually mundane – situation they say they were in and quite often who they were with at the time (cf. (4)). Then something happens: Most often, authors report that they suddenly see a strange light moving in the sky and such the 'actual' reporting of the sighting unfolds. This reporting is mostly an account of the author's cognitive processes. Reports typically end without them reaching a definitive conclusion with regard to what it was that they saw. This puzzlement can be seen as the prototypical resolution of these narratives.

(4) *I was inside my house with my wife, brother and his daughter. i thought i'd go out into my backyard. so i opened my door which faces east and walked out. i stopped about 6 feet from the door and felt like i needed to look up in the south direction.* **so i did and then i saw it right there in front of a low cloud.** *it like came out and went down about 25 feet then left about 25 feet then back then up and back again, then stopped and sat there. i was yelling at my wife and brother to come out here fast now! my wife was the first one yet could not see it cause she did not have her glasses on! then my brother came out and before he could actually look at it, it went into another cloud next to it. the funny thing is these clouds were kind of transparent so i do not know where it went it just went into it and vanished. at the time i saw it, it was a circular object just like a ufo to say. it was of a dark color yet you could see the sun hitting it. so it was there! but the moves were just to quick. it went from a straight down to a left turn in just a split second and did what i said above in the same time. but i did get to see it for the time mentioned above. it makes me wonder if it intended for me to see it. as when i walked out the door i had the urge to look in that direction. but who knows. this was what i saw and was amazed at what it did.* (Report 60500)

We sampled 496 texts from the larger corpus of reports. These texts were preprocessed using Stanza (Version 1.8.1) [13] for tokenization, sentence segmentation and part-of-speech tagging. Two of the authors annotated which of each report's sentences marks the turning

---

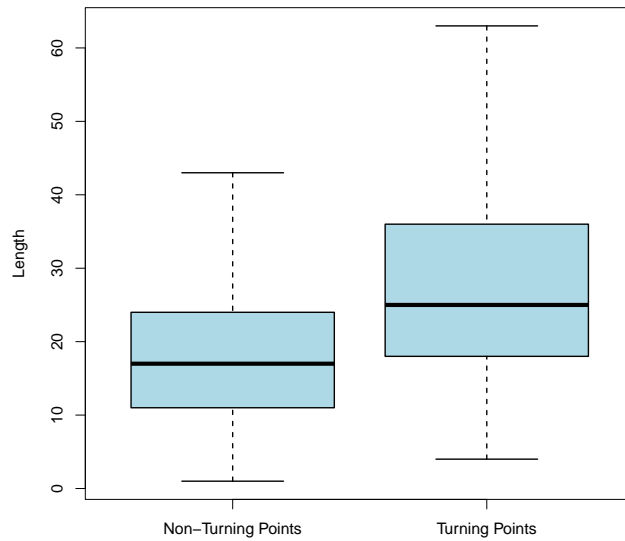[1]A similar dataset (that encompasses a different timespan) is available at Kaggle.
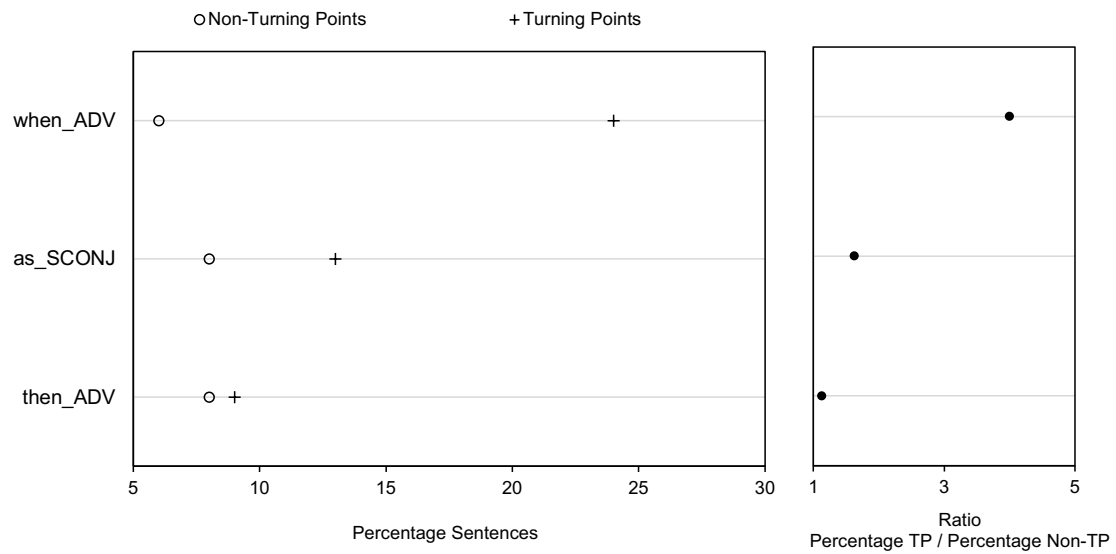
Figure 1: Percentages of sentences including *when_ADV*, *as_SCONJ* and *then_ADV* for turning points (TP) and non-turning points (left) as well as the ratios of these percentages (right).
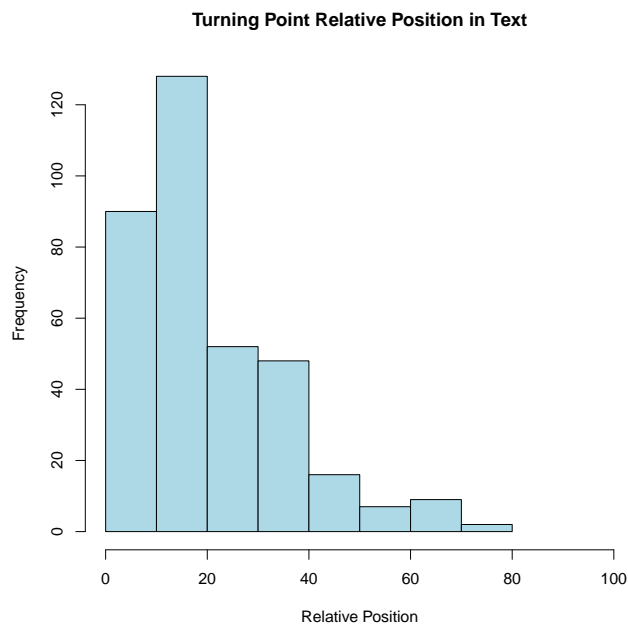
Figure 2: Distribution of relative positions of turning points within texts as a percentage.

Figure 3: Distribution of sentence lengths for turning point and non-turning point sentences. Outliers not shown.

point which we operationalized as the one sentence where it becomes clear that the narrative is about a UFO sighting, i.e., we only annotated one turning point per text. Inter-annotator agreement was good (ICC(2,1) = 0.808, 95%-CI [0.766, 0.843]; ICC(3,1) = 0.81, 95%-CI [0.769, 0.845]). Disagreement was resolved via discussion. Reports that consisted of fewer than three sentences were discarded, in line with the data preparation done by Ouyang/McKeown [9] following Prince's definition of a minimal narrative [12]. Also excluded were reports that were written in languages other than English, described something other than a UFO sighting, were a mere description of photos or videos that were provided along with the report, did not feature any narrativity or did not include a discernible turning point. Finally, 352 reports consisting of 5,346 sentences were included in the analysis. Texts contained up to 81 sentences (Median = 12, IQR = 10).

4. Modeling

To test the hypothesis laid out above, we fit binary logistic regression models with the probability of a sentence being the turning point as the outcome variable. As predictor variables we used dummy variables coding for whether the words *when_ADV*, *then_ADV* or *as_SCONJ* occurred in a given sentence. Further, we opted to include two more structural variables. First, we added the sentence's relative position within the text (the sentence's index divided by the text's length) as a percentage. Since we assumed a certain narrative structure, we knew that position within the narrative would play a role: We observed beforehand that the turning point is
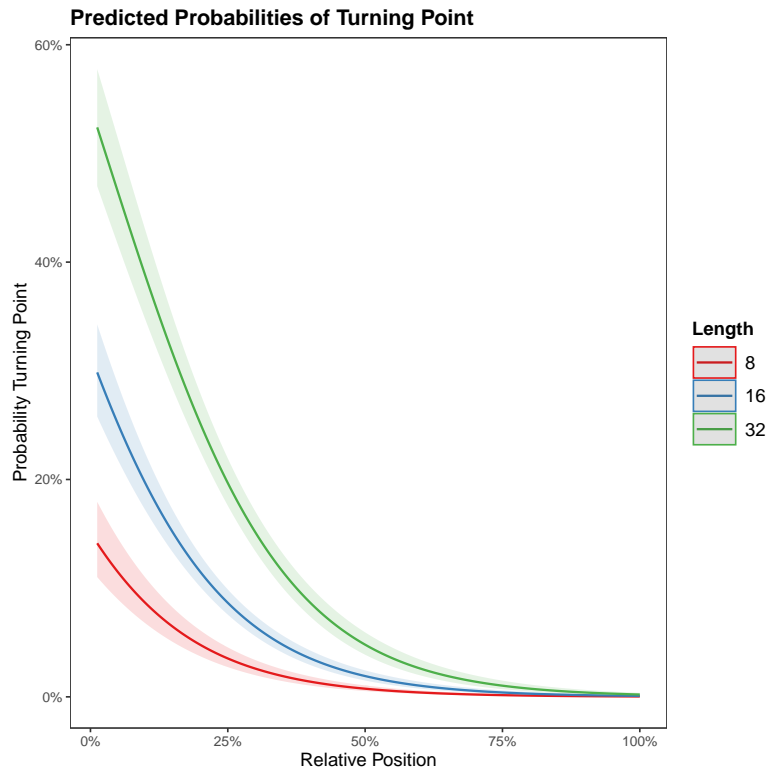
Figure 4: Predicted probabilities of turning points for all possible values of relative position and three different sentence lengths.

usually located toward the beginning of the story. Second, we included logged sentence length as a predictor. Sap et al. found that what they call *major events* are usually expressed in longer sentences [15] and a similar pattern has been observed by Ouyang/McKeown [9]. Importantly, the context of sentences is not included in any way – no information on what was written in the preceding or following sentence was used in the model, i.e., sentences were assumed independent. Thus, we do not measure any kind of *change* from one sentence to the next (like, e.g., Ouyang/McKeown do [9]), but rather compare 'global' differences between turning points and non-turning points. Note that even though sentences are naturally clustered at the text level, a multilevel model is not warranted in our case since we decided to only select one sentence per text as the turning point. Thus, the turning point probability does not vary between texts.

Looking at descriptive evidence, the three selected words do exhibit different occurrence distributions depending on whether sentences are turning points or not (Fig. 1). The percentage of sentences that include *when_ADV* is four times higher for turning points than for non-turning points. The same tendency, though less pronounced can be observed for the word *as_SCONJ*, whereas *then_ADV* occurs in a similarly sized share in both subsets of the corpus. Turning point sentences have a median relative position within the text of 16.7 (IQR = 19.4), so they are usually present in the earlier parts of the narrative (Fig. 2). Turning point sentences are also longer than non-turning point sentences in our data (Median$_{TP}$ = 25 vs. Median$_{\text{non-TP}}$ = 17;
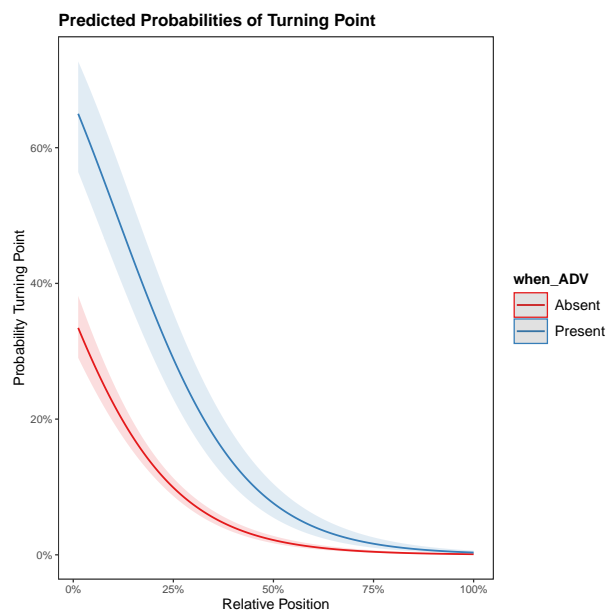
Figure 5: Predicted probabilities of turning points for all possible values of relative position and *when_ADV* absent vs. present.

Fig. 3).

5. Results

We fit three separate models. In a first step, regressing the probability of a sentence being a turning point on a sentence's relative position within the text, we estimate a negative association. This model described a small amount of variance (Tjur's R^2 = 0.113). In a second step, we added the logged sentence length (Tjur's R^2 = 0.183). Fig. 4 plots the predicted probabilities of a sentence being a turning point against its relative position within the text (a percentage value close to 0 for the very beginning and 100 for the end of a text) for different sentence lengths. As can be seen, the model assigns very low probability to sentences after half of the narrative has passed whereas sentences that lie in the first quarter of the text are assigned probabilities between around 0.15 and 0.04 for shorter sentences and between 0.54 and 0.20 for very long sentences.

Adding our main variables of interest, namely the occurrence of temporal markers resulted in improved model fit (Tjur's R^2 = 0.213; for full model comparison, see Data Availability). The estimated coefficients for *as_SCONJ* and *then_ADV* were not statistically significant, which is in accordance with the descriptive patterns presented above. The word *when_ADV*, however, was associated with an increased probability of a sentence being a turning point (Fig. 5). Again, different sentence lengths predict different probabilities for a sentence being the turning point with longer sentences being associated with higher probabilities (Fig. 6).

It is important to note that adding content words which we know a priori to be discriminative

Predicted Probabilities of Turning Point
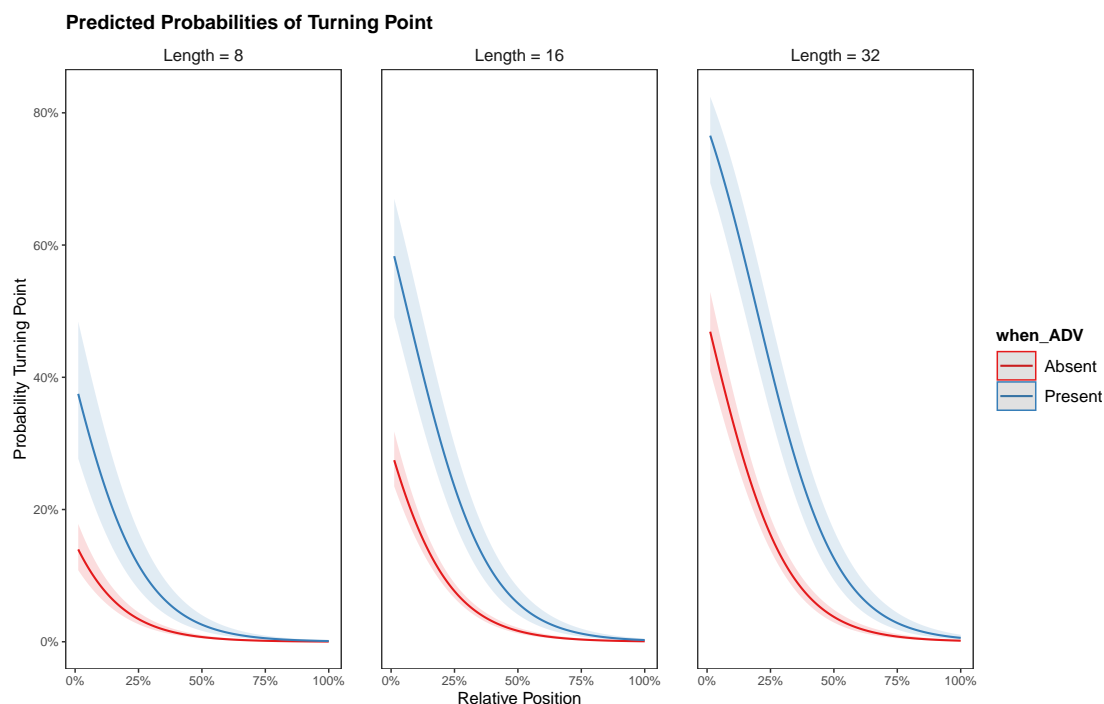
Figure 6: Predicted probabilities of turning points for sentences containing vs. not containing *when_ADV* and three different sentence lengths.

of turning points for our specific genre of text would also result in better model fit – e.g., adding a binary variable that captures whether the word *sky* appears in a given sentence (which is presumably typical of turning points in our texts since that is the locus of the extraordinary event) improves model fit (Tjur's R^2 = 0.242). While it is clear that including more or even all word occurrences into the model will result in better model fit or predictive power, respectively, it was not our aim to design a model that discriminates turning point sentences and non-turning point sentences perfectly – i.e., solve a classification problem – but rather test the theoretical question laid out above in a very specific exemplary genre of texts.

6. Discussion

Our investigation of turning points in UFO sighting narratives was driven by a hypothesis on the role of certain content-independent characteristics of turning points and had a relatively narrow scope: Not only does our corpus consist of a very particular and, possibly, idiosyncratic genre of texts. Also, our study only used a small hand-annotated sample and focused on few variables that were situated at different levels: Position within texts and sentence length did already account for some variation in the probability of sentences being turning points. Regarding the role of temporal function words, we found that while *when_ADV* is predictive of turning points, *then_ADV* and *as_SCONJ* are not. Thus, we were not able to identify a whole

957

group or class of words that are used to mark turning points, but we did corroborate that the use of *when_ADV* is predictive of a turning point. This finding supports our general hypothesis that turning points are characterized not only by their content, but also by structural properties such as temporal adverbials. Whether this also holds true for other types of narratives remains subject to further investigation.

Using state-of-the-art NLP methodology, there may be huge advances in the prediction of event types in narrative texts over the next few years. Another question, however, is how well these NLP models will serve us to understand what makes a turning point a turning point (or an event an event, for that matter). On a theoretical level, one can think about approaches like ours as modeling the reader, but also as modeling the author: What hints enable readers to place the content of a given sentence within the greater narrative? What hints does the author deem viable to trigger said interpretation? Do these cues vary between different genres that feature different narrative structures or schemas? These and many more questions should be addressed by future research from the vantage point of different disciplines – such as literary studies, linguistics, and psychology. This will help us gain a quantitatively informed understanding of (literary) narratives. We hope to have exemplified with this study how focusing on individual linguistic characteristics can complement prediction-focused approaches, aiding the development of a more thorough, corpus-based understanding of narrativity.

Data Availability

The data and code for our analysis are available at: https://osf.io/vd9pu/.

References

[1] A. Berhe, C. Guinaudeau, and C. Barras. "Survey on Narrative Structure: from Linguistic Theories to Automatic Extraction Approaches". In: *Traitement Automatique des Langues 63*. Ed. by C. Fabre, E. Morin, S. Rosset, and P. Sébillot. France: ATALA (Association pour le Traitement Automatique des Langues), 2022, pp. 63–87. URL: https://aclanthology.org/2022.tal-1.3.

[2] R. L. Boyd, K. G. Blackburn, and J. W. Pennebaker. "The Narrative Arc: Revealing Core Narrative Structures through Text Analysis". In: *Science Advances* 6.32 (2020), pp. 1–9. DOI: 10.1126/sciadv.aba2196.

[3] S. Field. *Screenplay: The Foundations of Screenwriting*. Revised. New York: Random House, 2005.

[4] E. Gius and M. Vauth. "Towards an Event Based Plot Model. A Computational Narratology Approach". In: *Journal of Computational Literary Studies* 1.1 (2022), pp. 1–20. DOI: 10.48694/jcls.110.

[5] P. Hühn. "Event and Eventfulness". In: *Handbook of Narratology*. Ed. by P. Hühn, J. C. Meister, J. Pier, and W. Schmid. Berlin/Boston: De Gruyter, 2014, pp. 159–178. DOI: 10.1515/9783110316469.159.

[6]   W. Labov. *Language in the Inner City*. Philadelphia: University of Pennsylvania Press, 1972, pp. 354–396.

[7]   V. D. Lai, T. N. Nguyen, and T. H. Nguyen. "Event Detection: Gate Diversity and Syntactic Importance Scores for Graph Convolution Neural Networks". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020, pp. 5405–5411. DOI: 10.18653/v1/2020.emnlp-main.435.

[8]   C. Liu, M. Last, and A. Shmilovici. "Identifying Turning Points in Animated Cartoons". In: *Expert Systems with Applications* 123 (2019), pp. 246–255. DOI: 10.1016/j.eswa.2019.01.003.

[9]   J. Ouyang and K. McKeown. "Modeling Reportable Events as Turning Points in Narrative". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Ed. by L. Màrquez, C. Callison-Burch, and J. Su. Lisbon, Portugal: Association for Computational Linguistics, 2015, pp. 2149–2158. DOI: 10.18653/v1/D15-1257.

[10] A. Piper. "Computational Narrative Understanding: A Big Picture Analysis". In: *Proceedings of the Big Picture Workshop*. Ed. by Y. Elazar, A. Ettinger, N. Kassner, S. Ruder, and N. A. Smith. Singapore: Association for Computational Linguistics, 2023, pp. 28–39. DOI: 10.18653/v1/2023.bigpicture-1.3.

[11] A. Piper and S. Bagga. "Toward a Data-Driven Theory of Narrativity". In: *New Literary History* 54.1 (2022), pp. 879–901. DOI: 10.1353/nlh.2022.a898332.

[12] G. Prince. *A Grammar of Stories: An Introduction*. Vol. 13. De Proprietatibus Litterarum. Series Minor. The Hague: De Gruyter, 1973.

[13] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning. "Stanza: A Python Natural Language Processing Toolkit for Many Human Languages". In: *arXiv preprint arXiv:2003.07082* (2020). DOI: https://doi.org/10.48550/arXiv.2003.07082.

[14] A. J. Reagan, L. Mitchell, D. Kiley, C. M. Danforth, and P. S. Dodds. "The Emotional Arcs of Stories are Dominated by Six Basic Shapes". In: *EPJ Data Science* 5.1 (2016), pp. 1–12. DOI: 10.1140/epjds/s13688-016-0093-1.

[15] M. Sap, A. Jafarpour, Y. Choi, N. A. Smith, J. W. Pennebaker, and E. Horvitz. "Quantifying the Narrative Flow of Imagined versus Autobiographical Stories". In: *Proceedings of the National Academy of Sciences* 119.45 (2022), pp. 1–12. DOI: 10.1073/pnas.2211715119.

[16] W. Schmid. *Elemente der Narratologie*. 3rd, revised. De Gruyter Studium. Berlin/New York: De Gruyter, 2014. DOI: https://doi.org/10.1515/9783110350975.

[17] W. Schmid. "Narrativity and Eventfulness". In: *What is Narratology? Questions and Answers Regarding the Status of a Theory*. Ed. by T. Kindt and H.-H. Müller. Vol. 1. Narratologia. Berlin/New York: De Gruyter, 2003, pp. 239–275.

[18] M. Sims, J. H. Park, and D. Bamman. "Literary Event Detection". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 3623–3634. DOI: 10.18653/v1/P19-1353.

[19] T. Todorov. "Die Grammatik der Erzählung". In: *Strukturalismus als interpretatives Verfahren*. Ed. by H. Gallas. Vol. 2. Collection Alternative. Darmstadt/Neuwied: Luchterhand, 1972, pp. 57–71.

[20] Z. Wang, A. Jafarpour, and M. Sap. "Uncovering Surprising Event Boundaries in Narratives". In: *Proceedings of the 4th Workshop of Narrative Understanding (WNU2022)*. Seattle, United States: Association for Computational Linguistics, 2022, pp. 1–12. DOI: 10.18653/v1/2022.wnu-1.1.

A. Model comparison

Table 1
Model Comparison

	Dependent variable:		
	TurningPoint		
	(1)	(2)	(3)
RelativePosition	−0.06***	−0.06***	−0.06***
	(−0.07, −0.05)	(−0.07, −0.06)	(−0.07, −0.06)
log(Length)		1.37***	1.22***
		(1.15, 1.59)	(0.98, 1.46)
as_SCONJ			0.28
			(−0.11, 0.66)
then_ADV			−0.27
			(−0.61, 0.07)
when_ADV			1.31***
			(0.85, 1.77)
Constant	−0.51***	−4.58***	−4.28***
	(−0.70, −0.32)	(−5.27, −3.89)	(−5.01, −3.55)
Tjur's R^2	0.11	0.18	0.21
Observations	5,346	5,346	5,346
Akaike Inf. Crit.	2,046.34	1,877.00	1,826.43

Note: *p<0.1; **p<0.05; ***p<0.01