Zero Prompting Segmentation of Movie Dialogues into Narrative Units*

José Luis López Calle[1], Pablo Gervás[1]

Facultad de Informática, Universidad Complutense de Madrid, Madrid, 28040 Spain

Abstract

If a sufficiently detailed and accurate description of a structured representation of narrative is available, using it in a zero-prompt to ask an LLM to construct the corresponding representation for a fragment of movie dialogue ought to return a reasonably correct transcription. The present paper reports on-going work to explore the possibility of asking ChatGPT to build a structured representation of simple texts extracted from the script for a movie. The input consists of an unannotated fragment of screenplay dialogue, and the output is a sequence of structured representations for an Action Unit, understood as the smallest narratively meaningful unit that can be independently annotated. Through empirical comparison with an existing dataset where the movie The Princess Bride, was annotated in terms of Actions Units, we show that the model provides reasonable responses only when applied to small fragments of text. For longer fragments of text the process omits relevant information from the input.

Keywords

Narrative representation protocol, dialogue segmentation, GPT-based execution, large language models, audiovisual narrative analysis

1. Introduction

Understanding narrative at a computational level would require being able to construct something like the complex set of nested and overlapping models that evolves in our minds as we process a story. Existing computational models of narrative compete in complexity but they do not yet advance towards a consensus. Significant progress would be achieved if we could start building extensive corpora of existing narratives modeled in terms of the various frameworks. However, the very complexity of the models and the inherent length of stories of any merit makes the task of annotating even one story a significant ordeal [1].

At the same time, we hear frequently in the media that Generative Artificial Intelligence (GenAI) – in the form of Large Language Models (LLMs) has revolutionised information technologies to the point that General Artificial Intelligence has already been achieved and GenA solutions will soon out-perform humans at any task. If this were the case, existing LLMs ought to be capable of undertaking the task of constructing structured models of a narrative from the corresponding text.

The present paper reports on-going work to explore the possibility of asking ChatGP to build a structured representation of simple texts extracted from the script for a movie. The input consists of an unannotated fragment of screenplay dialogue, preserving speaker attribution and original ordering, with no analytical markup, as shown in Figure 1, where the original transcript is reproduced.

The movie in question – The Princess Bride (Rob Reiner, 1985) [2] – has been the focus of an effort to design a representation framework for narrative structures and a systematic protocol for constructing

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): Proceedings of the Text2Story'26 Workshop, Delft (The Netherlands), 29-March-2026.

*You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

*Corresponding author.

†These authors contributed equally.

joslop11@ucm.es (J. L. L. Calle); pgervas@ucm.es (P. Gervás)

http://nil.fdi.ucm.es// (J. L. L. Calle); http://nil.fdi.ucm.es// (P. Gervás)

0000-0002-4160-9408 (J. L. L. Calle); 0000-0003-4906-9837 (P. Gervás)

GRANDFATHER

Hey! How's the sickie? Heh?

MOTHER

I think I'll leave you two pals alone.

GRANDFATHER

I brought you a special present.

THE KID

What is it?

GRANDFATHER

Open it up.

THE KID

A book?

GRANDFATHER

That's right. When I was your age, television was called books. And this is a special book. It was the book my father used to read to me when I was sick, and I used to read it to your father. And today, I'm gonna read it to you.

THE KID

Has it got any sports in it?

Figure 1: Fragment of the original screenplay.

such representations from text. This constitutes a valuable opportunity to contrast the already existing structured representation for the movie with the representations produced by an LLM when prompted with descriptions of the representation structures desired and appropriate paraphrases of the systematic protocol for constructing them.

2. Previous Work

Two lines of research need to be reviewed to provide an adequate context for the rest of the paper: existing efforts to model the structure of narrative computationally, and the framework for representing narrative structures used as target format.

2.1. Computational Representation of Narrative-Relevant Structural Elements

Computational modeling of narrative started drawing the attention of researchers in recent years. Initial efforts address issues such as the the structure of character plans and goals, the relation between narratological theories and proposals for temporal annotation schemes, and mechanisms for automatically inferring and analyzing temporal relationships [3].

More recent work has addressed the task of automatically detecting prominent elements of the narrative structure by analyzing the role of characters' inferred mental state along with linguistic information at the syntactic and semantic levels [4]. By combining embeddings and transformers, and considering information about the protagonist's mental state, the proposed model improves upon existing baselines when trying to identify climax and resolution.

Action Unit ID	OSA / DL	Character	On Screen Action Dialogue Line Action	Addressee	Subject	Verb	Direct Object	Indirect Object
27	2	Grandfather	How's the sickie?	The kid	Grandfather	ask	The kid's health	
29	2	Mother	I think I'll leave you two pals alone.	the kid, the grandfather	Mother	leave	the kid, the grandfather	
31	2	Grandfather	I brought you a special present.	The kid	Grandfather	bring	a special present	The kid
33	2	The kid	What is it?	Grandfather	The special present	be		
34	2	Grandfather	Open it up.	The kid	The kid	open	the special present	
36	2	The kid	A book?	Grandfather	The special present	be	a book	
37	2	Grandfather	That's right. When I was your age,	The kid	Grandfather	be	the age of the kid	
38	2	Grandfather	television was called books.	The kid	Television	be called	books	
39	2	Grandfather	And this is a special book.	The kid	The book	be	special	
41	2	Grandfather	It was the book.	The kid	The book	be	39	
42	2	Grandfather	My father used to read to me	The kid	Grandfather's father	read	39	Grandfather
43	2	Grandfather	when I was sick.	The kid	Grandfather	be sick		
45	2	Grandfather	and I used to read it to your father.	The kid	Grandfather	read	39	the kid's father
46	2	Grandfather	And today, I'm gonna read it to you.	The kid	Grandfather	read	the book	
47	2	The kid	Has it got any sports in it?	Grandfather	The book	contain	any sports	

Figure 2: Corresponding entries from the manual annotation.

Tian et al. [5] tested the ability of Large Language Models (LLMs) to analyze narratives in terms of three discourse-level aspects: story arcs, turning points, and affective dimensions, including arousal and valence. The results they report indicate that LLMs at that point performed significantly below human standards on that task, and that related shortcomings were observed in the related task of generating stories.

2.2. Representation of Narrative Used as Reference

The present paper builds on a comprehensive methodology for narrative representation [1]. That methodology proposed a representation for a narrative as a sequence of Action Units, where each Action Unit is the smallest narratively meaningful unit that can be independently annotated. Each Action Unit is itself structured in terms of a breakdown of the linguistic form of its text into subject, verb, direct object, indirect object. For the cases of narratives underlying a movie, Action Units would be either instances of On Screen Action or Dialogue Line. Dialogue lines also come annotated with the Character speaking the line and the Addressee. Each Action Unit is assigned a unique identifier, and AUID. An example of the fragment of the movie shown earlier represented in terms of Action Units is shown in Table 2.[1]

The methodology was originally proposed to allow detailed analysis of the structure of the narrative in terms of complex chronology [6] and on the representation of evolution of potentiality and truth status over narrative discourse [7]. For those purposes, the ability to assign distinct values of either temporal tagging or potentiality and truth status provided significant advantages over the raw text without requiring more elaborate representations in terms of semantics, character intentionality, or character affect.

The present paper focuses on the problem of whether the description of the representation protocol as originally described is sufficient for an LLM to construct the Action Units for a given sample of movie dialogue.

3. LLM Zero Prompting as Acid Test for Description of a Narrative Representation

The hypothesis underlying the experiment reported in this paper is that, if the description of the desired representation is specific and accurate, using it in a zero-prompt to ask an LLM to construct the corresponding representation for a fragment of movie dialogue ought to return a reasonably correct set of Action Units. The hypothesis was tested against the online version of ChatGPT[2] with a condensed

[1]For convenience, On Screen Actions have been excluded from this example, not being relevant for the purposes of the paper.
[2]https://chatgpt.com/

Table 1
Extract of the prompt used.

```
Your purpose is to process dialogue transcripts from
film scripts and convert them into
structured Action Units(AUIDs).
You must strictly follow the protocol defined below.
------------------------------------
SCOPE AND OPERATIONAL LIMITS
------------------------------------
This system operates exclusively on dialogue
transcripts. At this stage you must only:
• segment dialogue into minimal Action Units (AUIDs)
• populate the restricted annotation fields defined below
You must not:
• infer chronology
• infer narrative levels or universes
• infer potentiality or truth status
• infer narrative function
• speculate about analytical subsets beyond those
explicitly listed
------------------------------------
ACTION UNIT (AUID)
------------------------------------
An Action Unit (AUID) is the smallest narratively
meaningful unit that can be independently annotated.
For dialogue material:
• a dialogue line may correspond to one or multiple AUIDs
• segmentation is required whenever a dialogue line contains:
- more than one predicate with independent narrative force
- more than one speech act
- a statement followed by explanation
- a request followed by justification or condition
Rules:
• AUIDs must be assigned sequential integers
• each AUID corresponds to exactly one output row
• numbering occurs only after segmentation is complete
Guiding principle: narrative independence and
annotative clarity.
------------------------------------
ACTION SUBSET
------------------------------------
CHARACTER
The Character field records the speaker of the
dialogue unit.
Rules:
• use canonical character names
• multiple speakers may be separated by commas
• collective entities may be used where appropriate
------------------------------------
ADDRESSEE
The Addressee field specifies the target of the dialogue.
Rules:
• may be left blank in clear two-character exchanges
• must be filled when:
- more than two characters are present
- segmented AUIDs address different targets
- a collective group is addressed
------------------------------------
GRAMMATICAL BREAKDOWN
------------------------------------
Each AUID must be decomposed into:
SUBJECT
• the grammatical subject of the utterance
• may be explicit or implicit
• placeholders allowed if necessary
VERB
• the primary lexical verb
• expressed in bare infinitive
• auxiliary and modal verbs excluded
DIRECT OBJECT
• entity directly affected by the verb
• may be noun phrase or reference
INDIRECT OBJECT
• secondary recipient or contextual complement
• flexible interpretation allowed if strict object
distinction is inadequate
------------------------------------
EXPLICIT PROHIBITIONS
------------------------------------
You must not:
• infer tense
• infer aspect
• infer modality
• infer truth status
• infer narrative chronology
• interpret meaning beyond grammatical structure
• introduce commentary or explanation
------------------------------------
OUTPUT FORMAT
------------------------------------
For every AUID output exactly the following fields:
AUID
OSA/DL
Character
On-Screen Action / Dialogue Line Action
Addressee
Subject
Verb
Direct Object
Indirect Object
No additional fields are allowed.
------------------------------------
OUTPUT STYLE
------------------------------------
Return the result as structured CSV rows.
Do not include commentary, explanations, headings,
or narrative text.
Output only the annotation rows.
```

version of the description of the relevant parts of the representation – as described in Section 2.2 – and extracts of different sizes taken from the dialogues in the script for the movie The Princess Bride. Because this movie had already been transcribed manually into that representation in earlier work [1], this provides a simple Gold-standard against which to compare, which allows for objective evaluation.

3.1. Zero-Prompting ChatGPT with a Description of a Representation

A summarised version of the type of prompts employed is shown in Table 1. For simplicity and to comply with the size restrictions of the format for papers for this workshop, parts of the prompt dealing with aspects of the representation not addressed in the paper have been omitted. The version of the underlying LLM that the online version of ChatGPT reported as being in use at the time is GPT 5.2. The temperature parameter is set to `temp=0`.

The prompt encodes the rules of the narrative protocol in procedural form, specifying how dialogue must be segmented into minimal narrative units, how speakers and addressees are attributed, how grammatical roles are populated, and how references are normalized. These rules include, among others, constraints on segmentation (one independent speech act or predicate per unit), strict attribution of the Character field to the speaker, normalization of pronouns, possessives, and demonstratives to canonical referents, and the replacement of discourse-level references with Action Unit identifiers –AUIDs– when

Table 2
Summary of numerical results for the three tests.

Test	# words	Cleaned	% Action Units	% words
0	94	no	100.00	100.00
1	1,016	no	44.00	22.64
2	929	yes	64.18	35.20

prior statements are explicitly invoked. The instruction set does not contain training examples or probabilistic preferences; it consists exclusively of normative constraints derived from the protocol. Under these constraints, the GPT-based execution process operates as a procedural transformer. Given a dialogue fragment and the instruction set, it produces a linearized sequence of narrative units, each corresponding to a single AUID. Each unit is expressed as a structured row populating the protocol's required fields, including unit type, speaker, dialogue content, and grammatical decomposition. The output is deterministic in structure: every unit conforms to the same representational schema and is directly suitable for integration into an existing dataset of narrative action units without further transformation.

A crucial aspect of this execution model is that it explicitly limits analytical scope. The model is not permitted to infer narrative function, interpret character intention, resolve ambiguity beyond what is grammatically explicit, or assign temporal, modal, or truth-related values. Those dimensions are handled elsewhere in the protocol and are deliberately excluded at this stage. By restricting execution to dialogue segmentation and grammatical encoding, the process preserves methodological transparency and prevents errors and analytical overreach, and enables subsequent control and validation checks.

3.2. Testing Over Samples of Different Sizes

Experiments are reported over samples extracted from the dialogue of the movie of different sizes.

The first test – Test 0 – was conducted on a short text of 94 words (453 characters, blanks included), comprising a dialogue excerpt from The Princess Bride (Reiner 1985). Camera movements and any contextual information beyond the dialogue itself were removed from the input in order to prevent interference during processing, as such elements were present in the original version of the script used. Test 0 was carried out to validate the sequence of prompts used to guide the execution process. It was applied to the short text shown in Figure 1.

In this case, 100% of the input text was returned and processed. Several aspects of the execution are worth noting. First, segmentation produces one narrative unit per independent speech act, without collapsing or expanding units beyond protocol constraints. Second, pronouns and demonstratives ("I", "you", "it") are normalized to canonical referents whenever unambiguous, ensuring that no surface pronouns remain in grammatical fields. Third, all output units conform to the same representational schema and are immediately suitable for integration into an existing dataset of narrative action units.

Subsequent tests were conducted using the complete dialogue from The Princess Bride corresponding to Universe A –where the grandfather reads the book to the kid–. A series of executions were performed on this material.

Test 1 comprised 1,016 words (5,472 characters, blanks included). The execution returned 59 rows, compared to the 134 rows obtained through manual annotation, and accounted for 230 words, less than one quarter of the original text. The break did not occur at a clearly identifiable point; instead, the model extracted dialogue segments scattered –though sequentially– from the beginning to the end of the input, skipping intervening sections in what appeared to be a non-systematic manner. Upon revision, formatting issues were identified in the input, including off-screen tags and double paragraph breaks, which may have interfered with processing.

Test 2 was conducted on a cleaned version of the same dialogue, comprising 929 words (5,022 characters, blanks included). In this case, the output returned 86 rows, approximately half again as many as in Test 1, and closer to the original 134 rows obtained manually. The processed text accounted for 327 words, slightly over one third of the original input. As in Test 1, no single formatting issue or structural feature could be clearly identified as the cause of the remaining omissions.

In all three cases, the execution model is locally stable. The results are summarised in terms of percentages in Table 2. Test 0 works perfectly, validating both the instruction set and the GPT-as-interpreter framing. Failure is progressive: when overloaded, the GPT model does not "hallucinate" structure or invent protocol-compliant units, but instead skips, truncates, or thins out the output.

4. Discussion

A number of issues need to be discussed regarding the outlined experiments.

First, cleaning dialogue transcripts and removing extraneous markup leads to measurable improvements in performance; results for test 2 are significantly better than for text 1, with comparable input sizes. This may limit applicability to noisier sources unless additional preprocessing steps are introduced.

Second, the process is very sensitive to the length of the input. The results of Test 0 show that, when the execution process operates on bounded dialogue fragments, segmentation and reference handling closely match manually curated annotations. The results of Tests 1 and 2 show that increasing input length degrades results. However, it is important to note that the degradation observed involves progressive omission of certain parts of the input, but no uncontrolled reinterpretation or structural hallucination is observed: units are skipped or truncated, but incorrect or invented structures are not introduced. From the perspective of protocol-driven annotation, this is a preferable failure mode. Missing units can be detected and recovered through reprocessing or manual intervention, whereas false positives would compromise the integrity of the dataset and obscure later analytical stages. This observation nevertheless sets limits on how much of the material can be reliably processed in a single pass, leading to the need for batch-based execution if the procedure is to be applied over longer texts.

Third, certain forms of ambiguity remain irreducible at the level of automated execution. In cases where the protocol itself allows or requires under-specification, the execution process correctly leaves fields unresolved rather than attempting to resolve them heuristically. This behaviour preserves analytical integrity, but also confirms that human judgment remains necessary for a subset of decisions.

5. Conclusions

This paper reports on on-going work to test the hypothesis that a sufficiently detailed description of a structured representation format for narrative, when used as a zero-prompt for a large language model, may under limited circumstances produce correct instances of the desired representation. Over short input texts, the results do correspond with the set of Action Units produced by human annotators. However, over longer inputs the results exhibit progressive omission of relevant events in the input.

Several avenues for further work remain open. One-shot or few-shot prompting solutions may certainly improve results. Revision of the descriptions in the prompt based on observed shortcomings will also be considered.

Acknowledgments

This paper has been partially supported by project DARK NITE: Dialogue Agents Relying on Knowledge-Neural hybrids for Interactive Training Environments, Grant No. PID2023-146308OB-I00 (Spanish Ministry of Science and Innovation).

Declaration on Generative AI

The authors have not employed any Generative AI tools in the writing of this paper.

References

[1] J. L. López-Calle, A Comprehensive Computational Representation of Narrative-Relevant Structural Elements in a Film, Ph.D. thesis, Universidad Complutense de Madrid, Madrid, Spain, 2025.

[2] R. Reiner, The Princess Bride, Twentieth Century-Fox Film Corporation, 1985. Film.

[3] I. Mani, Computational modeling of narrative, Springer Nature, 2022.

[4] P. Vijayaraghavan, D. Roy, M-sense: Modeling narrative structure in short personal narratives using protagonist's mental representations, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, 2023, pp. 13664–13672.

[5] Y. Tian, T. Huang, M. Liu, D. Jiang, A. Spangher, M. Chen, J. May, N. Peng, Are large language models capable of generating human-level narratives?, in: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, 2024, pp. 17659–17681.

[6] P. Gervás, J. L. López-Calle, Representing complex relative chronology across narrative levels in movie plots., in: Text2Story@ ECIR, 2024, p. 65–76.

[7] P. Gervás, J. L. López-Calle, Accounting for the importance of changes in event actuality in the representation of narrative, in: Text2Story@ ECIR, volume 3964, 2025, p. 117–125.