

UNDERSTANDING NARRATIVE

Redefining the frame

Eva Portelance (B.A. Hons Linguistics) and
Andrew Piper (Associate Professor and Director of .txtLab@McGill)

PROJECT GOAL

- Understand the nature of scene/frame changes in literary narratives with the help of computational methods.
- Predict scene/frame changes to study the rhythm of narrative progression and plotline distribution.

ABSTRACT

- Understanding narrative structure at large scale remains a challenging problem within the field of cultural analytics and computational linguistics. Our aim with this project is to develop novel methods to study the pacing of narrative scene changes and the overall distribution of different plotlines within novels. Being able to analyze such narrative features at large scale can give us insights into the way different genres, time periods, or cultures favor different modes of storytelling. In this project we formalize definitions of narrative scenes and implement new methods of detecting them using computational methods.

WHAT IS A NARRATIVE ?

- A formal definition from the *Dictionary of Narratology*, by Gerald Prince, 1987:
 - “The recounting of one or more real or fictitious **EVENTS** communicated by one, two, or several (more or less overt) **NARRATORS** [...]”Revised in 2003:
 - “...narrative is essentially a mode of verbal presentation and involves the **LINGUISTIC** recounting or telling of **EVENTS**.”
- This can be reduced to two fundamental definitions:
 - **Narrative Discourse (Sujet) = HOW the EVENTS are recounted**
 - **Story (Fabula) = WHAT the EVENTS recounted are.**

At its most bare form, what are we really looking at in a narrative?

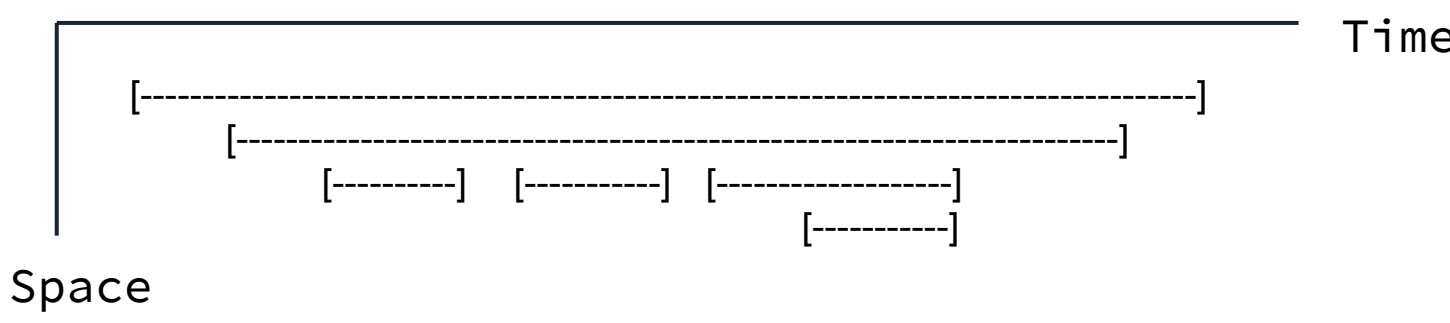
A SERIES OF EVENTS

COMPUTERIZING THIS THOUGHT PROCESS

- The first step: define the variables which compose an event.
 - A simple definition from semantics:
EVENT = relationship between an AGENT of an event (VERB) and its THEME (Object)
 - For it to be have a truth value, it must have a domain of discourse (D), or world:
D = Entities x Time x Space
- HYPOTHESIS:** a frame is a sequence of events within its own domain of discourse. Therefore, a frame boundary would consist in a change in the three components of D.
- Algorithm breakdown:
 - Uses parts of speech (POS) to determine in what category a content word belongs.
 - Compares sets of words from adjacent passages of a narrative.
 - Records the level of dissimilarity over a determined threshold as frame boundaries.

WHAT MAKES A FRAME ?

- The classic definition of a frame narrative: a story told within a story, e.g. Arabian Nights, represented below.
 - However, frame narratives can be considered more of a GENRE.
- If we want to include **ALL** narratives, we need much broader definition.
 - A parallel between the SCENE in theater and the FRAME in literature
scene change = disruption in the sequencing of events > FRAME BOUNDARY



RESULTS

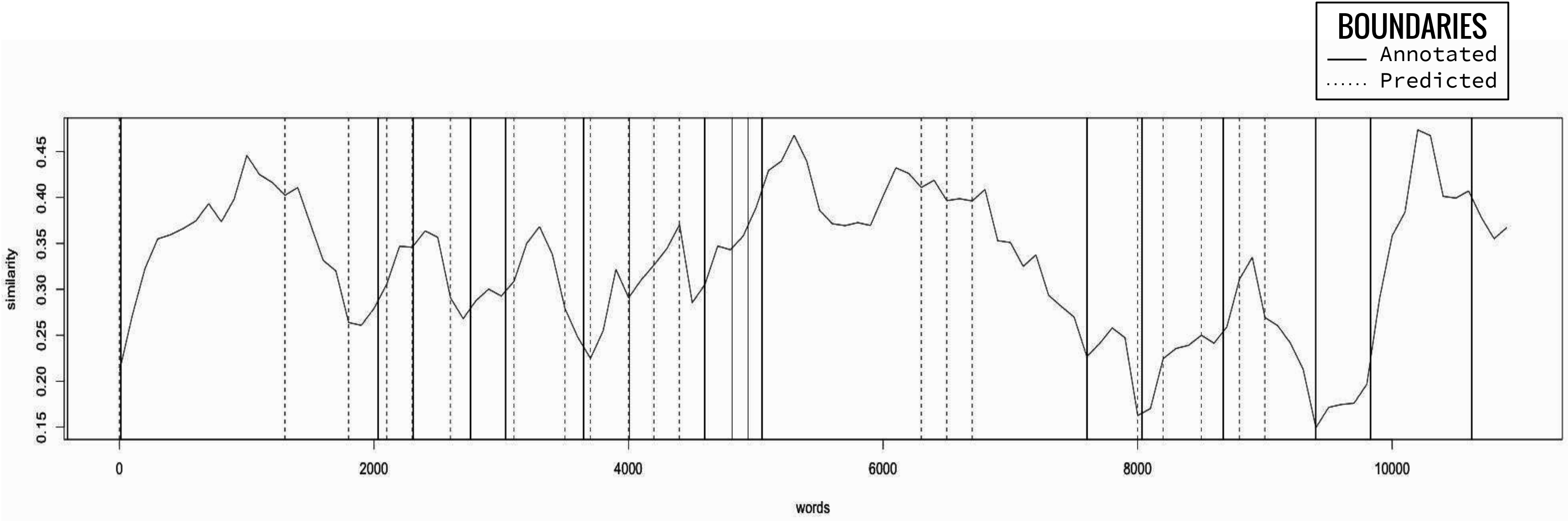
- DATA SET of 12 000 word passages from 9 books
Arabian Nights, Frankenstein (1818), Wuthering Heights (1850), A Tale of Two Cities (1859), The Kreutzer Sonata (1890), Heart of Darkness (1899), Ethan Frome (1911), Catch-22 (1961), A Dance with Dragons (2011)
- FRAME TYPES:
Books were hand-annotated for:
 - character change
 - setting change
 - time change
 - parallel plot
 - flashback or flash forward
 - narrator change
 - narrative digression

- PERFORMANCE:
 - Hearst's Texttiling algorithm from 1994-97 (HST)
 - Character driven model (NNP)
 - Time and Setting model (ALL)
 - combination of NNP and ALL (COMB)Result over all annotated frames (9 books)

Models	Precision	Recall	F1 Accuracy
HST	0.182320442	0.1907514451	0.186440678
NNP	0.7037037037	0.4393063584	0.5409252669
ALL	0.7054263566	0.5260115607	0.6026490066
COMB	0.6231884058	0.7456647399	0.6789473684

DETECTING FRAME BOUNDARIES

- Best performance:
 - 2000 word rolling windows, with a step of 100 words for choosing sets of words
 - measure similarity using a cosine similarity test
 - select local minimums under the 95 percentile confidence interval as frame boundaries
 - **COMBINATION** of frame boundaries predicted from changes in the set of proper names (i.e. Entities) and those of all content words (exclude stop words and includes proper names), mostly verbs and other contentful nouns (i.e. Time and Setting).
- Our highest performing model was able to successfully detect all frame types with 68% accuracy within +/- 200 words.
- The following graph shows the predicted and actual frame changes in Dickens' *A Tale of Two Cities* (1859) based on sliding window of cosine similarity between text chunks



PLOT DISTRIBUTION

- “In this portion we try to predict the overall number of plotlines in a novel based on scene divisions.”
-
- Clustering of frames from R.R. Martin's *A Dance with Dragons* (2011) into two distinct plotlines
 - The coloured numbers indicate actual plot groups, while the coloured leaves indicate predicted plot groups.
 - Uses Ward's method of hierarchical clustering

WHAT'S NEXT

- Comparing our results to human agreement on these same tasks
- Studying narrative on a larger scale, comparing genres, eras and authors for rhythmic patterns in storyline progression and plot distribution