



# Making BERT Feel at Home

Modelling Domestic Space in 19th-Century British and Irish Fiction



Svenja Guhr, Jessica Monaco,  
Alexander J. Sherman, Matt Warner,  
and Mark Algee-Hewitt

*fortext lab*



## Theoretical Background



Evans, Elizabeth, and Matthew Wilkens. "Nation, Ethnicity, and the Geography of British Fiction, 1880-1940." *Journal of Cultural Analytics*, 2018, 1–48. 10.22148/16.024.

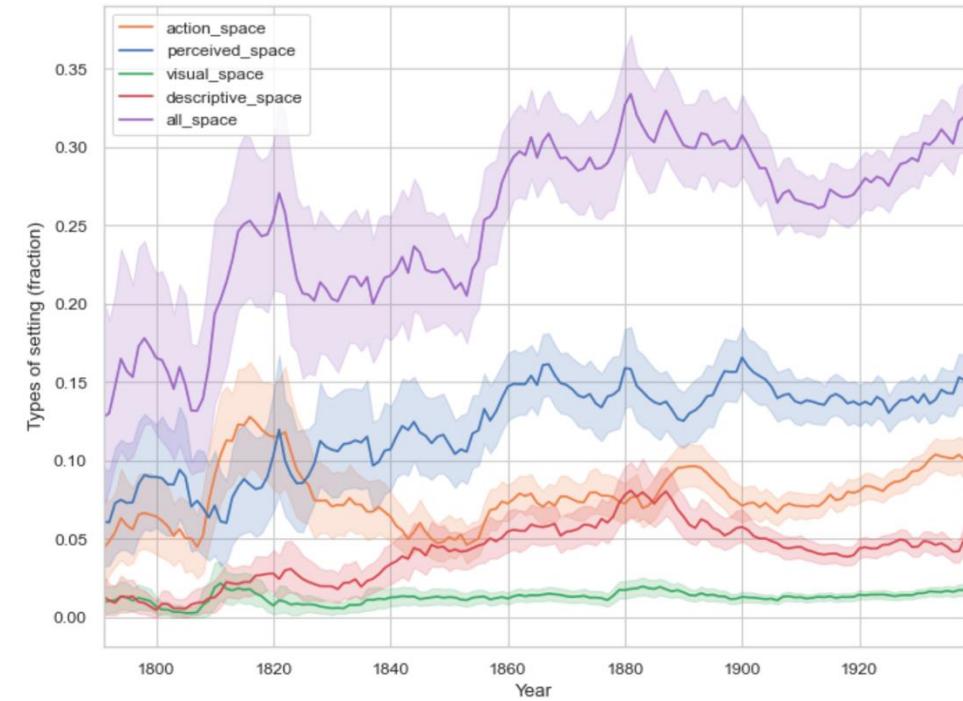
#### Examples of words for each category of NNSE (Non-Named Spatial Entities)

Category	Exampels
interior	<i>Abstellkammer</i> (storage room), <i>Wohnzimmer</i> (living room), <i>Küche</i> (kitchen)
urban	<i>Bibliothek</i> (library), <i>Kloster</i> (abbey), <i>Vorstadt</i> (suburb)
rural	<i>Bauernhaus</i> (farmhouse), <i>Garten</i> (garden), <i>Schweinestall</i> (pigsty)
natural	<i>Berg</i> (mountain), <i>Fluss</i> (river), <i>Wald</i> (forest)

Kababgi, Daniel, Giulia Grisot, Federico Pennino, and Berenike Herrmann (2024). "Recognising Non-named Spatial Entities in Literary Texts: A Novel Spatial Entities Classifier". In: *CHR 2024*. Vol. 3834, 472–481.

Entity type	Count	Examples
PER	9,383	my mother, Jarndyce, the doctor, a fool, his companion
FAC	2,154	the house, the room, the garden, the drawing-room, the library
LOC	1,170	the sea, the river, the country, the woods, the forest
GPE	878	London, England, the town, New York, the village
VEH	197	the ship, the car, the train, the boat, the carriage
ORG	130	the army, the Order of Elks, the Church, Blodgett College

Bamman, David, Sejal Popat, and Sheng Shen. "An Annotated Dataset of Literary Entities." In *Proceedings of the 2019 Conference of NAACL*, 2138–44. ACL Minneapolis, 2019. 10.18653/v1/N19-1220.



Katrin Rohrbacher (2025). "Opening Worlds: Narrative Beginnings and the Role of Setting". In: *CCLS2025 Preprints* 4 (1). 10.26083/tuprints-00030149.

# Challenge

- Spatial setting often mentioned only once  
e.g., often no space indication during long dialogues  
→ the bigger picture is missing
- Implicit domesticity is hard to detect using rule-based or keyword methods

“How likely is it that the passage is set in domestic space?”



# Our Approach

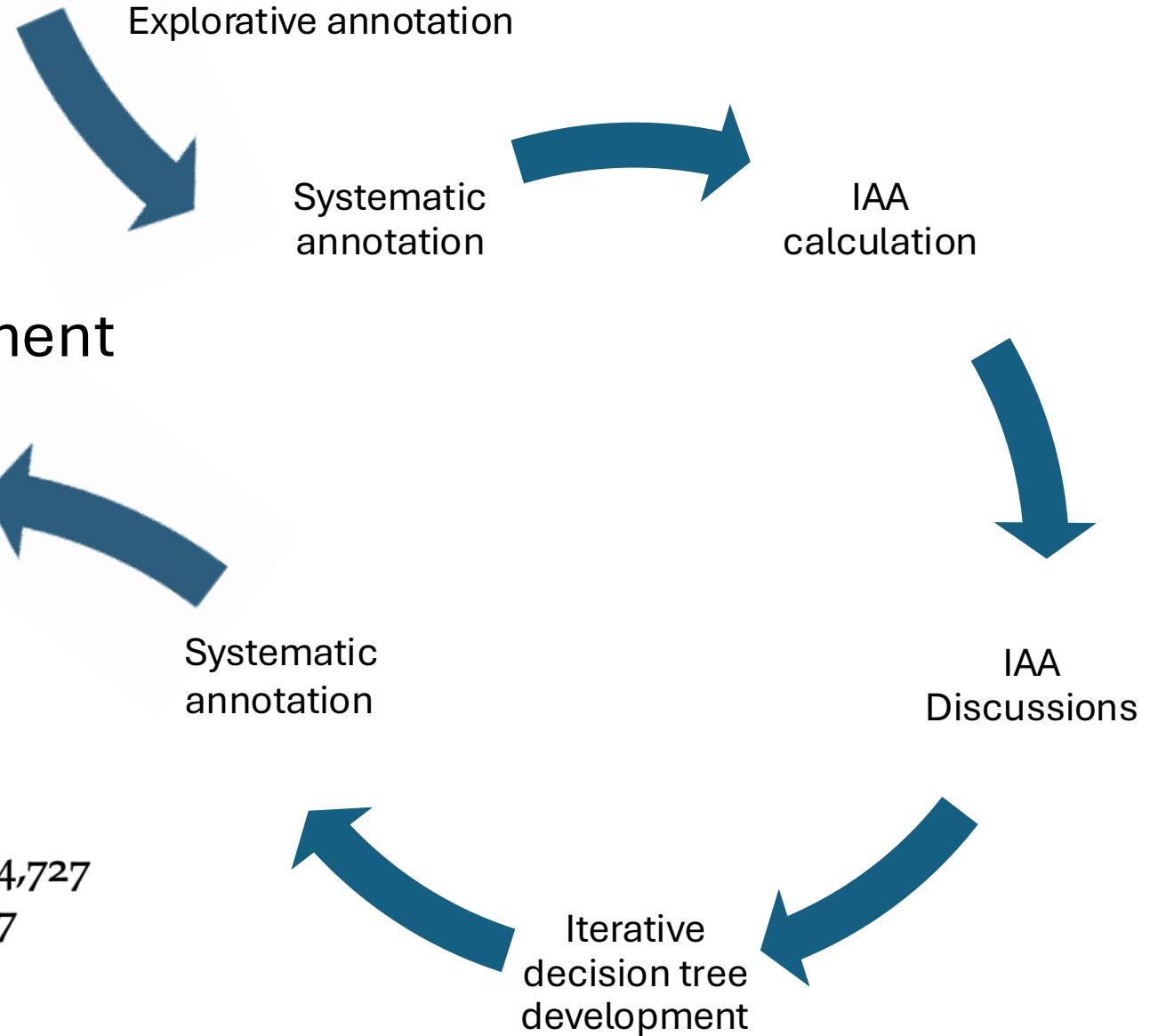


- “*How likely is this passage set in domestic space?*”
- Shift from binary classification to *probabilistic modeling*
- Introduce a **domesticity score** using fine-tuned BERT model
- Captures both **explicit and implicit** markers of domesticity
- Reveals **liminal domestic spaces** (e.g., gardens, carriages, ships)

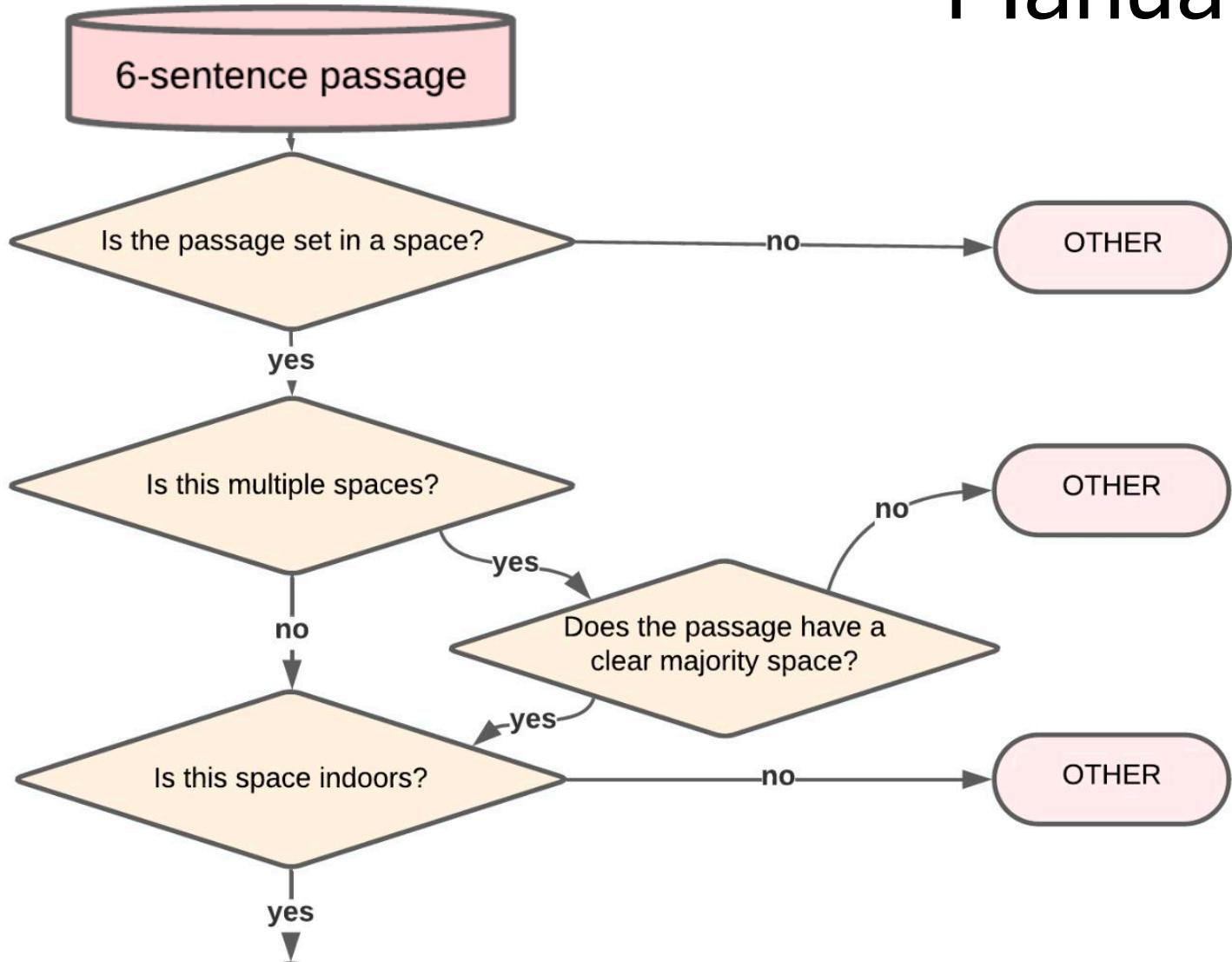
# Operationalization

- Iterative guideline development
- 3 categories:
  - Trash
  - Domestic space
  - Other

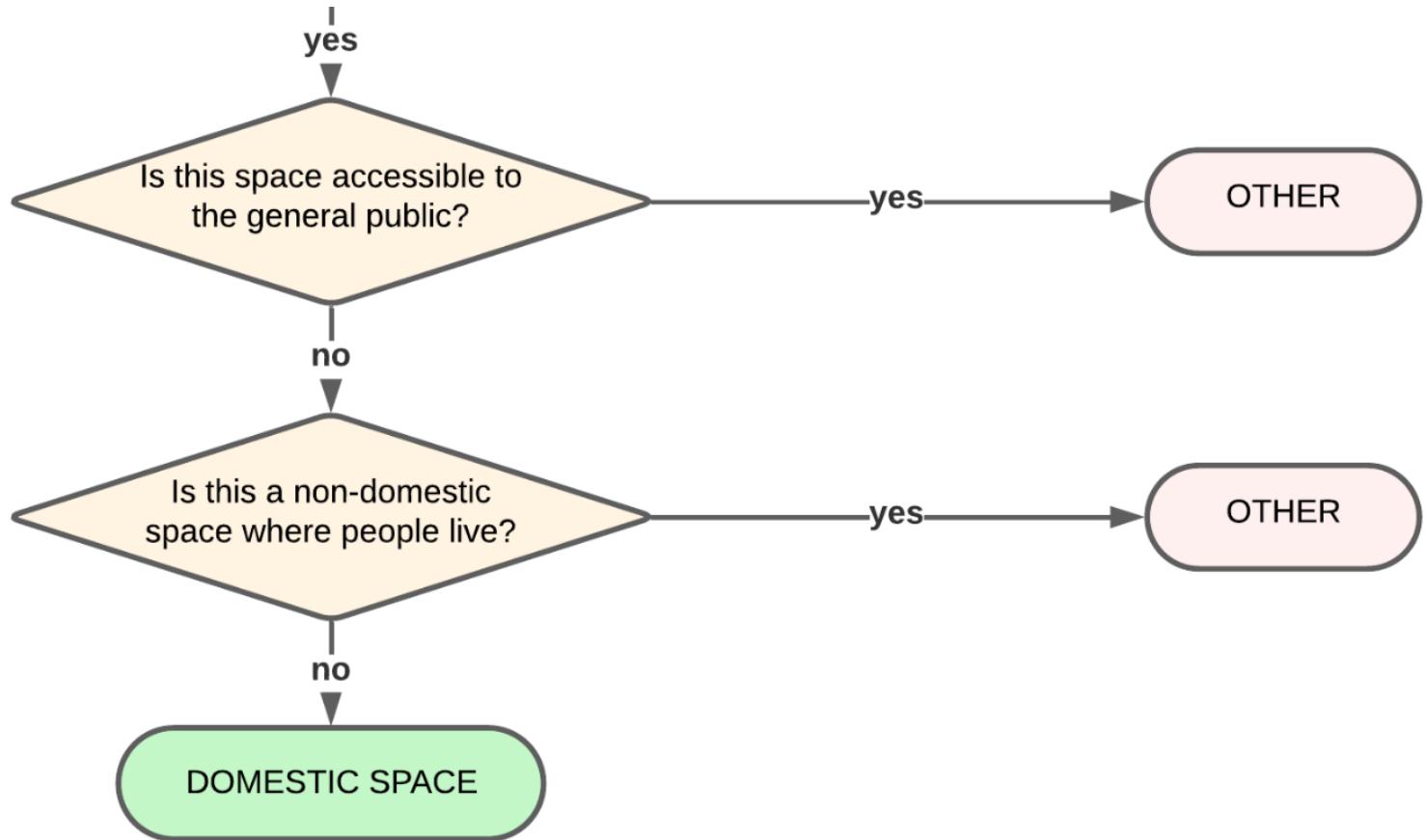
Six-sentence passages	3,684,727
Manually annotated passages	1,227
“domestic space” passages	521
“other” passages	678
“trash” passages	28



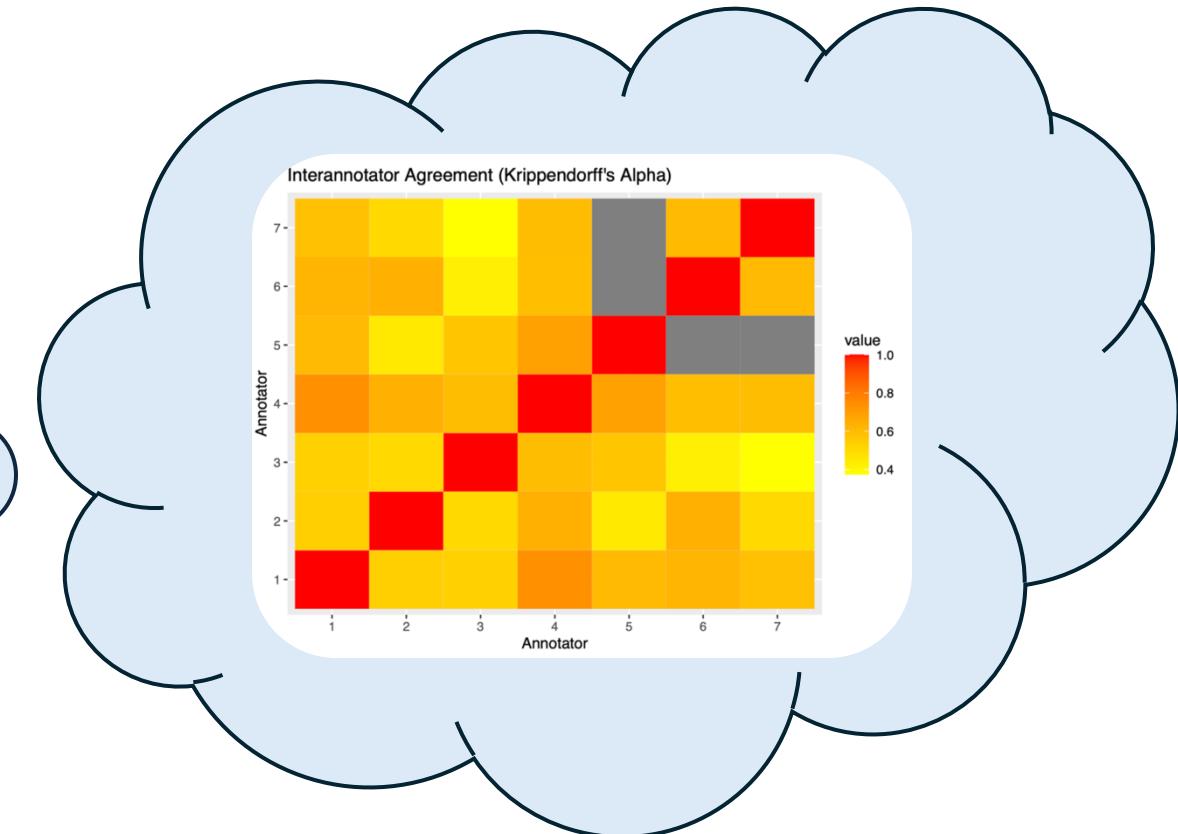
# Manual Annotation



# Manual Annotation: Domestic Space



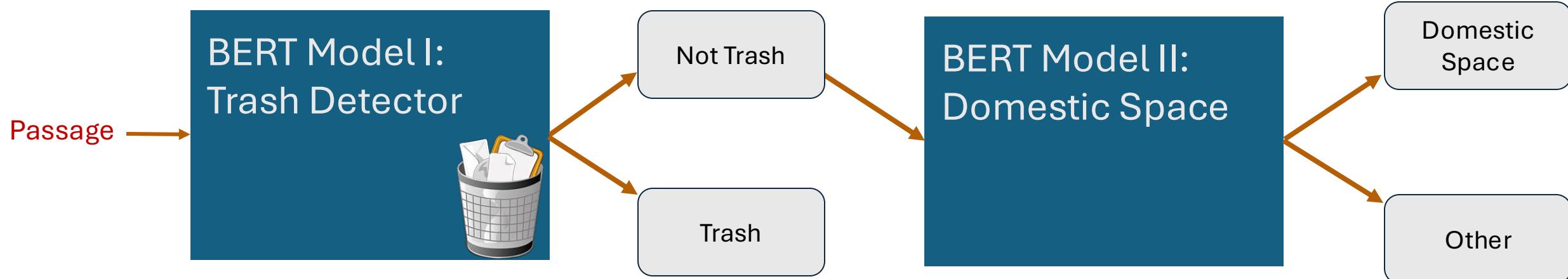
# Groundtruth for Literary Fiction?



# Automation: Making BERT Feel at Home

- BERT-family model USE for sequence classification task with contextual sentence embeddings
- Majority agreement for single classification
- 2 model approach:

	1.	2.
Trash Detector	Domestic Space Prediction	
Accuracy	0.8159	1.00
Recall	0.8105	1.00
Precision	0.8092	1.00
F1-score	0.8097	1.00
Loss	0.4288	0.0012



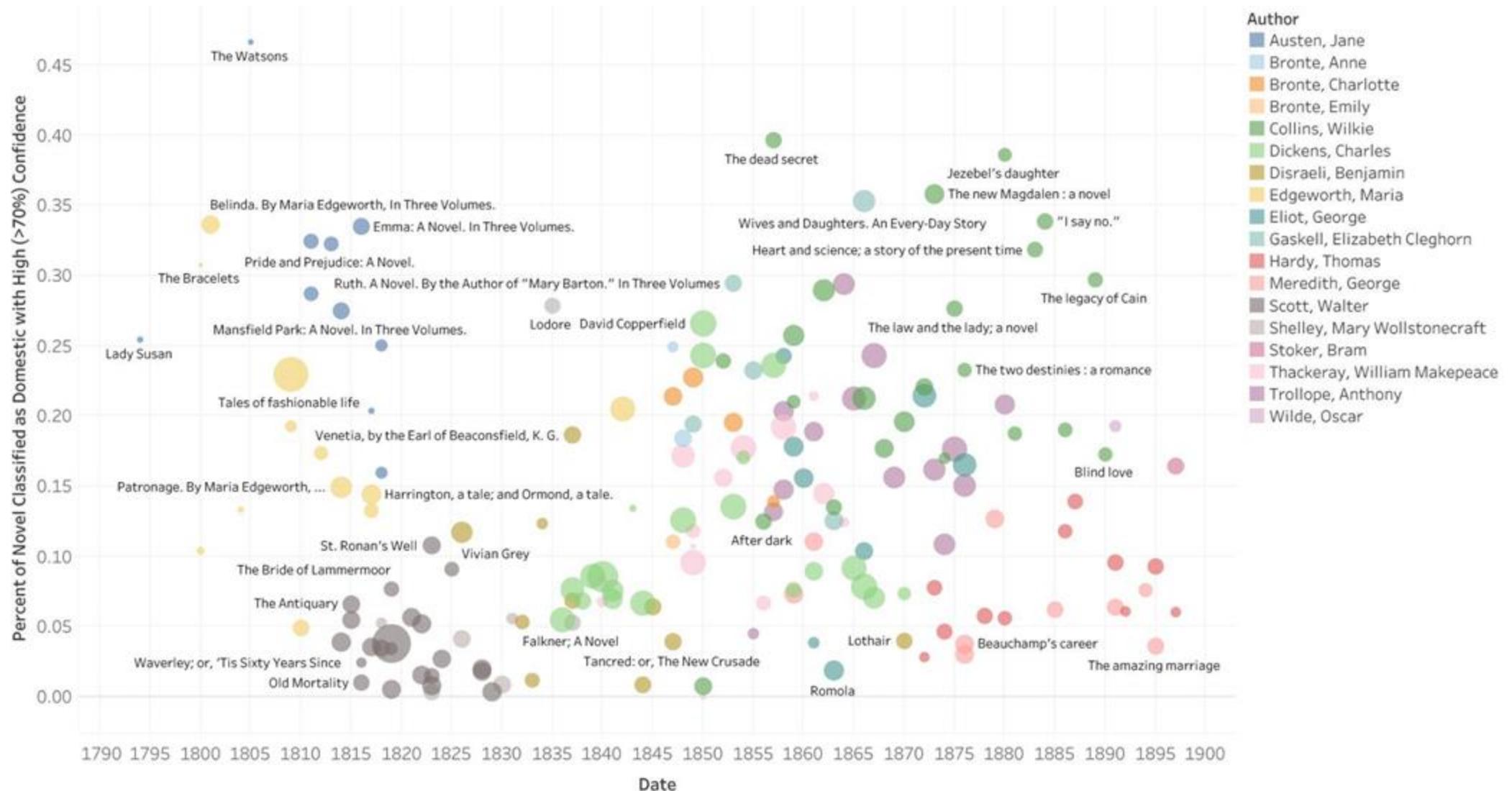
Even if blurry don't worry!



# Domestic Space in 19<sup>th</sup>-Century British and Irish Fiction



# Domestic Space in Texts by Canonic Authors



# Summary and Outlook

- Focus shift from keyword-based prediction to detecting implicit domestic space
- Six-sentence segments balance context and scalability  
→ though the approach remains heuristic
- Current and future work:
  - Domestic space prediction for scene segments
  - Ongoing: Prediction of automatically scene-segmented 20<sup>th</sup> century US romance novels
  - Future: Discover what actually happens in domestic spaces by looking at character behavior and gender distribution



# References

## Images:

Muppet Fandom: [https://muppet.fandom.com/wiki/Muppet\\_Wiki](https://muppet.fandom.com/wiki/Muppet_Wiki).

## Bibliography:

Baledent, Anaëlle, Yann Mathet, Antoine Widlöcher, Christophe Couronne, and Jean-Luc Manguin (2022). “Validity, Agreement, Consensuality and Annotated Data Quality”. In: *Proceedings of the 13th LREC*, 2940–2948. <https://aclanthology.org/2022.lrec-1.315>.

Bamman, David, Sejal Popat, and Sheng Shen. “An Annotated Dataset of Literary Entities.” In; *Proceedings of the 2019 NAACL*, 2138–44. ACL, 2019. 10.18653/v1/N19-1220.

Cer, Daniel et al. (2018) “Universal Sentence Encoder for English”, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, ACL, 169–174. 10.18653/v1/D18-2029.

Chadwyck-Healey Literature Collections and ProQuest (no date) “Nineteenth-Century British Fiction”. [https://collections.chadwyck.com/marketing/products/about\\_ilc.jsp?collection=ncf](https://collections.chadwyck.com/marketing/products/about_ilc.jsp?collection=ncf).

Evans, Elizabeth, and Matthew Wilkens. “Nation, Ethnicity, and the Geography of British Fiction, 1880-1940.” *Journal of Cultural Analytics*, 2018, 1–48. 10.22148/16.024.

Kababgi, Daniel, Giulia Grisot, Federico Pennino, and Berenike Herrmann (2024). “Recognising Non-named Spatial Entities in Literary Texts: A Novel Spatial Entities Classifier”. In: CHR 2024. Vol. 3834, 472–481.

Rohrbacher, Katrin (2025). “Opening Worlds: Narrative Beginnings and the Role of Setting”. In: CCLS2025 Preprints 4 (1). 10.26083/tuprints-00030149.



# Making BERT Feel at Home

Modelling Domestic Space in 19th-Century British and Irish Fiction



Svenja Guhr, Jessica Monaco,  
Alexander J. Sherman, Matt Warner,  
and Mark Algee-Hewitt

*fortext lab*

# Questions to Discuss with the Community

- What does "groundtruth" mean for CLS?
  - With LLMs, is it better to have a small amount of well-curated data, such as groundtruth data, or are gold or silver annotations sufficient?
  - What is your experience with silver annotations?

# Bibliography

- Algee-Hewitt, Mark / Porter, J.D. / Walser, Hannah (2020) "Representing Race and Ethnicity in American Fiction, 1789-1920", *Journal of Cultural Analytics*, 5(2). Available at: <https://doi.org/10.22148/001c.18509>.
- Armstrong, Nancy (1987) *Desire and Domestic Fiction: A Political History of the Novel*. Oxford: Oxford University Press.
- Auerbach, Erich (2003) *Mimesis: The Representation of Reality in Western Literature*. Translated by William R. Trask. Princeton, New Jersey: Princeton University Press.
- Bachelard, Gaston (1994) *The poetics of space*. Translated by Maria Jolas. Boston: Beacon Press.
- Bakhtin, M.M. (1981) "Forms of Time and of the Chronotope in the Novel: Notes toward a Historical Poetics", in Michael Holquist (ed.), Caryl Emerson and Michael Holquist (trans.) *The Dialogic Imagination: Four Essays*. Austin: University of Texas Press (University of Texas Press Slavic Series, 1), pp. 84–258.
- Bologna, Federica (2020) "A Computational Approach to Urban Space in Science Fiction", *Journal of Cultural Analytics*, 5(2). Available at: <https://doi.org/10.22148/001c.18120>.
- Bushell, Sally et al. (2017) *Chronotopic Cartographies*, *Chronotopic Cartographies*. Available at: <https://www.lancaster.ac.uk/chronotopic-cartographies/>.
- Cer, Daniel et al. (2018) "Universal Sentence Encoder for English", in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Brussels, Belgium: Association for Computational Linguistics, pp. 169–174. Available at: <https://doi.org/10.18653/v1/D18-2029>.
- Chadwyck-Healey Literature Collections and ProQuest (no date) "Nineteenth-Century British Fiction". Available at: [https://collections.chadwyck.com/marketing/products/about\\_ilc.jsp?collection=ncf](https://collections.chadwyck.com/marketing/products/about_ilc.jsp?collection=ncf).
- Chollet, François (2015) "keras".
- Cohen, Monica F. (2017) "Domesticity in Victorian Literature", in Cohen, Monica F., *Oxford Research Encyclopedia of Literature*. Oxford University Press. Available at: <https://doi.org/10.1093/acrefore/9780190201098.013.252>.

# Bibliography

- Cuming, Emily (2013) “‘Home is home be it never so homely’: Reading Mid-Victorian Slum Interiors”, *Journal of Victorian Culture*, 18(3), pp. 368–386. Available at: <https://doi.org/10.1080/13555502.2013.826424>.
- David, Deirdre (2002) “Empire, Race, and the Victorian Novel”, in P. Brantlinger and W.B. Thesing (eds) *A companion to the Victorian novel*. Oxford, UK ; Malden, MA: Blackwell (Blackwell companions to literature and culture), pp. 84–100.
- De Swert, Knut (2012) “Calculating inter-coder reliability in media content analysis using Krippendorff’s Alpha”, *Center for Politics and Communication*, 15, pp. 1–15.
- Devlin, Jacob et al. (2019) “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, in *Proceedings of the 2019 Conference of the North. Proceedings of the 2019 Conference of the North*, Minneapolis, Minnesota: Association for Computational Linguistics, pp. 4171–4186. Available at: <https://doi.org/10.18653/v1/N19-1423>.
- Evans, Elizabeth / Wilkens, Matthew (2018) “Nation, Ethnicity, and the Geography of British Fiction, 1880–1940”, *Journal of Cultural Analytics*, pp. 1–48. Available at: <https://doi.org/10.22148/16.024>.
- Gilbert, Sandra M. / Gubar, Susan (2000) *The madwoman in the attic: the woman writer and the nineteenth-century literary imagination*. 2nd ed. New Haven: Yale University Press.
- Grisot, Giulia / Herrmann, Berenike (2023) “Examining the representation of landscape and its emotional value in German-Swiss fiction between 1840 and 1940”, *Journal of Cultural Analytics*, 8(1). Available at: <https://doi.org/10.22148/001c.84475>.
- Heuser, Ryan / Le-Khac, Long (2012) *A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method*. 4. Stanford Literary Lab. Available at: <https://litlab.stanford.edu/LiteraryLabPamphlet4.pdf>.
- Jameson, Fredric (1991) *Postmodernism, or, The Cultural Logic of Late Capitalism*. Durham: Duke University Press.
- Krippendorff, Klaus (2018) *Content analysis: an introduction to its methodology*. Fourth Edition. Los Angeles: SAGE.
- Mani, Inderjeet et al. (2010) “SpatialML: annotation scheme, resources, and evaluation”, *Language Resources and Evaluation*, 44(3), pp. 263–280. Available at: <https://doi.org/10.1007/s10579-010-9121-0>.
- Marcus, Sharon (2009) “At Home with the Other Victorians”, *South Atlantic Quarterly*, 108(1), pp. 119–145. Available at: <https://doi.org/10.1215/00382876-2008-026>.

# Bibliography

McKeon, Michael (2005) *The Secret History of Domesticity: Public, Private, and the Division of Knowledge*. Baltimore: Johns Hopkins University Press.

Meyer, Susan (1996) *Imperialism at home: race and Victorian women's fiction*. Ithaca, N.Y: Cornell University Press (Reading women writing).

Piatti, Barbara (2016) "Mapping Fiction: The Theories, Tools and Potentials of Literary Cartography", in D. Cooper, C. Donaldson, and P. Murrieta-Flores (eds) *Literary mapping in the digital age*. London ; New York: Routledge (Digital research in the arts and humanities), pp. 88–101.

Reiter, Nils (2020) "Anleitung zur Erstellung von Annotationsrichtlinien", in Nils Reiter, Axel Pichler, and Jonas Kuhn (eds) *Reflektierte algorithmische Textanalyse*. De Gruyter, pp. 193–202. Available at: <https://doi.org/10.1515/9783110693973-009>.

Schumacher, Mareike K. (2023) *Orte und Räume im Roman: Ein Beitrag zur digitalen Literaturwissenschaft*. Berlin, Heidelberg: Springer Berlin Heidelberg (Digitale Literaturwissenschaft). Available at: <https://doi.org/10.1007/978-3-662-66035-5>.

Soni, Sandeep et al. (2013) "Grounding Characters and Places in Narrative Texts", in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto. Available at: <http://arxiv.org/abs/2305.17561> (Accessed: 9 December 2023).

Spivak, Gayatri Chakravorty (1985) "Three Women's Texts and a Critique of Imperialism", *Critical Inquiry*, 12, pp. 243–61. Available at: <https://doi.org/10.7916/D8KD1WSW>.

Tally, Robert T., Jr. (2013) *Spatiality*. New York: Routledge.

Taylor, Joanna E. / Gregory, Ian N. (2022) *Deep Mapping the Literary Lake District: A Geographical Text Analysis*. Rutgers University Press. Available at: <https://doi.org/10.36019/9781684483792>.

Wagner, Tamara S. (2021) "Laboring in a Place of Rest: Reconsidering Victorian Domestic Work and Its Narratives", *Victorians: A Journal of Culture and Literature*, 140(1), pp. 103–114. Available at: <https://doi.org/10.1353/vct.2021.0011>.

Wall, Cynthia S. (2006) *The Prose of Things: Transformations of Description in the Eighteenth Century*. Chicago: University of Chicago Press.

Wilkens, Matthew (2013) "The Geographic Imagination of Civil War-Era American Fiction", *American Literary History*, 25(4), pp. 803–840.

Xu, Jun (2011) "Austen's fans and fans' Austen", *JLSE*, 40(1), pp. 81–97. Available at: <https://doi.org/10.1515/jlse.2011.005>.

Zoran, Gabriel (1984) "Towards a Theory of Space in Narrative", *Poetics Today*, 5(2), pp. 309–35.

# Corpus

- 19<sup>th</sup>-Century British and Irish novels
- Sourced from the Univ. of Illinois libraries and Chadwyck-Healey Nineteenth-Century Fiction collection (Chadwyck-Healey Literature Collections and ProQuest 2024)
- Canonical and lesser-known fiction
- Publication dates and author information for diachronic and comparative analysis

Texts	2,865
Words in total	557,097,804
Individual authors (+ 126 “unknown”)	1,250
British authors	1,226
Irish authors	24
Texts by Irish authors	118
Time period	1748–1899
Six-sentence passages	3,684,727
Manually annotated passages	1,227
“domestic space” passages	521
“other” passages	678
“trash” passages	28

# Research Question

→ Instead of asking: “Is this space domestic?”,  
we ask: “How likely is it that the passage is set in domestic space?”

Are there other indicators besides explicit spatial markers like “home,” “house,” or “living room” that can be used to detect if a passage is set in a “home”?

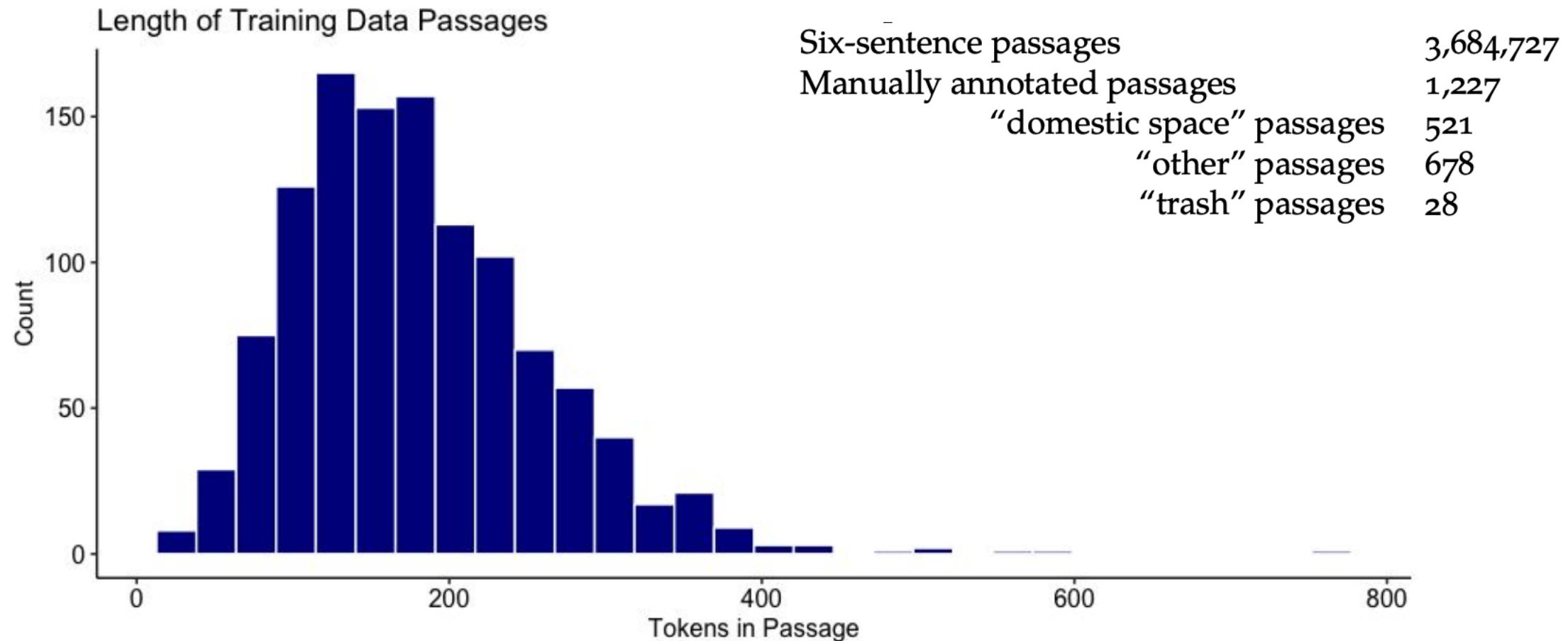
# Text Segmentation: 6-Sentence Segments

Heuristic approach dating back to 2022

- Chapters too broad, sentences too narrow
- Paragraphs too inconsistently varying in length, mark-up needed
- Six-sentence passages
  - Balanced context and granularity
  - Consistent input for annotators & BERT
  - Efficient for both:  
manual tagging and model training



# 6-Sentence Passages



# Evaluation Results

	1. Trash Detector	2. Domestic Space Prediction
Accuracy	1.00	0.8159
Recall	1.00	0.8105
Precision	1.00	0.8092
F1-score	1.00	0.8097
Loss	0.0012	0.4288

# Methods: Trash Detection

Our model is particularly good at tagging trash:

Precision: 1.00   Recall: 1.00      F1: 1.00

Example passages judged by the model to be trash with the highest probability:

Illustrated by the Author, i vol. crown Svo. 35-. CHRISTMAS BOOKS. Illustrated by the Author and.  
Richard DovLE. I vol. crown Svo. 3J. BURLESQUES.

0.9985 probability of being Trash

ESSAYS. 2 Vols. I. Classical. II. Modern. Crown Bvo. 4<sup>A</sup>. 6d. Each.

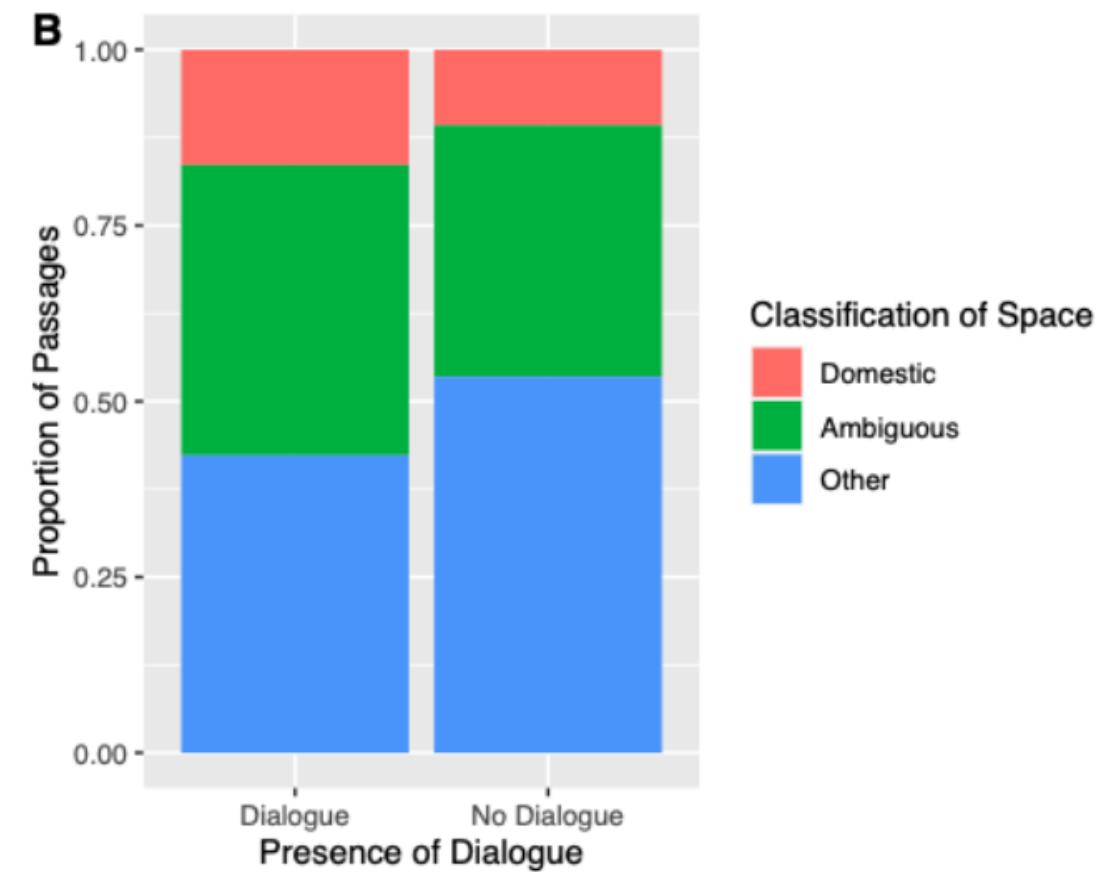
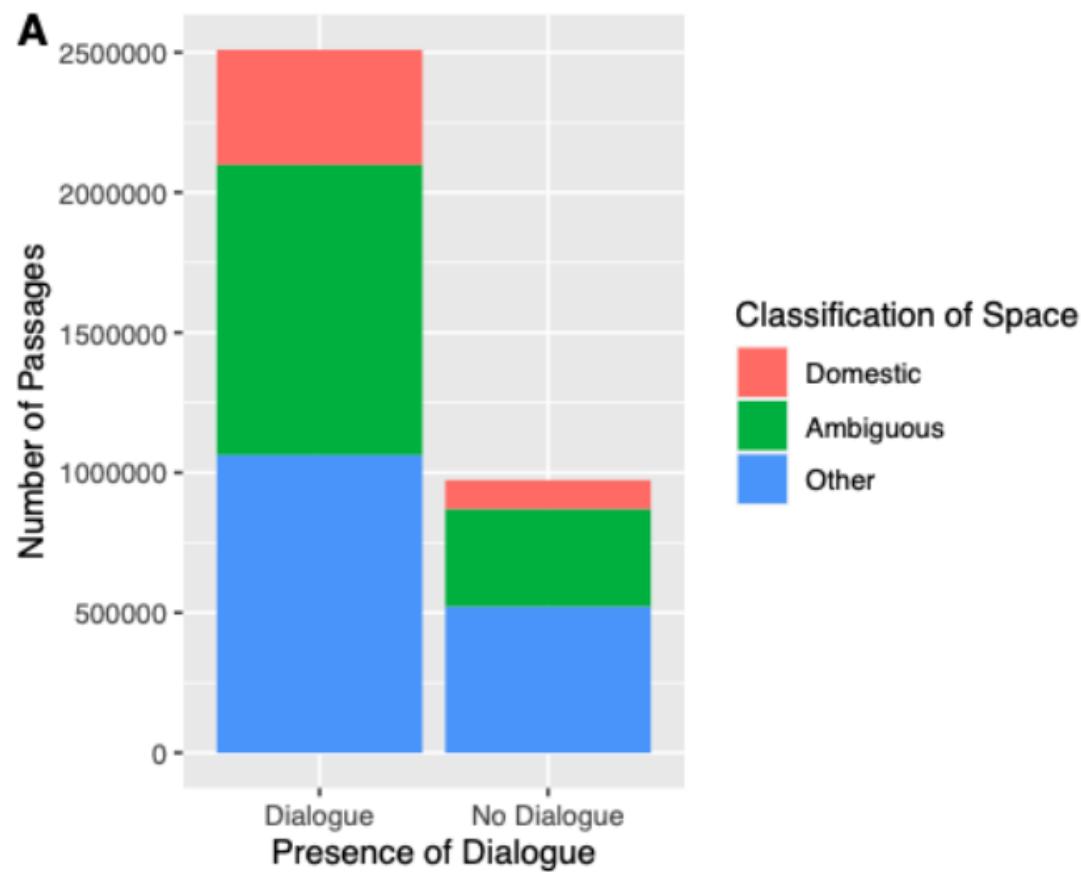
0.9984 probability of being Trash

# Example for a passage labeled as "domestic"

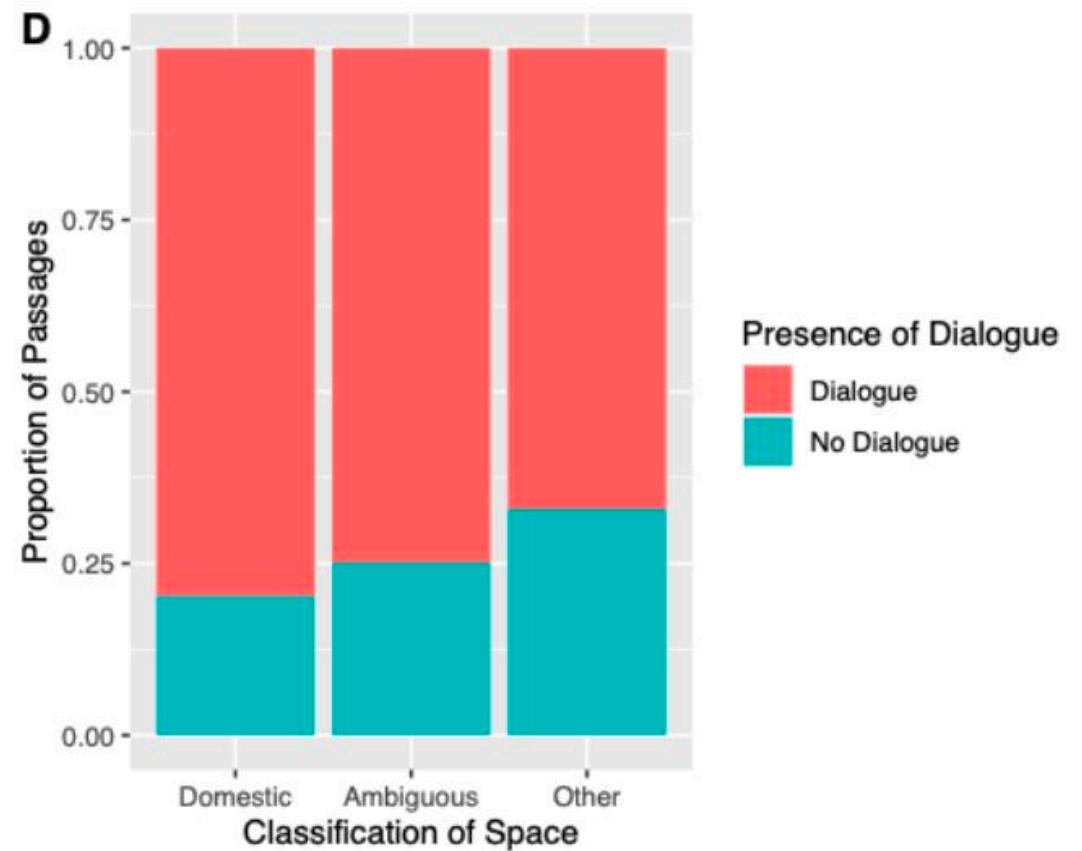
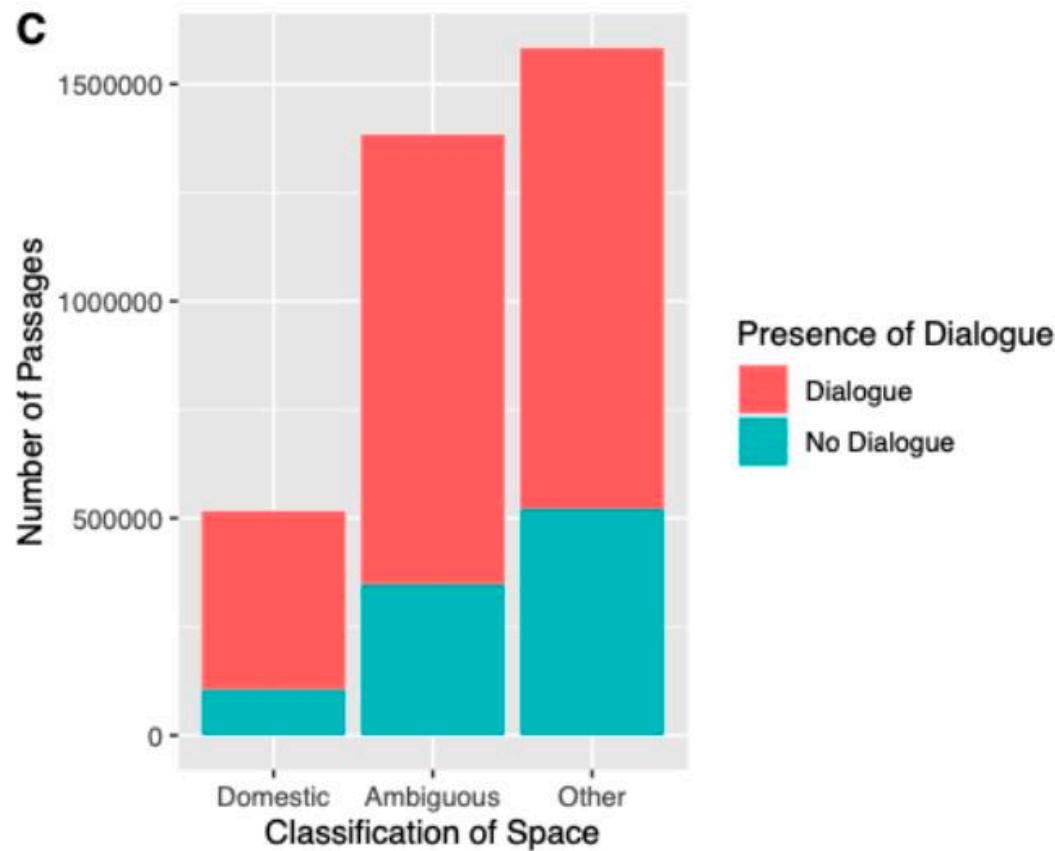
"It is warm and mild now, and we shall be back in time for luncheon, I will just get my hat." He went into his bedroom as he spoke, and after a moment came back with his hat in his hand. John had left the room and was standing just outside the door. As Sir Lionel came through the sitting-room, he watched him furtively, but closely; and as soon as he was fairly in the corridor, John shut the door, and, forgetting his usual deference, led the way briskly through the porch. They walked towards the gardens; but presently John said: "I fear you will have some further trouble with James, I hope he will go this afternoon." "I hope so, these scenes of howling and supplicating are very tiresome." "

- passage from Riding out the Gale by Annette Lyster
- partly outside of the home "walked towards the gardens"

# Proportions of Dialogue in Domestic Space



# Presence of Dialogue in Domestic Space



## Domestic Space and Canonicity

Percent of Novel Classified as Domestic with High (>70%) Confidence

1800 1825 1850 1875 1900

Date

### Author

- Austen, Jane
- Bronte, Anne
- Bronte, Charlotte
- Bronte, Emily
- Collins, Wilkie
- Dickens, Charles
- Disraeli, Benjamin
- Edgeworth, Maria
- Eliot, George
- Lewis, M. G. (Matthew Gregory)
- Maturin, Charles Robert
- Radcliffe, Ann Ward
- Scott, Walter
- Shelley, Mary Wollstonecraft
- Stoker, Bram
- Thackeray, William Makepeace
- Trollope, Anthony
- Wilde, Oscar