

The Scenification of the Novel in the Time of Jane Austen

Nicholas Paige^{1*}; Aaron Culich²

¹ University of California, Department of French, Berkeley CA 94530-2580 USA

² University of California, Eviction Research Network, Berkeley CA 94530 USA

Abstract

This paper presents a study of the evolving use of scenes in the English novel circa 1800. The distinction between summary and scene has been a basic one in classical narratology, but comparatively little research exists on the historical development of scenic narrative techniques. Although quantitative methods have identified growth in textual features that may be expected to be linked to changes in scenic narration, the typical bag-of-words approach has limited understanding of how techniques perceptible to human readers may change over time. Our approach uses a hierarchical tagging schema to annotate scene types and their internal components to assist in quantifying scenic architecture. The project described uses BeautifulSoup4 and custom Python scripts, while future analysis will integrate computational tools such as BookNLP with custom scene annotations in order to understand the evolution of narrative techniques.

Keywords

Narratology, novel, fiction, scene, summary, BookNLP

1. Introduction

This paper presents the rationale for and design of a study of the evolving use of scenes in the English novel circa 1800.

The distinction between *scene* and *summary* is likely one of the oldest in narratology. It first starts to be articulated by “Jamseians” such as Percy Lubbock [1] and Joseph Warren Beach [2]; the pair of terms themselves appear first in Phyllis Bentley [3]; they are then picked up by Wayne Booth [4] and then finally Gérard Genette [5]. Summary telescopes a temporally dilated series of important actions by excluding the proximate context and details of each: Caesar’s “*veni, vidi, vici*” could be taken as a paradigmatic summary. By contrast, scenic narration approaches the pace of recounted events by restoring detail, often plunging the reader into the sensory world of the characters. Commonly though by no means universally, scene, which purportedly occurs “before our eyes,” is reckoned by critics to be artistically superior to summary, made in the author’s voice.¹

Impressionistically, it might seem that over the course of relatively recent literary history, novelistic scenes have taken over from summary by dint of their ability to create a vivid, immersive reading experience. “We have become so accustomed to the minute-detailing technique of the scenic model that we have great difficulty experiencing the broad and generalized strokes of summary found in Butler, Dickens, Fielding, Hardy, Thackeray and Trollope,” asserts one scholar (20) [7]. Yet research into the historical practice of scenes in the novel is scant. Ted Underwood’s quantitative research [8] has come the closest to an assessment of historical trends; while renouncing any attempt to distinguish scene from summary, Underwood nonetheless argues that the slower pace of scene has indeed come to dominate fictional prose. Calculating elapsed narrative time in 250-word segments of 90 novels reveals a “200-year trend that transformed time in fiction, moving it steadily from a scale of days to a scale of minutes” (360) [8]. (Underwood correlates this change to twin increases, of dialogue on the one hand and of description, or concrete diction, on the other.) Gingrich [6], picking up on the suggestion of Genette [5] that Proust’s fiction destroys an earlier, “classical” alteration between scenes and summaries, offers the related argument that scene had already, in the nineteenth century,

¹8th International Workshop on Computational Models of Narrative

* Corresponding author.

✉ npaige@berkeley.edu (N. Paige); aculich@berkeley.edu (A. Culich)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Gingrich [6, pp. 9–22] gives a detailed account of the history of the scene-summary opposition.

pushed out summary, depriving it of its definitional contrast and thus precipitating a later breakdown of the scenic model. In unrelated research in the field of diachronic narratology, Monika Fludernik [9] has explored the dissolution of the medieval episode, which she understands as bearing the marks of oral structures, as well as the newer, novelistic practice of what she calls consciousness scenes. (It has been suggested [7] that summary characterizes oral narration, whereas scene is a development specific to written narration.) Finally, though a preliminary effort has been made to use NLP to identify boundaries between individual scenes as well as between scenes and non-scenes [10], this research does not have a historical dimension.

The in-progress project described in this paper will provide additional data on diachronic change in scenic practices. While acknowledging Underwood's misgivings about separating scenes from summaries (344) [8] — not only are they difficult to disentangle in practice but the binary itself is merely a critical convention — we assume that the feeling of narrative “slowing,” which for Genette characterizes the scene, is experienced by readers as much as any other rhetorical effect. It is certainly true that especially in earlier novels, many scenes do not approach the kind of “real time” narration Genette describes, where the time it takes to read the scene is close to the time it would take the scene to unfold in the fictional world. In such novels, summary often occurs inside scenes as well as outside them. Nonetheless, this intra-scenic summary does not cause the reader to feel that the scene has ended. The scene, then, has a sort of narrative coherence as a unit, a coherence based not only on speed but also on continuity of space, moment, and character. With all due allowance for problematic limit cases, the researcher should be able to identify scenes in novels — when (most) readers feel them start and when (most) readers feel them end.

2. Corpus and data collection

The temporal contours of the project, which concentrates on a forty-year span around 1800, were determined by observations in a co-author's previous work [11]. These years mark what could be called the reinvention of the third-person novel: while the third person dominates the genre in the seventeenth century (especially in France), first-person forms (memoirs and epistolary fiction) largely take over in the mid- to late-eighteenth century. “Reinvented” third-person novels can be distinguished from earlier variants by their length (they become longer), by their segmentation (they are chapterized), and by their beginnings (they start less with biographical portraits and more with scenes or action). Data-gathering from that study suggested that the use of scenes — not only their frequency with respect to summary but also their inner architecture, what we might call their “sceneyness” — changes greatly around 1800, and that by the 1820s the scenic repertoire of novelists is already very developed.

The present project is designed to capture such changes in scene use, in a manner that is quantitative and archival (that is, not privileging a small subset of canonical authors) while remaining congruent with the reading experience and traditional critical practices of “close reading.” With the exception of Underwood [8], previous approaches to the novel's increasing concreteness — that is, to novelists' increasing use of physical description and sensorial cues — have used a bag-of-words approach that does not “read” in any traditional sense [12][13][14]. In Underwood [8], meanwhile, computer-chosen passages are annotated by humans, who read them linearly to assess the amount of time passing. This produces significant preliminary results, and indeed the steadily decreasing average speed of narration suggests an increasing use of scenes. But the approach still leaves us far from accounting for the changing scenic architecture of novels. Decreasing speed of narration might be an effect of simply more scenes; but it is also possible that it reflects, e.g., a greater proportion in the corpus of “scene novels” as opposed to “summary novels.” One might also wonder if scenes themselves “slow down” over time due to increased dialogue, the excision of summary within scenes, or the growth of description (say, of the physical environment or of the characters). In addition, might summaries themselves slow down? Finally, previous work does not help us understand the possible proliferation of different scene types (while the dialogue scene is

prototypical, there are others, and this variety may increase with time) or the varying length of scenes.

The project is based on the manual annotation of a chronologically stratified sample of English novels from around 1800. To capture change over time while limiting the number of titles to be annotated, we randomly selected ten novels from each of the three years 1788, 1808, and 1828 using the authoritative bibliography [15]. All were available online via Gale-Cengage's ECCO and NCCO databases or GoogleBooks. In almost all cases clean txt files of the novels did not exist. OCR was used to produce a file that was then manually checked against the page images. Formatting (e.g., paragraphing, quotational practices) was retained, along with period spelling, although obvious misprints were corrected and the long s was regularized. Though changes in scenic practices are most likely observable in all novels, to minimize variables relating to narrative mode we limited the corpus both to third-person novels specifically and to those that do not contain extensive first-person intercalated narration. Because scenes are a form of narrative segmentation, we further limited the corpus to novels sharing another form of segmentation, that of the chapter. The annotation of whole novels being prohibitively time intensive, and in order to maximize the representativeness of the sample, we selected only the initial 12,000 to 15,000 words of each: the number is approximate, the effect of guaranteeing that each sample ends on a chapter break and, further, that no scenes are truncated.

3. Tagsets

Tagsets were developed through an iterative process that brought initial common-sense notions in line with the diversity of practices visible in the samples.

As noted above, scenes have been associated with narrative slowing, often produced by the introduction of dialogue, but also with temporal and spatial specificity — in common parlance, scenes are “set.” Certain cues for scene beginnings are familiar, such as the simple formula “One day” that effects the switch from summary to scene [6, pp. 57–59], or the shift from the descriptive past imperfect to the past perfect. But scenic practices vary widely both diachronically and synchronically. Many earlier dialogue scenes, for example, contain no indications of where or when a conversation is taking place, while samples from later decades contain scenes with no dialogue. The scene, therefore, is more of an effect produced by a variable mixture of speed and spatiotemporal specification. And that effect can vary in intensity, so that some scenes are more obviously (more consensually) scenes than others.

A limited number of scene tags have proved sufficient to describe practices in the sample. The substantial majority of scenes can be divided into *dialogue scenes* (revolving around the exchange of dialogue) and *action scenes* (recounting an event, though not excluding the exchange of speech). Rarer scene types include the *perception scene* (featuring the sensory apprehension of something, focalized through a character) and the *iterative scene* (recounting habitual actions). Also rare are *monologue scenes* and *consciousness scenes*, both featuring characters alone musing to themselves, verbally or not, and *reading-writing scenes*, in which characters read or write. All of the above scene types have relatively pronounced beginnings and endings, along with a duration that enhances the reader's sense of the narrative unit (see below). *Quasi scenes*, by contrast, are moments in a summary narration that briefly scenify due to detail not usually present in summary narration — e.g., the location, the appearance of a character — and (often) snippets of dialogue. It should be emphasized that these scene types are not intended to represent a complete taxonomy of all possible scenes in all novels. The hope, however, is that they can be used to adequately describe the corpus of these years. The category of *other scenes* is assigned to special cases — passages not captured by the above tags.

As noted previously, Genette describes classical narration as featuring alternation between scenes and summaries. Indeed, very often scenes are linked by summaries, or summaries are punctuated by scenes: our phrasing will vary depending on which of the two narrative modes is perceived as being dominant. However, some novels in the corpus are composed only of scenes, and even those that are not often feature scenes connected paratactically (without intervening summary). While such

paratactical scenic construction can move the reader from one place and set of characters to another, in many other cases contiguous scenes are connected. For instance, one character in a scene may part with other characters only to go on, in a temporally continuous fashion, to meet other characters in another location. In cases in which continuity of action or conversation is such that the reader may feel the initial scene is being blended into another, the tag *blend* is applied, with both scenes being considered as one scene (with one or more blends). In cases in which the only continuity is that of character and temporality, the tag *graft* is preferred, and both scenes are tagged separately. Finally, tags are applied to moments within dialogue scenes when characters arrive or depart from the group.

Additionally, texts are annotated for dialogue utterances, to permit the calculation of utterance length and the share of scenes taken up by characters' speech. (Previous research suggests utterance length decreases over time [16], while amount of dialogue grows [17][18, pp. 13–14].) Utterances are distinguished as tagged ("‘Yes,’ she said, looking at the floor"), plain (untagged) ("Oh, mamma, I'm so happy"), or introduced with a quotative ("He left, saying, ‘See you tomorrow’). Dialogue tags are themselves tagged, with a reasonable hypothesis being that those tags become increasingly detailed.

4. Annotation protocols

The above tags can be expected to be easier or more difficult to apply, depending not only on the texts sampled but on the researcher. (If scene is largely a rhetorical effect, it stands to reason that some observers will be more or less sensitive to it.) While at the present stage all annotations have been done by the corresponding author of this paper, a future step will involve other annotators and an assessment of inter-annotator agreement. The general assumption is that the fuzziness of scenic practice — the fact that scenes are a question of degree as much as kind — does not preclude the gathering of data, and that the interpretation of that data will be highly dependent on knowledge of how it was produced.

While certain areas of many narratives can unambiguously be characterized as scenes — readers know they are "in" a scene, for example, when characters are busy exchanging dialogue — annotation requires the specification of firm boundaries. Such boundaries may be uncontroversial when novels contain only scenes, with no linking summaries; or when chapters open with scenes, or scenes terminate at the end of chapters. Nevertheless, the alternation of summary and scene is a frequent occurrence in the corpus, and in many cases that alternation is more of a blending, as scenes "fade" in or out, as it were. In these cases, an explicit protocol must be developed for tagging the specific words said to constitute the scene. Some examples will help clarify the main challenges.

In the following passage, summary narration slowly becomes scenic narration.

[Excerpt 1]

Frederick Beaumont was, at this period, thirteen years of age, and stood very high, in character and situation, at Eton school. No time was lost in his vacations; for they were passed in reading, walking, riding, and conversing, with his beloved instructor and friend. His gayer hours were spent, for the most part, in frolicking with Helena, in her mother's garden. Sometimes, too, he would assist her improvements in drawing, in which he had, himself, made very considerable advances; but his greatest delight was in riding by her side with Mr. Melcombe, who was just beginning to teach her to hold her bridle, and sit a pretty little snow-white poney, which, at this time, was her newest and favourite plaything. Mr. Melcombe was extravagantly fond of her, and when she paced, and prattled along, between himself and Frederick, it was difficult to say which looked the happiest of the three. Now and then, when she was more inclined to laugh with Frederick, than attend to her horse, the grave preceptor would pretend to scold, and tell her to mind her bridle, or he would not take her out another time.

"O yes you will, Sir," said she, one day, "and, besides, I know you and Frederick will take care of me: — as for him, if poney does but trip the least in the world, he is always ten times more frightened than I am." [19, pp. 13–14]

Up until the words “favorite plaything,” the text presents a summary of the actions taking place, iteratively, in the general time frame. With the mention of horseback riding, however, detail encourages the reader to picture the iterative actions (we know where the characters are with respect to one another; we are told to think of them looking happy), following which we are given reported speech, again said to be iterative (Melcombe would tell Helena to mind her bridle). The text becomes fully scenified with the onset of direct dialogue and the common scene trigger of “one day.”

In Excerpt 1, then, the bridge from summary to scene is an action the characters are engaged in: we move from their horseback riding in general to one particular horseback ride. Excerpt 2 offers a different type of bridge, based on contiguity of time rather than action.

[Excerpt 2]

“To-morrow” came. Frederick rose earlier: the morning engagements were all fulfilled in good time, and the tutor and pupil went together to Mrs. Villiers’s. It was a beautiful, soft, summer day; and, after the first greetings, they all adjourned to the garden. The young ones scampered away to look after their flowers, and Mr. Melcombe and Mrs. Villiers walked quietly after them.

“That’s a charming boy, Mr. Melcombe!” [19, pp. 26–27]

As in Excerpt 1, summary gradually becomes scene, but now by a kind of slowing effect.

In both these cases, therefore, we might speak of scenes “fading in”: the shift to scene is not abrupt, but follows a kind of preparation withing summary. Conversely, though many scenes end abruptly — on lines of dialogue, or on a chapter break — others may be said to “fade out” — to move into a summary mode before the conclusion of their action. The scene that starts in Excerpt 1 ends in the following manner:

[Excerpt 3]

“Helena, Helena, why did you let it go?” — She mounted again with great composure, saying,

“So! now you are going to scold!” At last, all was adjusted, and they rode peaceably on to Mrs. Villiers’s house, where she was anxiously watching for them at the parlour window. Mr. Melcombe took Helena from her steed with the greatest care, gave her an affectionate kiss, and away she ran to her mamma. [19, p. 15]

Here, the last two sentences are in summary mode: the characters and action are continuous, but the time elapsed clearly grows with respect to the previous dialogue and described actions. Yet by continuity of action those summary sentences are still closely linked to the previous scene.

Fade-in and fade-out can be considered as part of the scene or tagged separately. In the current iteration of the project, annotation protocol is to group these moments of transition with the scene in cases where a particular scene trigger is not provided. Thus, in Excerpt 2, the scene is taken as beginning with “‘To-morrow’ came”; and in Excerpt 3, the scene ends with “she ran to her mamma.” Excerpt 1, by contrast, is marked as a scene starting with “‘O yes you will, Sir,’ said she, one day.” In most cases typography aids annotation by marking boundaries for fade-in and fade-out: paragraphs separate the summary of fade-in and fade-out from other summary. When this is not the case, the boundary must be set by annotators based on their sense of narrative logic.

A last example is that of *quasi scenes*, relatively short passages in which the pace of summary narration slows and proprioceptive details accumulate.

[Excerpt 4]

Deep, and silent was the grief of the father at the separation; and long was the interval, before either the tenderness of his daughter-in-law, who, from this time, quitted him no more, or the daily opening smiles of intelligence in the countenance of her child, could bring

him comfort. At length, in the second year of his son's absence, and just as he was again beginning to taste of hope, a newspaper, in which was recorded a battle against a powerful Nabob, in the interior parts of India, concluded its melancholy tale with a list of the officers who had fallen. In that list he had barely power to read the name of Edmund Villiers.

This stroke, which his parental anxiety had too often anticipated as probable, was not, on that account, the less severe. For a time, it left him stunned — but the sight of his poor Maria, who now came smiling into the room, unconscious of the coming blow, while it thrilled to his heart, irresistibly called on him to forget, if possible, himself in her. Instantly did she perceive that something had shaken him to the centre. A moment afterwards, he desired her to sit beside him; and, summoning at once the remaining energies of his mind, he tenderly prepared her for the wound he was enforced to give. She rivetted her aching eyes on his, as he proceeded through his mournful task; and when the dreaded truth was all unfolded, convulsively clasped her hands, and dropped lifeless at his feet! A few days afterwards [...] [19, pp. 4–6]

Elements of this passage feel like a scene, even if dialogue is absent (or rather, unreported): the “now” of Maria entering the room and other temporal indications (“instantly”; “a moment afterwards”); the physical orientation of the characters (who sit side by side); gestural notations (rivetted eyes and clasped hands). The move to this degree of specificity and to this at least somewhat scenic pace would appear to commence with the trigger “At length” in the first paragraph. The sentences following in that paragraph may still be qualified as summary, yet the slowing from the summary preceding the trigger is perceptible. Hence, the passage running from “At length” to “at his feet!” is tagged as a *quasi scene*. Note that, unlike in this example, some quasi scenes contain short exchanges of dialogue.

4. Data and future data analysis

The full data and code are available in our github repository at this link:

<https://github.com/aculich/novel-scenification/>

The hierarchical tagging schema is comprised of primary scene category tags and nested component tags. We process the tagged text using BeautifulSoup4 and custom Python scripts (`count_tags.py` available in our github repository) to quantify frequency and word count of each scene type (by year) and hierarchical relationships between primary and component tags.

A first set of queries involves basic numerical calculations: cumulatively, how do ratios of summary-to-scene change; do scenes lengthen, show greater use of blends, or feature more frequent character arrivals and departures; are some scene types more prevalent in different chronological strata of the sample? Modifications in the use of dialogue, including dialogue tags, can also be tracked.

Such questions, which presuppose the construction of a kind of scenic average for each of the sample's strata, need to be supplemented with qualitative work to distinguish possible novel types. That is, it may be possible that most novels are gradually changing with respect to their scenic variety, scene length, and use of dialogue; but it is also possible that novels cluster in their use of scene, with some continuing to alternate between scene and summary while others abandon summary entirely.

Finally, we mentioned at the outset of this paper that scenes may be said to be more or less “sceney.” Scenes with little summary — thus, being closer to unfolding in real time — may seem more sceney than those where parts of a conversation or an action are summarized. Likewise, scenes with more proprioceptive queues — descriptions of place, objects, human bodies that permit readers to picture and situate people and things in space — will be experienced as “more” of a scene than those that present an exchange of dialogue whose spatiotemporal situation is not posited beyond, e.g., “tea time” in “so-and-so's home.” To be sure, simple ratios can offer clues to this kind of difference: a dialogue scene in which 95 percent of its total words is taken up by character utterances may well be experienced as less sceney than one in which only 60 percent are. But it may not be so experienced: in

such a hypothetical scene, it is equally possible that much of the remaining 40 percent of text is taken up by summary, rather than descriptive and proprioceptive information.

As such, we will apply to the tagged corpus various computational tools, with special attention to differences between text tagged as scene and the surrounding summary. Foremost, we will use BookNLP for entity tagging, supersense tagging (of lexical/semantic categories), part-of-speech tagging, event annotation, and speech attribution. We also compare dictionary-based methods (like LIWC) with unsupervised clustering (“semantic cohorts” [12]) to highlight lexical trends in scenes and nonscenes. We anticipate that results will underscore subtle style transitions across decades that contribute to “sceneyness,” as well as variations in character portrayals not captured by off-the-shelf NLP alone.

Acknowledgements

Support from the Susanne Carson Jessen Faculty Fellowship (UC Berkeley) is gratefully acknowledged.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] P. Lubbock, *The Craft of Fiction*. New York: Scribner's, 1921.
- [2] J. W. Beach, *The Twentieth-Century Novel: Studies in Technique*. New York: Appleton-Century-Crofts, 1932.
- [3] P. Bentley, *Some Observations on the Art of Narrative*. New York: Macmillan, 1947.
- [4] W. C. Booth, *The Rhetoric of Fiction*, 2nd ed. Chicago: University of Chicago Press, 1983.
- [5] G. Genette, *Narrative Discourse: An Essay in Method*. Ithaca: Cornell University Press, 1980.
- [6] B. Gingrich, *The Pace of Fiction: Narrative Movement and the Novel*. Oxford: Oxford University Press, 2021.
- [7] R. Linhares-Dias, *How to Show Things with Words: A Study on Logic, Language and Literature*. Berlin: Mouton de Gruyter, 2006.
- [8] T. Underwood, “Why Literary Time is Measured in Minutes,” *ELH*, vol. 85.2, pp. 341–65, 2018.
- [9] M. Fludernik, *Towards a “Natural” Narratology*. London: Routledge, 1996.
- [10] A. Zehe, L. Konle, L. Dümpelmann, and et al., in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 2021, pp. 3167–3177.
- [11] N. Paige, *Technologies of the Novel: Quantitative Data and the Evolution of Literary Systems*. Cambridge: Cambridge University Press, 2021.
- [12] R. Heuser and L. Le-Khac, “A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method,” *Stanford Literary Lab Pamphlet* 4, 2012.
- [13] A. Piper, *Enumerations: Data and Literary Study*. Chicago: University of Chicago Press, 2018.
- [14] T. Underwood, *Distant Horizons: Digital Evidence and Literary Change*. Chicago: University of Chicago Press, 2019.
- [15] P. Garside, J. Raven, and R. Schöwerling, *The English Novel 1770-1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*, 2 vols. London: Oxford University Press, 2000.
- [16] T. Menon, “Keeping Count: Direct Speech in the Nineteenth-Century British Novel,” *Narrative*, vol. 27.2, pp. 160–81, 2019.
- [17] O. Sobchuk, “The Evolution of Dialogues: A Quantitative Study of Russian Novels (1830–1900),” *Poet. Today*, vol. 31.1, pp. 137–54, 2016.
- [18] H. Katsma, *Morfologia del romanzo*. Milan: Nottetempo, 2024.
- [19] Anonymous, *The Master Passion; or, The History of Frederick Beaumont*. W. Miller: London, 1808.