LATECH-CLFL 2025, THE 9TH JOINT SIGHUM WORKSHOP ON COMPUTATIONAL LINGUISTICS FOR CULTURAL HERITAGE, SOCIAL SCIENCES, HUMANITIES AND LITERATURE AT NAACL 2025, ALBUQUERQUE, NM, USA

MAY THE 4TH, 2025

RETHINKING SCENE SEGMENTATION

ADVANCING AUTOMATED DETECTION OF SCENE CHANGES IN LITERARY TEXTS

Guhr, Svenja°*, Huijun Mao*, Fengyi Lin*
°*fortext lab*, Technical University of Darmstadt, DE
*Literary Lab, Stanford University, USA

SUMMARY

Automated scene segmentation remains a central challenge in computational literary studies, especially for narrative texts with fluid boundaries. This study presents a new approach to detecting scene changes in 20th-century US-English romance fiction. Building on manually annotated data from the Harlequin "Men Made in America" series and other popular fiction texts, we fine-tuned the pretrained transformer-based model Universal Sentence Encoder (USE) to identify scene changes in six-sentence segments.

DEFINITION OF SCENE

"A scene is defined as a segment of the discourse (presentation) of a narrative which presents a part of the histoire (chronologically ordered, causally connected events in the narrated world) in such a way that

a) **time** is more or less similar in discourse and histoire,
b) **place** stays - more or less - the same,
c) it centers around a particular **action**, and
d) the **character** configuration is (…) equal"

(Gius et al. 2021)

CORPUS

- 50 novels "Men Made in America"
- 20 translated popular fiction novels
- 30 novels manually scene-annotated

METHOD

0. Scene Theory
1. Manual annotation of scene segments
2. Discussion, IAA measurement
3. Preparation of gold annotations
4. Data preprocessing and splitting in six-sentence segments
5. Automated USE segment classification

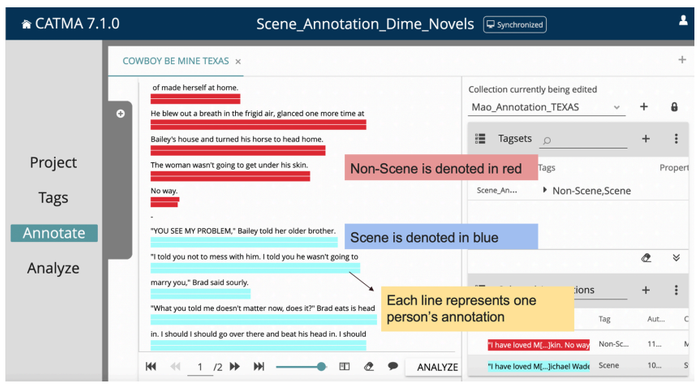

Figure 1: Manual annotation with the annotation software CATMA.

ANALYSIS

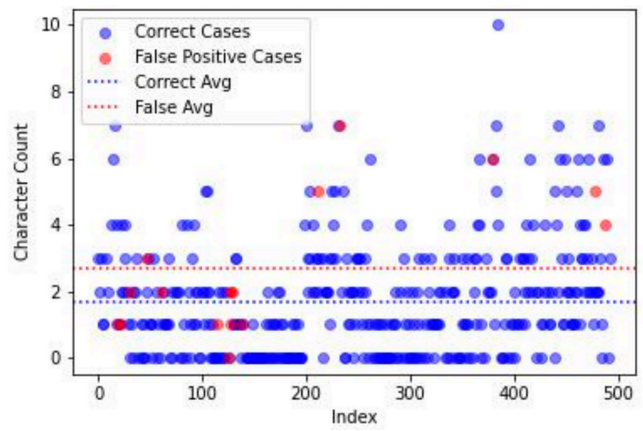

Figure 2: Scatter plot of character counts for correct and false positive cases.

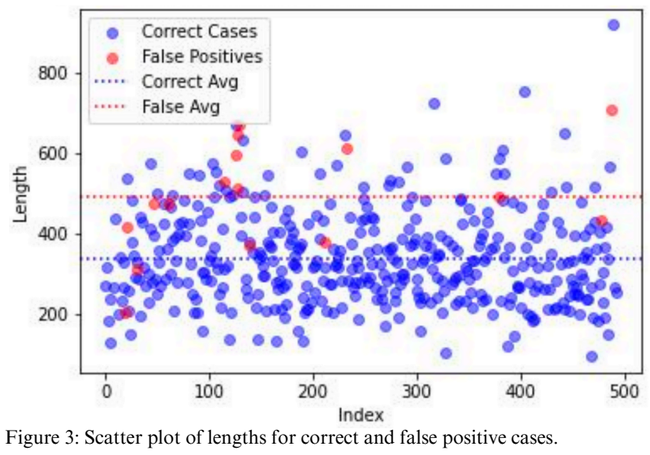

Figure 3: Scatter plot of lengths for correct and false positive cases.

Category	Average character count	Annot. differing from gold label
correct	1.67	0
incorrect	2.29	0.62
false negative	2.20	0.53
false positive	2.69	1.02

Table 3: Average character count comparison across different categories.

- More false positives than false negatives
- Model:
  - tends to miss scene changes rather than misidentifying them where none exist
  - more likely to make errors when processing longer inputs
- Key factors influencing the predictions:
  - presence of character names
  - ambiguity in manual annotations

SUMMARY OF MANUAL ANNOTATIONS

Author (Date)	Title	Cohen's *k*	Mathet's γ	Scene changes in gold annot.	words
Ferrarella (2000)	*Found: His Perfect Wife* AK	0.49	–	59	65,421
Broadrick (1986)	*Deceptions* CA	0.3	–	69	42,605
Stuart (1984)	*Tangled Lies* HI	0.3	0.31	88	69,124
Palmer (1985)	*Love By Proxy* IL	0.3	–	76	42,063
Campbell (1987)	*Pros and Cons* MA	0.38	–	91	75,527
Webb (2000)	*Warrior's Embrace* MS	0.21	0.34	118	58,903
McKenna (1984)	*Too Near the Fire* OH	0.78	0.41	39	43,319
Leonard (2000)	*Cowboy Be Mine* TX	0.52	0.53	65	62,421
Neggers (1989)	*Finders Keepers* VT	0.51	–	90	52,869
Cassidy (1997)	*Midnight Wishes* WY	0.39	–	100	60,655
10 Romance novels	from "Men Made in America"	ø0.4	ø0.4	795	572,907
20 translated novels	reuse from Zehe et al. (2025)	–	ø0.7	1,250	597,659
30 novels	total training set	–	–	2,045	1,170,566

Table 1: IAA between two expert human annotators.

AUTOMATION

- Data: 2,045 manually annotated scene changes
- Model: USE (Cer et al., 2018) using sentence-level semantic embeddings
- Input: Text split into six-sentence passages
- Output: Binary classification
  1 for scene change, 0 for no scene change
- Balancing: Class balancing via random oversampling to reduce majority bias

Model Performance	(first trial) BERT	(first trial) USE	(final training) USE	(final training) test set
Accuracy				
Training	0.92	0.94	0.81	0.83
Validation	0.92	0.95	0.81	
F1				
Training	0.48	0.69	0.66	0.5
Validation	0.48	0.71	0.65	
Precision				
Training	0.47	0.74	0.72	0.59
Validation	0.46	0.76	0.72	
Recall				
Training	0.50	0.65	0.62	0.44
Validation	0.50	0.67	0.61	
Loss				
Training	0.31	0.16	0.43	–
Validation	0.31	0.17	0.42	

Table 2: Evaluation metrics for model training.

CONCLUSION

- Probabilistic approach for scene change detection without exact sentence-level boundaries
- Six-sentence segment approach offers:
  - solution balancing granularity and ambiguity in narrative transitions
  - new perspective on meaningful segmentation of narrative texts for CLS

ACKNOWLEDGEMENTS

Many thanks to our student annotators / other project members Mallen Clifton, Agnes Hilger, Jessica Monaco, Alexander J. Sherman, and Ellen Yang, who supported the project with their close readings, manual scene annotations, and discoveries of unconventional scene structures. The conference travel was funded by the priority program SPP 2207 Computational Literary Studies (CLS) of the German Research Foundation (DFG).

REFERENCES

Cer, D., Y. Yang, S Kong, N. Hua, N. Limtiaco, R. St. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y. Sung, B. Strope, and R. Kurzweil. 2018. "Universal Sentence Encoder." 10.48550/arXiv.1803.11175

Gius, E., F. Jannidis, M. Krug, A. Zehe, A. Hotho, F. Puppe, J. Krebs, N. Reiter, N. Wiedmer, and L. Konle. 2019. "Detection of Scenes in Fiction." In *Book of Abstracts of DH 2019*. 10.34894/OOY9CE

Gius, E., C. Sökefeld, L. Dümpelmann, L. Kaufmann, A. Schreiber, S. Guhr, N. Wiedmer, and Fotis Jannidis. 2021. "Guidelines for Detection of Scenes." Zenodo. 10.5281/ZENODO.4457177

Zehe, A., E. Fischer, and A. Hotho. 2025. "Assessing the state of the art in scene segmentation." In *Proceedings of the NAACL 2025*. https://2025.naacl.org/program/accepted_papers/

Zehe, A., L. Konle, S. Guhr, L. Dümpelmann, E. Gius, A. Hotho, F. Jannidis, L. Kaufmann, M. Krug, F. Puppe, N. Reiter, and A. Schreiber, eds. 2021b. *Proceed. of the Shared Task on Scene Segmentation*, CEUR, KONVENS. https://ceur-ws.org/Vol-3001/

https://github.com/literarylab/scene_segmentation

fortext lab · Computational Literary Studies · Literary Lab