Word Category Arcs in Literature Across Languages and Genres

Winston Wu **Lu Wang** **Rada Mihalcea**
Computer Science and Engineering
University of Michigan
{wuws,wangluxy,mihalcea}@umich.edu

Abstract

Word category arcs measure the progression of word usage across a story. Previous work on arcs has explored structural and psycholinguistic arcs through the course of narratives, but so far it has been limited to *English* narratives and a narrow set of word categories covering binary emotions and cognitive processes. In this paper, we expand over previous work by (1) introducing a novel, general approach to quantitatively analyze word usage arcs for any word category through a combination of clustering and filtering; and (2) exploring narrative arcs in literature in eight different languages across multiple genres. Through multiple experiments and analyses, we quantify the nature of narratives across languages, corroborating existing work on monolingual narrative arcs as well as drawing new insights about the interpretation of arcs through correlation analyses.

1 Introduction

Throughout history, the narrative has been an essential medium for communicating and transferring information. The study of the structure of narratives has roots in the ancient Greek philosophers but did not gain much interest until the last few hundred years. One of the most well-known structures is Freytag's pyramid, the dramatic arc of German novelist and playwright Gustav Freytag (1894), which contains five stages: exposition, rising action, climax, falling action, and resolution. Many others have hypothesized sets of universal structures into which all narratives can be classified. For example, Foster-Harris (1959) argued that a story has three basic plots that end with a happy, unhappy, or tragic ending. Booker (2004) proposed seven basic plots: overcoming the monster, rags to riches, the quest, voyage and return, comedy, tragedy, and rebirth. Others have posited 20 plots (Tobias, 2012) and even 36 plots (Polti, 1917) that are universal across great stories.

Regardless of the actual number of different plots, one point is clear: the structures of plots naturally vary. A story's structure gives coherence to the entire plot and can be mathematically represented as a function over time, or a *narrative arc*. The American writer Kurt Vonnegut claimed in his famously rejected master's thesis (1947) that every story can be plotted as such a curve, where the x-axis is the duration of the story, and the y-axis is a character's "Ill Fortune – Great Fortune" (Vonnegut, 1999). This was a revolutionary notion at the time and only recently has been computationally investigated. Following existing computational work (Mohammad, 2011; Reagan et al., 2016; Boyd et al., 2020), we consider a *narrative arc* as a measure of word usage (count) across a story. We use the term *word category arc* to emphasize that this arc is measured by examining words that belong to certain categories. This count may be z-score standardized to better understand the relative usage of certain words across a story. Thus, an arc provides a high-level structural overview of a narrative.

All cultures tell stories, but the manner in which the stories are told differs. Narrative arcs are one method for quantifying the cultural differences in stories. We first describe a general framework for analyzing arcs in a narrative that follows closely from Vonnegut's claim. To compute arcs, we measure the usage of words in a given word category, such as positive emotion words in LIWC (Pennebaker et al., 2015), a popular dictionary of English words associated with various psychometric properties. However, LIWC is not available for many of the world's languages. Thus, we develop an automatic method to translate the English LIWC into other languages. Our automatic translations exhibit high overlap with an existing manual Chinese translation (Huang et al., 2012), indicating that machine translation is a viable alternative to human translations, which are often tedious and costly. Using our translated LIWC dictionaries,

Proceedings of the The 5th Workshop on Narrative Understanding, pages 36–47
July 14, 2023

we perform in-depth analyses of many categories of arcs, including those that represent structure and emotion, in eight different languages. Next, we investigate narrative arcs across stories in multiple languages from Project Gutenberg, a large repository of public-domain books. While different languages largely exhibit similar arcs on average, we find that different genres of stories follow diverse narrative arcs, which we concretely quantify through correlation analyses. Finally, we demonstrate how to interpret clusters of arcs, and how similar word categories can be identified by their arcs even when the categories have no words in common. Code to reproduce our experiments is available at `github.com/wswu/arcs`.

2 Related Work

Storytelling. Storytelling differences have largely been investigated in classroom settings (see McCabe (1997) for a survey). For example, the age and ethnicity of the storyteller are linked to differences in the stories' emotionality, relationality, and socialization (Pasupathi et al., 2002). However, such differences have not been investigated in novels and at the scale conducted in our work.

Narrative Arcs. The field of NLP disagrees on what exactly constitutes narrative (Piper et al., 2021). Narrative arcs are one method for studying the structure of narratives. They do not seek to capture traditional notions of narrative (e.g. sequences of events or interactions between characters) but rather measure changes in a story over time. Most previous work has focused on emotion or sentiment arcs. Mohammad (2011) study the occurrence of emotion words by applying the NRC Emotion Lexicon Mohammad and Turney (2013) to English novels and fairy tales. Reagan et al. (2016) study emotional arcs in English fiction books from Project Gutenberg using a variety of machine learning methods including principal component analysis, clustering, and self-organizing maps. Somasundaran et al. (2020) study emotional arcs in stories written by students. Boyd et al. (2020) compile a set of words associated with three narrative phases—staging, plot progression, and cognitive tension—and apply these lists to analyze a variety of texts including Project Gutenberg, self-published romance novels, sci-tech news articles, and Supreme Court opinions. Narrative arcs have also been applied to other downstream tasks, including predicting turning points in narratives (Ouyang and McKeown, 2015) and genre classification of novels (Kim et al., 2017). One common limitation in these works is their focus on English text, which we seek to remedy in our work.

LIWC Dictionaries. LIWC consists of a lexicon of word patterns associated with various psycholinguistic categories. Many previous efforts have translated earlier versions of LIWC into languages including (among others) Dutch (Boot et al., 2017; Van Wissen and Boot, 2017), German (Meier et al., 2019), and Romanian (Dudău and Sava, 2020). However, the process of translation often requires years of intensive manual effort. Computational approaches to LIWC translation are usually based on existing translation dictionaries, possibly with techniques such as triangulating through a third language (Massó et al., 2013). Van Wissen and Boot (2017) showed that using Google Translate to translate the LIWC dictionary word for word into Dutch is a viable solution. However, as of this writing, Google Translate supports only 113 languages. We develop a simple but effective automatic translation method using Wiktionary that can be applied to over 4,000 languages, and we show its effectiveness by comparing translations using this method with an existing Chinese LIWC dictionary (Huang et al., 2012).

3 Data and Dictionaries

Our analysis requires two main resources: a collection of narratives in multiple languages, and dictionaries with relevant word categories for the same set of languages.

3.1 Narratives

We utilize Project Gutenberg, a repository of over 60K public-domain books in many languages. We download the plaintext versions of books from Project Gutenberg, then remove Project Gutenberg headers and footers, lowercase, tokenize, and perform dependency parsing using spaCy.[1] Following existing work, we analyze novels within the Fiction genre, focusing on languages with the most number of books in Project Gutenberg (Figure 1) that also cover a wide range of cultures. Not shown in Figure 1 is the full English set of 13,656 fiction books. Given the uneven distribution of books across the languages, for our analyses described in Section 4.2, we downsample the set of English

[1] https://spacy.io

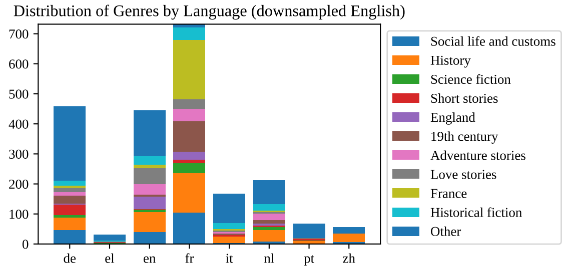

Figure 1: Number of fiction books compiled from Project Gutenberg, split by language and genres. Note that 13,565 English books were downsampled to form this set of 436 books shown here.

texts, keeping 436 books contained in the Project Gutenberg bookshelves "Best Books Ever Listings" or "Bestsellers, American, 1895-1923".

3.2 Word Dictionaries

We seek to quantify differences in narrative structure among stories of different languages. To this end, we study word category arcs using two sets of word lists: arc-of-narrative word lists (Boyd et al., 2020), and Linguistic Inquiry and Word Count (LIWC) dictionaries (Pennebaker et al., 2015). We describe each of these in turn.

Boyd et al. (2020) builds upon Gustav Freytag's pyramid of dramatic structure: exposition, rising action, climax, falling action, and resolution. They condensed Freytag's five-step model into three narrative phases: staging, plot progression, and cognitive tension. They find that the staging phase, associated with setting the scene of the story, is characterized by higher relative usage of function words such as prepositions and articles, which diminish as the story progresses. The plot progression phase is characterized by increased use of auxiliary verbs, pronouns, and connectives that help move the story forward. Finally, the cognitive tension phase is characterized by an increase in cognitive process words up until the climax of the narrative, at which point it then decreases. Boyd et al. (2020) constructed three lists of words correlated with these three patterns, which they call arcs of narrative.[2]

LIWC (Pennebaker et al., 2015) is a proprietary lexicon that associates word patterns with a range of psychological processes, including emotion, cognitive processes, perceptual processes, bodily processes, drives, personal concerns, and many others. LIWC is one of the most popular tools to analyze word usage in texts with respect to psychological processes.

We compute narrative arcs using word categories from both these dictionaries. By tracking the usage of a specific category of words (e.g. positive emotion words) longitudinally across the duration of the narrative, we can study the structure of narratives just as Vonnegut envisioned. Computationally, others have analyzed narratives in this way (Mohammad, 2011; Reagan et al., 2016; Boyd et al., 2020), but only on English text and with a limited number of word categories.

3.3 Translating Dictionaries

One goal of this work is to generalize the study of narrative arcs *across languages*. However, existing word lists are largely limited to English. In addition, some popular resources like LIWC are proprietary, and thus many researchers may not have access to LIWC and its translations. Thus, we develop a method to translate such dictionaries, including the arc of narratives list and LIWC, using Wiktionary,[3] a large, multilingual, crowdsourced dictionary freely available online.

Because these lists contain words as well as stem patterns (e.g. *happy* and *happi**), we first perform pattern expansion on each word, using the entries in Wiktionary as a comprehensive word list. Note that contrary to some traditional dictionaries, Wiktionary contains inflected forms as separate dictionary entries (e.g. *eat* and *eats*). Then, we use translations within Wiktionary (Wu and Yarowsky, 2020a,b) as a translation table to translate each word into seven target languages: German (de), Spanish (es), French (fr), Greek (el), Italian (it), Dutch (nl), and Chinese (zh). Each translation is then associated with the set of psychological categories of the original English word. This process is illustrated in Figure 2.

The process of pattern expansion on the three arc of narrative dictionaries expanded the original size of 916 words and patterns to 2,201 words. Pattern expansion on the English LIWC 2015 resulted in roughly 6.5K LIWC words and patterns expanded into 23K English words. The Wiktionary translation process generated a similar order of magnitude of translations into the target languages, as shown in Table 1. We use these translated dictionaries in the rest of this work.

Certain categories may be harder to translate: by

[2]Not to be confused with the broader term of *narrative arcs*.

[3]https://www.wiktionary.org

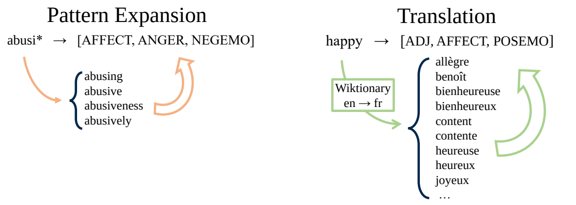

Figure 2: Illustration of the process of expanding LIWC asterisk patterns and performing automatic translation into French using Wiktionary. The resulting words inherit the original word's LIWC patterns.

Language	# Words
de	25k
el	11k
en	23k
es	23k
fr	20k
it	28k
nl	16k
zh	14k

Table 1: Translated LIWC dictionary sizes.

applying manually translated LIWC editions in English, Dutch, Romanian, and Brazilian Portuguese to analyze parallel texts in the four languages, Dudău and Sava (2021) found strong between-dictionary equivalences for function words that are not linguistically specific (e.g., negations, numbers, and I-statements), and several categories of content words (e.g., negative emotions, perceptual processes, biological processes, and personal concerns), while finding a weak correlation between many grammatical categories (e.g. third-person singular pronouns, auxiliary verbs, adverbs, conjugations, adjectives), the reward category, and the informal language category. Because of this, we ignore grammatical categories and limit our analysis of narrative arcs to psychological and cognitive categories, which are stable between languages (Dudău and Sava, 2021).

Case Study on Chinese. The Simplified Chinese version of LIWC (Huang et al., 2012) was created by manually translating the English LIWC 2007 and includes eleven new Chinese-specific categories that do not exist in the English version, as well as 106 words that occurred in the top 2000 most frequent words in the Sinica Corpus 3.0 (Chen et al., 1996). To validate our translation approach, we apply our method to automatically translate the English LIWC into Chinese and compare with the existing human-translated Chinese LIWC. The Chinese LIWC contains 6,828 words, while our translation of the English LIWC into Chinese contains 14,849 words. Because our translation contains both simplified and traditional characters, we convert all traditional Chinese characters to simplified characters using character conversion tables,[4] resulting in a total of 9,937 translations. In addition, because the English and Chinese word lists have slightly different LIWC categories, we remove the following categories that do not exist in both lists: all function words (FUNCT and subcategories); from the Chinese version, tense words (TENSEM and subcategories) and HUMANS; and from the English version: MALE, FEMALE, and certain informal words (INFORMAL, NETSPEAK, NONFLU, and FILLER).

Words in the Chinese LIWC have a mean number of categories of 2.46 (std. 1.04), while our translated list has a mean number of categories of 3.01 (std. 1.75), indicating that our automatic translation is slightly overproductive. The two lists' intersection contains 3,301 words, with a Jaccard distance of 0.54 indicating moderately high overlap. We compare a random selection of words in Table 2.

Though our translations are overproductive, the new word categories are often valid additions. For example, 哇 'wow!' is annotated as AFFECT and ASSENT in (Huang et al., 2012), but our translation adds the categories INFORMAL, NETSPEAK, and POSEMO. We believe these categories are actually omissions from the manual translation. Often the differences in categories lie at the superclass level because LIWC categories are hierarchical: a word labeled as WORK also falls under PERSONAL (the superclass of WORK). Similarly, all POWER words are DRIVES words by definition. For words that exist in our translation but do not exist in (Huang et al., 2012), a manual analysis indicates that many of them should be valid inclusions.

This case study on Chinese indicates that word-level translation of the English LIWC dictionaries is practical and feasible. Thus, we release our translations of the English LIWC into the seven non-English languages investigated in this paper, as well as the code to generate translations into over 4,000 languages supported by Wiktionary, in order to encourage further research in this area. We believe these automatically translated lexicons will serve as excellent starting points, saving hundreds of hours of manual translation. These can then be

[4] https://github.com/BYVoid/OpenCC

Word	Translation	LIWC Categories in Huang et al. (2012)	LIWC Categories in Our Translation
客人	guest	FRIEND, SOCIAL	FRIEND, SOCIAL
舐	to lick	PERCEPT	PERCEPT
消沉	depressed	AFFECT, NEGEMO, SAD	AFFECT, NEGEMO, SAD
哇	wow	AFFECT, ASSENT	AFFECT, INFORMAL, NETSPEAK, POSEMO
公安	public safety	PERSONAL, WORK	DRIVES, POWER, WORK
律师	lawyer	PERSONAL, WORK	DRIVES, POWER, WORK
节食	to diet	BIO, INGEST	BIO, HEALTH, INGEST
孙女	granddaughter	FAMILY, SOCIAL	FAMILY, FEMALE, SOCIAL
套房	hotel suite	HOME, PERSONAL	HOME
作孽	to sin	—	AFFECT, NEGEMO, RELIG
好站	warlike	—	ADJ, AFFECT, ANGER, NEGEMO
落败	to be defeated	—	ACHIEV, AFFECT, DRIVES, NEGEMO, POWER
滴	to drip	—	MOTION, RELATIV
乐观主义	optimism	—	AFFECT, DRIVES, POSEMO, REWARD
难吃	unpalatable	AFFECT, NEGEMO, PERCEPT	—
确立	to establish	CERTAIN, COGMECH	—
北面	northern side	RELATIV, SPACE	—
怒视	to glower	AFFECT, ANGER, NEGEMO, PERCEPT, SEE	—
远视	far-sighted	BIO, HEALTH	—

Table 2: Comparison of a random selection of words in the Chinese LIWC (Huang et al., 2012) and our automatic translation of the English LIWC into Chinese. Our translation tends to be overproductive but produces words that are associated with valid categories. Note that some categories are hierarchical. For example, PERSONAL encompasses WORK and HOME, while DRIVES encompasses POWER.

verified by human annotators to form larger, broad-coverage lexicons.

4 Quantifying Narrative Arcs

With our translated word dictionaries, we now investigate narrative arcs across languages.

4.1 Methods for Narrative Arcs

A narrative arc, also known as a word category arc, timeline, or trajectory, is a collection of word counts measured across segments of a narrative. Mathematically, a narrative arc is a word usage time series and can be conveniently visualized as a line plot, where the x-axis spans equally-spaced segments of the narrative, and the y-axis indicates the word usage computed within each segment. In previous work, the number of segments within a narrative varies from 5 (Boyd et al., 2020) to 20 (Mohammad, 2011), to a fixed window size of 10k words Reagan et al. (2016). For our experiments, we use 10 segments, a happy medium that balances granularity and computational cost. In addition, we follow Boyd et al. (2020) in z-score standardizing the word usage across each story in order to better analyze the difference in relative (rather than absolute) usage of words as a function of time.

4.2 Clustering and Interpreting Arcs

After computing arcs on all narratives in our dataset, we perform clustering of arcs within a word category to characterize stories that follow a particular arc. Reagan et al. (2016) discovered six arcs that correspond with Vonnegut's predictions (Vonnegut, 1999): 'Rags to riches' (rise), 'Tragedy' or 'Riches to rags' (fall), 'Man in a hole' (fall-rise), 'Icarus' (rise-fall), 'Cinderella' (rise-fall-rise), 'Oedipus' (fall-rise-fall). We ask: do these arcs also exist in non-English stories? To answer this question, we partition similar stories by their arcs using unsupervised clustering methods and then identify features of each group, a process reminiscent of topic models (Blei and Lafferty, 2009; Blei, 2012).

We perform k-means clustering on arcs of a specific LIWC category calculated on Fiction stories in Project Gutenberg across multiple languages, but using the downsampled English set (see Section 3.1), otherwise clustering will overemphasize English's contribution. We select the optimal number of clusters based on the elbow method with cluster inertia (the sum of squared distance between each point and the cluster centroid), a common metric for identifying the goodness of clusters. For many LIWC categories, we find that five to seven clusters are optimal.

Case Study on Positive Emotion Arcs. As a case study, we consider clusters of positive emotion (POSEMO) word usage trajectories. The elbow method indicates an optimal number of 5 clusters. The centroids of each cluster are shown in (Figure 3). To understand and interpret these clusters, a visual examination of each arc's peak pinpoints the

#	Shape	Size	Genres	Languages	Examples
0	rise-fall	236	History (18.2%), France (12.3%), Social life and customs 29 12.3%	en (29.7%), fr (28.8%), de (9.3%), nl (8.1%)	水滸傳 (Shi Nai'an), *L'île mystérieuse* (Jules Verne), *The Private Memoirs and Confessions of a Justified Sinner* (James Hogg), *Scaramouche: A Romance of the French Revolution* (Rafael Sabatini)
1	fall	222	History (17.6%), France (11.7%), Social life and customs (11.3%)	en (34.2%), fr (32.0%), es (8.6%), de (8.6%)	*The Awakening of Helena Richie* (Margaret Wade Campbell Deland), *Coniston — Volume 04* (Winston Churchill), *Trois contes* (Gustave Flaubert), *Elpénor* (Jean Giraudoux)
2	fall-rise-fall	218	History (21.1%), France (13.3%), Social life and customs (11.0%)	fr (34.9%), en (26.6%), de (12.4%), nl (8.3%)	*Die Klerisei* (N. S. Leskov), *The Reign of Law; a tale of the Kentucky hemp fields* (James Lane Allen), *The Right to Read* (Richard Stallman), *The Monk: A Romance* (M. G. Lewis)
3	fall-rise-fall-rise	211	History (18.5%), Historical fiction (10.4%), Love stories (10.4%)	en (50.7%), fr (18.5%), de (10.4%), it (7.1	狂人日記 (Lu Xun), *Jane Cable* (George Barr McCutcheon), *Robinson Crusoe (II/II)* (Daniel Defoe), *Le nabab, tome II* (Alphonse Daudet)
4	rise-fall-rise	224	History (18.3%), France (12.1%), Social life and customs (11.6%)	en (32.1%), de (32.1%), fr (22.3%), it (6.7	*The Expedition of Humphry Clinker* (Tobias Smollett), *La Marquise* (George Sand), *Les petites alliées* (Claude Farrère), *Du côté de chez Swann* (Marcel Proust)

Table 3: Interpretation of clustering on Positive Emotion arcs of stories across languages.

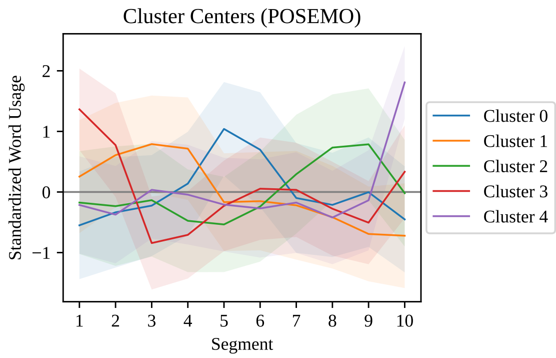

Figure 3: Cluster centroids of positive emotion (POSEMO) arcs computed on Fiction stories in Project Gutenberg (rebalanced English). Error bars indicate standard deviation.

location in the story where the most frequent use of positive emotion words occurs. We now dive deeper within each cluster, characterizing specific aspects including the languages and genres of the stories within in Table 3.

We find that the five clusters closely correspond to the following Vonnegut shapes: cluster 0 (blue) corresponds to 'Icarus' (rise-fall), cluster 1 (orange) corresponds to 'riches to rags' (fall), cluster 2 (green) corresponds to 'Oedipus' (fall-rise-fall), cluster 3 (red) corresponds to 'double man in a hole' (fall-rise-fall-rise), and cluster 4 (purple) corresponds to 'Cinderella' (rise-fall-rise). We do not see a 'man-in-the-hole' (fall-rise) -shaped arc, although at a high level, cluster 3 can be interpreted as fall-rise. If we specify six clusters, we find a sixth arc with a rise-fall-rise-fall shape that again may be a more specific form of the more general rise-fall shape.

In terms of cluster size, k-means tends to generate similarly sized clusters. We performed additional experiments clustering with HDBSCAN (McInnes et al., 2017), a hierarchical density-based clustering algorithm. HDBSCAN automatically identified 11 optimal clusters when computing POSEMO clusters. However, the majority of narratives were considered noise by this algorithm, and were thus not assigned a cluster, so we do not further analyze the HDBSCAN results here.

When analyzing genres, we find that History, France, and Social life are the top three genres in the entire dataset. Within a cluster, the only cluster that stands out is cluster 3, which is characterized by a larger portion of Historical fiction and Love stories, indicating that these genres tend to prefer this story structure. This cluster is also made up of over 50% English novels.

For non-English stories, the highest percentage of French novels appeared in cluster 2, while the highest percentage of German novels appeared in cluster 4. This may indicate a preference for these arc shapes by speakers of these languages. Such a preference could be cultural: from France and Germany originated Charles Perrault and the Grimm Brothers, respectively, whose fairy tale compilations have been read by children of numerous generations. Thus, the clustering of arcs allows us to examine similarities and differences between groups of narratives. While we consider positive emotion arcs here, due to their similarity with Vonnegut's story structure, future work will investigate other categories and their relevance to narrative structure.

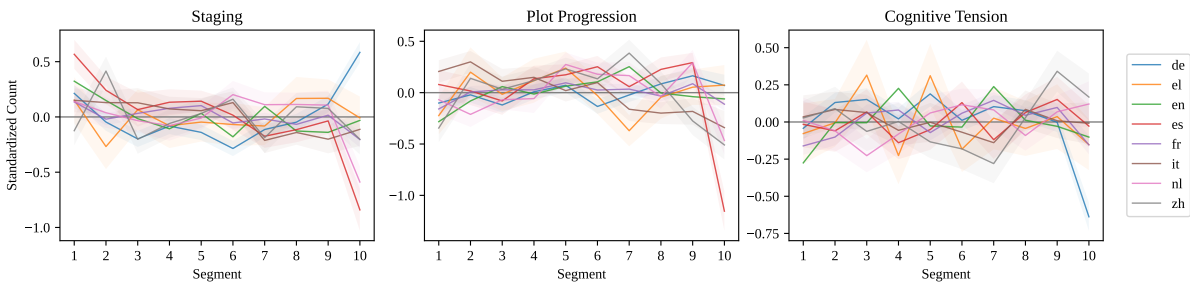

Figure 4: Boyd et al. (2020)'s word categories translated and computed on fiction narratives in other languages.

	Staging		Plot Prog		Cog Ten	
Lang	r	p	r	p	r	p
de	-0.075	0.836	0.241	0.502	0.604	0.064
es	0.723	0.018*	0.727	0.017*	-0.543	0.105*
fr	0.779	0.008**	0.801	0.005**	0.435	0.209
it	0.634	0.049*	0.789	0.007**	0.009	0.981
nl	0.721	0.019*	0.756	0.011*	0.086	0.813
zh	0.693	0.026*	0.858	0.002**	-0.198	0.583

Table 4: Correlation between narrative arcs to the English arcs in fiction stories. r is the Pearson correlation coefficient, and p is the p-value. A single asterisk indicates p-values ≤ 0.05, while double asterisks indicate p-values ≤ 0.01.

Category 1	Category 2	Corr.	Overlap
DISCREP	PLOTPROG	0.987	0.05
SOCIAL	YOU	0.986	0.01
FEMALE	I	0.973	0
AFFECT	REWARD	-0.972	0.04
AFFILIATION	WE	0.970	0.01
FEEL	WE	0.967	0
FILLER	NONFLU	0.962	0
FILLER	RELIG	-0.957	0
DEATH	NONFLU	-0.955	0

Table 5: Most strongly correlated narrative arcs (including negatively correlated). All correlations are significant ($p < 0.001$). PLOTPROG is from Boyd et al. (2020) and is not a LIWC category. Corr is the Pearson correlation coefficient, and Overlap is the Jaccard similarity between the words in each category.

5 Narrative Arcs Across Languages and Genres

5.1 Story Structure Processes

We now investigate narrative arcs' implications on narrative structure across languages by comparing them with an established study of narrative structure in English. Boyd et al. (2020) constructed three word categories corresponding to primary story structure processes: staging, plot progression, and cognitive tension. They then computed narrative arcs using these word categories, experimenting on various domains of text. We examine whether these three categories also apply to stories in languages other than English. We translate Boyd et al. (2020)'s word lists and apply them to a set of Fiction stories, standardizing the word counts within each story in order to allow fair comparison of relative word usage across stories. We compute the mean narrative arcs for fiction stories (shown in Figure 4), where error bars indicate standard error, and we calculate the Pearson correlation between each non-English narrative arc in Table 4.

Overall, we find strong support for Boyd et al. (2020)'s notion of staging and plot progression across languages, with most languages except for German showing a strong, statistically-significant correlation with the English narrative arc. For cognitive tension, we find that German, Spanish, and French arcs are weakly correlated, with Spanish surprisingly negatively correlated.

5.2 Arcs by Category

In addition to identifying similar stories, word usage arcs can also inform us about similarities between *word categories*, especially those with seemingly little or no overlap. Such analysis is similar to the idea of burstiness (Schafer and Yarowsky, 2002), where similar words occur at similar frequencies across time, an idea that was one of the precursors to the modern notion of embeddings computed based on some aspect of word usage.

We compute narrative arcs on all books within the Fiction genre in Project Gutenberg for each LIWC category and compute the Pearson correlation between the means of the arcs within each category. We show the most correlated categories in Table 5; the correlations between all categories are shown in Figure 7 in the Appendix.

Most of these correlations have a natural explanation. PLOT PROGRESSION words (from Boyd et al.

(2020)) are strongly correlated with LIWC DISCREPANCY words (*should*, *would*, *could*), which help to drive the plot forward. SOCIAL words (including social actions as well as relationships) already encompass a large percentage of You words (*you*, *y'all*), so high correlation is expected. However, some pairs of categories have zero overlap. FEMALE words (*girl*, *her*, *mom*) and I words (*I*, *me*) have high correlation; these words tend to occur in similar contexts (a paradigmatic relationship), as do FILLER words (*anyway*, *y'know*) and NONFLUENCIES (*er*, *um*). FEELING words (related to the perceptual process of touch, such as *feel*, *touch*, *cool*, *warm*) and WE words (*we*, *us*, *our*) in contrast have a syntagmatic relationship: they occur together but cannot be substituted for one another. The negatively correlated category pairs are also interesting. AFFECT words (related to emotion) and REWARD words (*take*, *prize*, *benefit*) have slight overlap and a strong negative correlation, the explanation of which needs further investigation. FILLER words and RELIGION words, as well as DEATH words and NONFLUENCIES, can be considered complete opposites: death and religion are heavy topics not often discussed with inconsequential or informal language such as filler words, and thus show a negative correlation.

5.3 Arcs By Genre

While certain plot structures may be universal, different genres may prefer different narrative structures. In this section, we discover structural differences between genres through the lens of word category arcs.

Consider Figure 5, containing all arcs computed on the LIWC category PERCEPT, which includes perception processes (e.g. seeing, hearing, and feeling). Through a visual inspection, we find that a large number of narratives in the History genre (total 1.2k books) exhibit a downward usage in perception words between segments 9 and 10, while in Science fiction (total 1.6k books), a visible portion of books have already dropped their usage of Perception words starting around segment 7.

To concretely quantify the difference between genres, we compute narrative arcs for all word categories over the eight most frequent genres within Fiction in Project Gutenberg: Social life and customs, History, Science fiction, Short stories, England, 19th century, Adventure stories, and Love stories. We identify word categories that maximally separate these genres by minimizing the mean absolute correlation between each pair of genres $\mathrm{MAC}_d = \frac{1}{n}\sum_{g_1,g_2}|\mathrm{r}(\overline{\mathrm{arc}}_d(g_1), \overline{\mathrm{arc}}_d(g_2))|$ for word category d, n pairs of genres g_1 and g_2 in the set of top 8 genres, $\overline{\mathrm{arc}}_d(g)$ indicating the mean narrative arc computed on word usage of dimension d on stories in genre g, and r is the Pearson correlation coefficient.

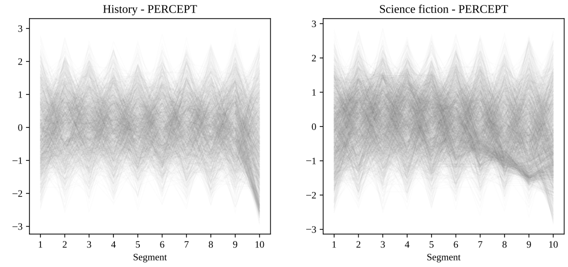

Figure 5: All Perception narrative arcs plotted for the genres History and Science fiction. Notice the clear difference in where usage of Perception words drops off.

The LIWC categories that maximally separate the top eight genres are SEXUAL (MAC = 0.29), ADVERBS (MAC = 0.37), and FILLER (MAC = 0.42), shown in Figure 6. We see, for example, that science fiction and short stories on average have a higher usage of SEXUAL words (*love*, *lust*) at the beginning of the narrative, which subsequently declines. The inclusion of love scenes at the beginning of a novel is a technique frequently used by authors to hook the reader. On the other hand, love stories on average are more likely to use Sexual words both at the beginning and the end of the story, perhaps indicating a happy ending. The next most distinguishable categories, Adverbs and Filler words, are harder to interpret due to their non-content nature. The categories that have the least distinguishing power are WE words (MAC = 0.91), CAUSE words (MAC = 0.92), and AFFILIATION words (MAC = 0.94); these arcs are very similar regardless of the genre.

5.4 Arcs by Language

Finally, we investigate how arcs differ with respect to language. We perform this analysis by correlating arcs computed for different word categories on stories in different languages, grouping stories by language. Correlation between languages for the same category is presented in Table 6.

When evaluating arcs across languages, we find that the most highly correlated categories are mem-

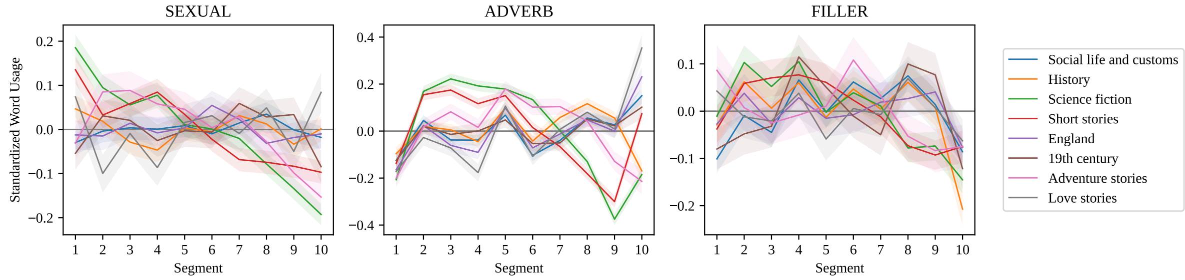

Figure 6: Top three word categories that maximally separate genres. Error bars indicate standard error.

Lang1	Lang2	Category	r
en	fr	DEATH	0.956
es	zh	NONFLU	0.956
es	nl	INFORMAL	0.956
es	fr	INFORMAL	0.94
fr	nl	INFORMAL	0.931
fr	nl	FOCUSPAST	0.931
es	fr	NUMBER	0.921
es	fr	NETSPEAK	0.909
es	nl	NETSPEAK	0.898
es	fr	ASSENT	0.896
de	es	ASSENT	0.884
es	fr	STAGE	0.881
en	fr	MOTION	0.88
en	fr	INFORMAL	0.878
de	fr	NEGATE	0.876
en	zh	NONFLU	0.876

Table 6: Most correlated categories across languages. All categories are from LIWC except STAGE, which is from Boyd et al. (2020). All correlations are statistically significant ($p < 0.001$).

bers of the INFORMAL category (including ASSENT, NONFLUENCIES, and NETSPEAK). For the other prominent categories, we already showed in Section 5.2 that DEATH words are highly correlated. For FOCUSPAST, a category that includes words that indicate focusing on past action (e.g. *was*, *has*, *been*), the high correlation between French and Dutch may be due to the fact that French and Dutch have some similarities in their past tenses.[5] For NEGATE, in both French and German, the negation word often comes after the verb (e.g. French *nous ne mangeons pas* vs. German *wir essen nicht*. Thus, narrative arcs also enable the study of language typology through careful selection of word categories.

[5]The French *passé simple* and *imparfait*, along with the Dutch *onvoltooid verleden tijd* (OVT), are morphologically simplex, while the French *passé composé* and Dutch *voltooid tegenwoordige tijd* (VTT) are morphologically complex.

6 Conclusion

Narrative arcs, operationalized as word category arcs, model word usage across the timeline of a narrative. They are powerful tools that allow us to not only gain a high-level overview of a narrative's structure but also enable us to identify similarities across languages and genres. In order to quantify narrative arcs across languages, we present a method for automatically translating wordlists such as LIWC, which we validate with an existing Chinese translation of LIWC. We then apply our translated dictionaries in eight languages to analyze narrative arcs in Project Gutenberg fiction books.

We first investigate clustering to interpret narrative structure according to Kurt Vonnegut's claims. Next, we investigate story structure, showing that Boyd et al. (2020)'s created word categories findings largely hold across languages. We then perform correlation studies, interpreting narrative arcs with respect to word categories, genres, and languages. Analyzing categories, we discover and explain positive correlations between several categories, even when they have no words in common. Analyzing genres, science fiction and short stories have a higher usage of SEXUAL words at the beginning of the story in order to hook the reader. Analyzing languages, we find that a high correlation between certain categories like DEATH and INFORMAL words can indicate a typological relation.

This work investigates how narrative arcs differ across various dimensions; we leave the question of *why* to future work.

Limitations

Corpus. In this paper, we use fiction novels from multiple languages in Project Gutenberg. One assumption of this work is that the text is representative of the culture surrounding the language. While this may or may not be true (e.g. Handler and Segal, 1999), our investigation's focus is on the structure, or narrative arc, of stories and how arcs may differ across languages. Naturally, our findings may differ for other genres, such as history or self-help. We focus on fiction because the vast majority of research on narratives has focused on fiction, though we believe non-fiction and other genres would be interesting for future work. Future work can also consider the addition of other corpora to enhance Project Gutenberg, such as MegaLite Moreno-Jiménez et al. (2021), a corpus of about 5,000 Spanish, French, and Portuguese narrative texts, poetry, or plays. However, multilingual corpora of this kind are few and far between, even for high-resource languages like Spanish and French.

Dictionaries. This work heavily relies on LIWC, which is proprietary software. Many researchers (including ourselves) may not have access to all LIWC dictionaries. In addition, as a dictionary of psychometric properties, LIWC is constantly evolving and improving with new research in psychology and linguistics.

Acknowledgements

This work is supported in part by the Air Force Office of Scientific Research (#FA9550-22-1-0099) and by the Templeton Foundation (#62256). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of AFOSR or of the Templeton Foundation. We would also like to thank the members of the LIT and LAUNCH labs at the University of Michigan, and especially Aylin Gunal, for helpful feedback on an early draft of this work.

References

David M Blei. 2012. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

David M Blei and John D Lafferty. 2009. Topic models. In *Text mining*, pages 101–124. Chapman and Hall/CRC.

Christopher Booker. 2004. *The seven basic plots: Why we tell stories*. A&C Black.

Peter Boot, Hanna Zijlstra, and Rinie Geenen. 2017. The dutch translation of the linguistic inquiry and word count (liwc) 2007 dictionary. *Dutch Journal of Applied Linguistics*, 6(1):65–76.

Ryan L Boyd, Kate G Blackburn, and James W Pennebaker. 2020. The narrative arc: Revealing core narrative structures through text analysis. *Science advances*, 6(32):eaba2196.

Keh-Jiann Chen, Chu-Ren Huang, Li-Ping Chang, and Hui-Li Hsu. 1996. Sinica Corpus : Design methodology for balanced corpora. In *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation*, pages 167–176, Seoul, Korea. Kyung Hee University.

Diana Paula Dudău and Florin Alin Sava. 2020. The development and validation of the romanian version of linguistic inquiry and word count 2015 (ro-liwc2015). *Current Psychology*, pages 1–18.

Diana Paula Dudău and Florin Alin Sava. 2021. Performing multilingual analysis with linguistic inquiry and word count 2015 (liwc2015). an equivalence study of four languages. *Frontiers in Psychology*, 12:2860.

William Foster-Harris. 1959. *The basic patterns of plot*. University of Oklahoma Press.

Gustav Freytag. 1894. *Die technik des dramas*. S. Hirzel.

Richard Handler and Daniel Segal. 1999. *Jane Austen and the fiction of culture: An essay on the narration of social realities*. Rowman & Littlefield.

Chin-Lan Huang, Cindy K Chung, Natalie Hui, Yi-Cheng Lin, Yi-Tai Seih, Ben CP Lam, Wei-Chuan Chen, Michael H Bond, and James W Pennebaker. 2012. The development of the chinese linguistic inquiry and word count dictionary. *Chinese Journal of Psychology*.

Evgeny Kim, Sebastian Padó, and Roman Klinger. 2017. Investigating the relationship between literary genres and emotional plot development. In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 17–26, Vancouver, Canada. Association for Computational Linguistics.

Guillem Massó, Patrik Lambert, Carlos Rodríguez Penagos, and Roser Saurí. 2013. Generating new liwc dictionaries by triangulation. In *Asia Information Retrieval Symposium*, pages 263–271. Springer.

Allyssa McCabe. 1997. Cultural background and storytelling: A review and implications for schooling. *The Elementary School Journal*, 97(5):453–473.

Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11):205.

Tabea Meier, Ryan L Boyd, James W Pennebaker, Matthias R Mehl, Mike Martin, Markus Wolf, and Andrea B Horn. 2019. "liwc auf deutsch": The development, psychometrics, and introduction of de-liwc2015. *PsyArXiv*, (a).

Saif Mohammad. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114, Portland, OR, USA. Association for Computational Linguistics.

Saif M Mohammad and Peter D Turney. 2013. NRC emotion lexicon. *National Research Council, Canada*, 2.

Luis-Gil Moreno-Jiménez, Juan-Manuel Torres-Moreno, et al. 2021. Megalite: a new spanish literature corpus for nlp tasks. In *Computing Conference*.

Jessica Ouyang and Kathleen McKeown. 2015. Modeling reportable events as turning points in narrative. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2149–2158, Lisbon, Portugal. Association for Computational Linguistics.

Monisha Pasupathi, Risha M Henry, and Laura L Carstensen. 2002. Age and ethnicity differences in storytelling to young children: Emotionality, relationality and socialization. *Psychology and Aging*, 17(4):610.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.

Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative theory for computational narrative understanding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 298–311, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Georges Polti. 1917. *The thirty-six dramatic situations*. Editor Company.

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1):1–12.

Charles Schafer and David Yarowsky. 2002. Inducing translation lexicons via diverse similarity measures and bridge languages. In *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*.

Swapna Somasundaran, Xianyang Chen, and Michael Flor. 2020. Emotion arcs of student narratives. In *Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events*, pages 97–107, Online. Association for Computational Linguistics.

Ronald B Tobias. 2012. *20 master plots: And how to build them*. Penguin.

Leon Van Wissen and Peter Boot. 2017. An electronic translation of the liwc dictionary into dutch. In *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference*, pages 703–715. Lexical Computing.

Kurt Vonnegut. 1999. *Palm Sunday: an autobiographical collage*. Dial Press.

Winston Wu and David Yarowsky. 2020a. Computational etymology and word emergence. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3252–3259, Marseille, France. European Language Resources Association.

Winston Wu and David Yarowsky. 2020b. Wiktionary normalization of translations and morphological information. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4683–4692, Barcelona, Spain (Online). International Committee on Computational Linguistics.

A Appendix

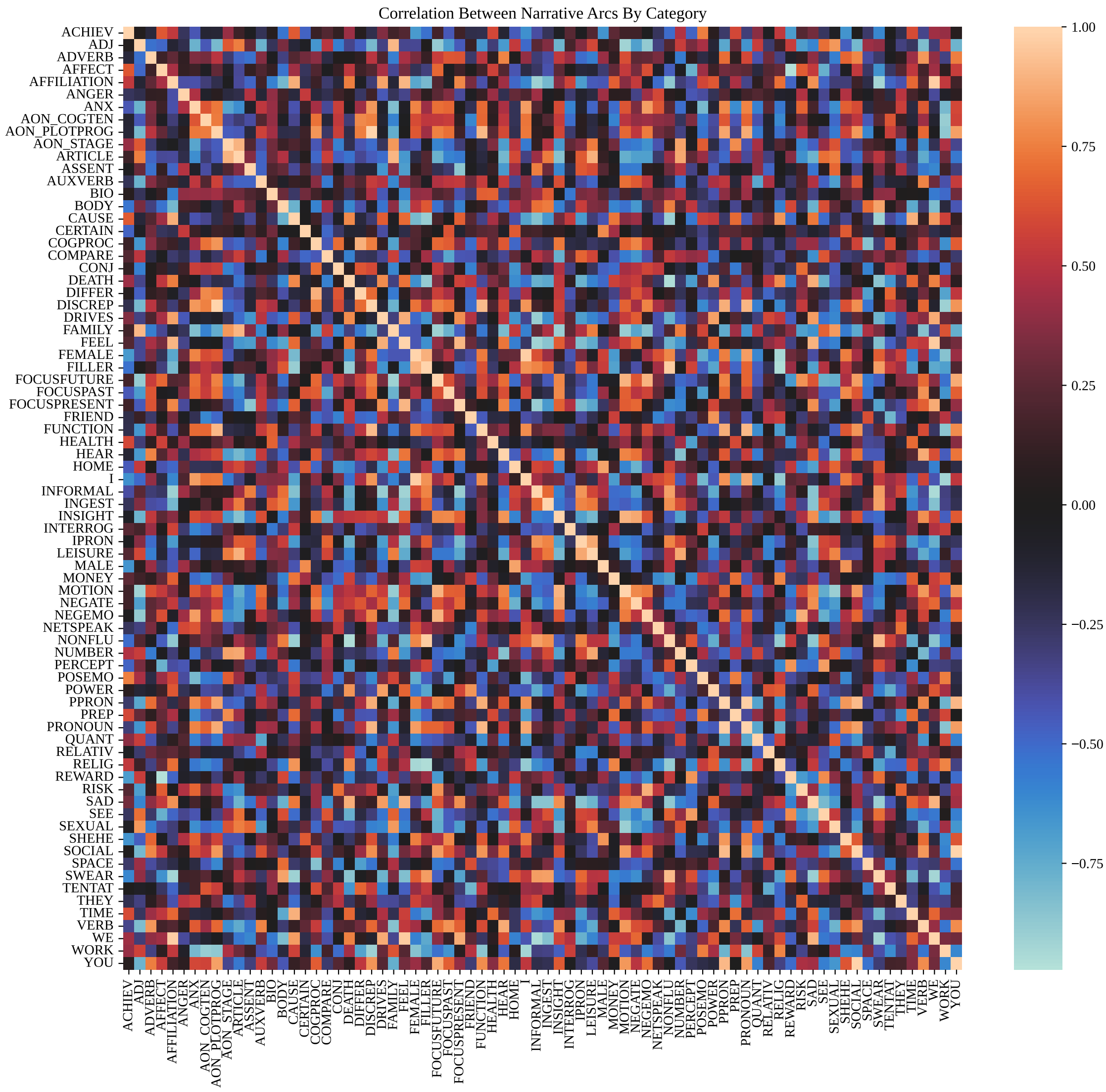

Figure 7: Correlation between narrative arcs for each LIWC category, with the addition of the three categories starting with AON_ from Boyd et al. (2020). The most highly correlated categories (including negative correlation) are in light blue and light red and are analyzed in Section 5.2.