

Harshita Gupta

Humanities Colloquium

April 26, 2017

Cover Letter

In developing this revision, I used Franco Moretti's "Network Theory, Plot Analysis" from *Distant Reading* as a model of a work in a similar genre. The day after our conference, I was finally able to get my network code to work, and figured out how to visualize topics as a graph connecting different words and elements across texts. This gave me a very productive experimental setup for understanding what I'd been interested in all along, the unique thematic 'structure' of each of the texts. As Moretti says in his analysis of Hamlet through a network of its plot,

Finally - and it is the most important thing of all, but also the most difficult - one can intervene on a model; make experiments. Take the protagonist again. For literary critics, this figure is important because it is a very meaningful part of the text; there is always a lot to be said about it; we would never think of discussing Hamlet without Hamlet. But this is exactly what network theory tempts us to do: take the Hamlet-network, and remove Hamlet, to see what happens.

I engage in this process of "making experiments" on the networks that I produce, by deleting different variations of words like 'women', 'daughter', 'mother', and 'mistress', and subsequently understanding their effect on the network: to put it differently: I attempt to understand each translation with questions like "What remains of the of each translation's depiction of women when we remove mentions of all male - associated women, like mothers, wives, and daughters?" "Do women remain in the networks, and to what extent, when we remove mentions of courtly duties and performance, versus when we remove mentions of feeling, emotion, and thought?" How does each translation's network respond differently to these interventions?

Title

In Metonymy in The Tale of Genji: An Analysis of Translation Strategies, Janel R. Goodman Murakami compares occurrences of metonymy in a passage across translations of the Tale of Genji to assess how domesticated or foreign Suematsus, Waleys, Seidenstickers, and Tylers translations are, concluding that the more modern translators retain foreign elements of the text more faithfully than their earlier counterparts. In “Going to Bed with Waley: How Murasaki Shikibu does and does not become world literature,” Valerie Henitiuk uses a similar microanalytic approach, more commonly referred to as close-reading, to critique Waleys translation of Genji and his portrayal of women. I build on the preexisting body of scholarship on Japanese to English translations of Genji, instead using the macroanalytic approach of topic modeling, to compute the themes across the translations, and analyze what discrepancy between translations reflects about the texts portrayal of women. TODO EXPAND ON WHAT EXACTLY I ultimately build upon Murakamis conclusion, determining that more recent translations not only portray Heian Japan more faithfully, but also elevate the position of women as individuals without “censor[ing],” as Henitiuk put it, male-female interactions to the same degree as older ones. TODO WHAT TYPE OF CENSORING, STAKES BECOME MORE CLEAR W DETAILS

Figure ?? shows a boat.

I apply digital topic modeling to break down and analyze three translations of Genji: Arthur Waleys from 1925, Edward G. Seidensticker’s from 1976, Royall Tylers from 2001, and Dennis Washburns from 2015¹². Topic modeling is a natural language processing technique that

¹My greatest interest was originally in comparing the work of male translators with Helen McCulloughs partial translation. Unfortunately, McCulloughs translation, published only in *Genji and Heike: selections from the Tale of Genji and the Tale of the Heike*, is heavily copyrighted and unavailable in any digital format.

²To prepare the text for topic modeling, I removed all annotations. Introductions, publishers notes, footnotes, and endnotes were removed. In the Tyler translation, this included the deletion of all chapter titles that are not from the

uses the principle of distributional semantics, or the common cooccurrence of two words, to group words into “topics.” I use the Latent Dirichlet Allocation (LDA) topic modeling method, which operates on 1000-word sequential chunks of the tale of Genji. LDA treats each “chunk” of words as an entity composed of “topics”—each topic is composed of words that commonly occur together. A statistical explanation of LDA modelings assumptions and mechanisms is beyond the scope of this paper, and one can turn to Rhodys “Unpacking the Assumptions of LDA” or Jockers Macroanalysis for a simplified analysis of the method. Crucial to this paper and its discussion, however, is LDA modelings unsupervised nature. Topics are not produced based on the programs understanding of the words meanings or potential similarity, but creates buckets for topics based on the position of words relative to each other, and their common cooccurrence, i.e. distributional similarity, to determine that they belong to a similar topic. Its unsupervised nature makes LDA modeling useful for identifying topics in a more “objective” fashion by identifying authors subconscious placement of words and themes and therefore reflecting their cultural inclinations.

Topic modeling is useful, therefore, in uncovering topics that might be outside our microanalysis-based understanding of discoverable topics. TODO: matthew’s note says ”A powerful justification for the method and for the virtues of using a DH approach more generally, especially for a massive work such as Genji (difficult to keep in anyones head all at once)”, emphasize this

Topic modeling is traditionally applied to multiple documents that are dissimilar, as in the

original, chapter introductions, and the “Persons” and “Relationship to Previous Chapters” sections. Additionally, all words are reduced to their stems - therefore treating “respect” and “respectfully” as identical semantic units and indistinguishable to the LDA algorithm. I also exclude all proper nouns from the corpus used for the topic model, so that topics are identified not upon the basis of scene (which correlates highly with specific characters) but upon common occurrence of motifs. Special care was taken to exclude all Japanese proper nouns, which are not identified by western natural language processing tools. For a list of all Japanese proper nouns removed, see the code in Appendix A.

Figure 1: Topics in Waley's Translation

Topic 1: said now look even man old come make seem mother tree young hous
 Topic 2: said time now letter ladi long even much feel littl last go visit way far
 Topic 3: cat udoneri princess palac seen still fine creatur
 Topic 4: said go now thing come know think way see seem look say time thought get make much long feel
 tell peopl back last well day night inde even girl moment hous old felt quit take
 Topic 5: ladi side
 Topic 6: mistleto rebuilt
 Topic 7: year inde cold dress imagin
 Topic 8: now time day ladi said much emperor inde father great even way seem princ year thought palac
 daughter feel made make long matter girl old mother see littl
 Topic 9: now ladi time said day even seem thought great long littl look come poem inde
 Topic 10: time day now ladi inde much even said seem come felt long thing way feel
 Topic 11: now boy place emperor princ hous
 Topic 12: emperor palac great present
 Topic 13: emperor ladi princess art

Figure 2: Topics in Seidensticker's Translation

Topic 1: stag
 Topic 2: seem come said t even see littl bishop dream came child look old girl
 Topic 3: seem robe paint red princess ladi women line littl white string master nose now
 Topic 4: girl ladi said seem thought think come now t look littl even know daughter good young time see
 women mother make came much go well man way princ
 Topic 5: seem thought even think come see now said go ladi princess time thing littl women make know
 world long want much day way say made feel look came still noth father
 Topic 6: emperor princ thought seem ladi court chines crown daughter present palac time son said year royal
 Topic 7: princ ladi emperor said blossom kobai daughter thought carriag even
 Topic 8: seem thought women littl governor even young room ladi look light said came back open away
 door wind way see day veranda
 Topic 9: emperor time now matter tear princ minist empress
 Topic 10: seem ladi safflow quarter white look
 Topic 11: ladi seem thought now time day even said year come see old princess came made much littl
 Topic 12: seem ladi thought daughter year come old think said son day thing well great made man
 Topic 13: t man thing good governor say seem girl think make look know go happen said woman daughter
 wife one want young day come even just don ve sort someon see everyth peopl m let ask
 Topic 14: emperor ladi court cat princ son crown mother third new father go see royal

Figure 3: Topics in Tyler's Translation

Topic 1: high paper great shoot book write cup scroll command son
 Topic 2: now high even look said way well still thought never time see littl go come long know feel seem day just thing ladi
 Topic 3: play captain old tale now flute music never ladi feel said seem life day hear say even
 Topic 4: majesti now flower light even wind long morn said dew day
 Topic 5: blossom look command made flower ladi littl spring beauti year
 Topic 6: now monk day world holi long mountain come adept wind littl captain thought citi know rever sent look way time
 Topic 7: rever nun young woman lordship go die come mother said even someon tell now noth high see look seem ask never sure just back told heard well happen talk sister women mistress
 Topic 8: majesti flower spring autumn ladi nun music year pine east quarter made consort wing long blossom high garden color
 Topic 9: captain young excel high play look lieuten son music blossom said right just seem man dress one even well make daughter come ladi secretari advis
 Topic 10: even now old play novic high noth often see biwa
 Topic 11: ladi gown look dress comb box one never mistress even high far day now cathay plum
 Topic 12: now long majesti littl thought day mani ladi made year see said well citi time still even come go look

Figure 4: Topics in Washburn's Translation

Topic 1: go ladi just woman feel come girl know even way said make think now nurs tell ask littl man lord night well look time dont attend see thought place
 Topic 2: carriag go feel wife process men view attend ladi just way blind space
 Topic 3: feel husband now even ladi emot heart deep woman say
 Topic 4: made lotu attend third minist flower make
 Topic 5: even capit boat day governor sea attend provinc ladi wind perform lord dream wave
 Topic 6: robe look ladi poem even day garden now seem blossom color tree made women
 Topic 7: cat third princess princ son crown littl blind robe felt see look close game face never contest think tri remark
 Topic 8: even feel now time world look thought see come ladi live heart felt just princess long villa day go
 Topic 9: father hardli wife
 Topic 10: time daughter son even emperor palac princ princess now father court look day year majesti third minist mani feel ladi world left
 Topic 11: play koto string instrument perform princess even music young son boy littl seem flute hear time daughter look major just robe women feel
 Topic 12: woman make way even man young just time now thing matter go mani
 Topic 13: princess even look feel young daughter woman man just ladi now go think women make time way captain letter thought see thing say come still
 Topic 14: ladi wind princess day ceremoni daughter father come carriag

work by Jockers, and has been trained on corpuses containing multiple texts, 4500 poems in Rhodys case and 4500 texts in Jockers case. These models develop generalized, often easily identifiable themes that can apply to texts across time and author, like Rhodys “night light moon stars day dark” and “tree green summer flowers grass” topics. Topic modeling of the form they practice is useful when the purpose of the topics is to be used later in topic distribution comparison across texts, but is less useful when working with a single text. I train the LDA model, instead, on a single translation of Genji at a time, thereby developing topics that are far more specific to each translations “thematic world,” and useful in identifying the thematic differences underlying each translation. The interpretive advantage to this method is evident in the difference between the two topic word clouds reproduced side-by-side. One is from Jockers paper “Theme” in Microanalysis, and one is a theme discovered in Tylers translation of Genji. TODO add figure here. The topics discovered, with the words that have the highest percentage of appearance within a given topic, are included in figures ??, ??, ??, and ??³.

The themes discovered in the four translations, reproduced in the figures above, reveal clear differences between the three texts. The themes are reproduced with their theme number and theme descriptor (a few words from the theme that encompass how the theme is interpreted). TODO describe how themes are split up and process to identify female-related themes and non-female-related theems. In Washburns translation, 12 of the 20 themes discovered have female references —words like “woman”, “lady” (stemmed as ladi), “daughter” — as one of their top 10 words. In Tylers, 8 of the 20 themes do, and in Waleys, 6 of the 20 themes do. An analysis of the evolution of these themes over the text, visualized in the heat maps below, reveals that not only

³All words in a given topic with signal strength greater than 0.0040 are included, i.e. words that comprise more than four percent of a given topic. If a topic had no words with signal strengths high enough, i.e. the topic was statistically insignificant, those topics were excluded from the final model and not used in subsequent discussion.

are there dramatically fewer female-related themes in the Tyler than the Washburn and in the Waley than the Tyler, but that the frequency of these female-related themes is lower in the same order as well. TODO: if i stick with this, elaborate on the frequency bit.

In the Washburn text, the word “feel” is the second-most salient one in the entire novel. The topic-modeling results reveal that the word “feel” occurs exclusively in the female-related themes, and is mentioned not once in a female-unrelated theme. Additionally, while five of the female-related themes also mention male-related words in them, only one of these contains the word “feel.” [This demonstrates that the inner life of women is given priority in Washburns translation, and more importantly, in contexts where women are not surrounded by other men. Females are afforded a level of individuality that they are not in the other translations.] TODO comment was”This is where “microanalysis” (using human brain power) can come back in to make a contribution in supporting the contours of the big data [see end comments]”. This is evident in the Tyler translation, where the word majestic appears in four out of eight of the female themes (specifically the ones on courtship and ceremony), and only three times in all the other twelve, thereby indicating the presence of females as a cultural object for the purposes of courtly procedure and order. The Waley translation, by far, exhibits the most striking objectification and minimization of women —women occur explicitly in themes about ceremony, palaces, and governance, thereby indicating a position in the text that is secondary to their male fathers and husbands. [TODO COMMENT: Similarly insightful. To extend the point for maximum impact: what do these additional associations tell us about the nature of that “objectification and minimization”? i.e., the switch from an adjective or quality of class (“majestic”) to a set of physical places and cultural practices (“ceremony”). Ceremony appears in both Tyler and Waley, so how might it function differently in each?]

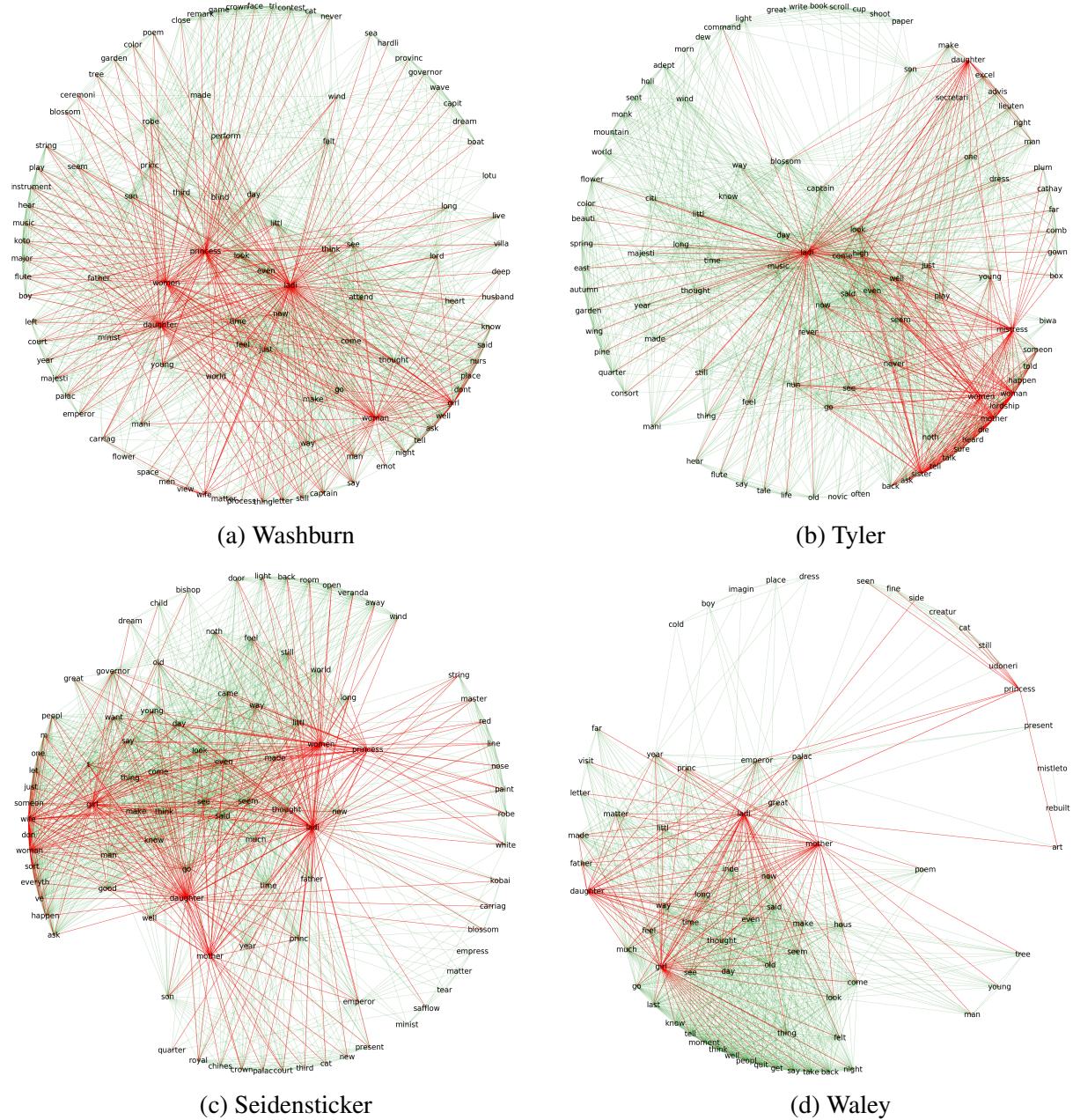


Figure 5: Thematic Structure: Female-Associated Connections in Red



Figure 6: Thematic Structures Without Women

An analysis of each thematic network without women reveals the centrality of their role in all four translations of the Genji: the structure of Washburn collapses entirely, leaving a virtually nonexistent graph, (See Figure ??), Seidensticker

The most intriguing changes, however, are when we remove individual female words rather than all female words in their entirety. Removal of the word 'mother' has different effects on each network, reducing Seidensticker's to a notably more romantic, sentimental, and courtship-heavy structure. In Waley's case, removal of the word "mother" results in

Removing feeling and thinking words, specifically 'feel', 'felt', 'think', 'thought', 'emot', removes all mention of females from Seidensticker's network.

Removing duty and palace-related words reduces the female network of Tyler almost exclusively to independent women-words, and removal of feeling-thinking words reduces the network almost entirely to male-independent women.

Works Cited

- Jockers, M. L.. *Macroanalysis: Digital Methods and Literary History*. Champaign: University of Illinois Press, 2013. Project MUSE.
- Rhody, Lisa M. "Topic Modeling and Figurative Language." *Journal of Digital Humanities*, vol. 2, no. 1, 2012, pp. 19—38.
- Henitiuk, Valerie. "Going to Bed with Waley: How Murasaki Shikibu does and does not become world literature." *Comparative Literature Studies*, vol. 45, no. 1, 2008, pp. 40—61., www.jstor.org/stable/25659632.
- Meeks, Elijah. "Using Word Clouds for Topic Modeling Results." *Digital Humanities Specialist*. N.p., 15 Aug. 2012. Web. 16 Apr. 2017. <https://dhs.stanford.edu/algorithmic-literacy/using-word-clouds-for-topic-modeling-results/>.
- Shikibu, Murasaki, and Dennis C. Washburn. *The Tale of Genji*. New York: W.W. Norton, 2015. Amazon.com. Web.
- Shikibu, Murasaki, and Arthur Waley. *The Tale of Genji*. Boston and New York: Houghton Mifflin, 1925. Amazon.com. Web.
- Murasaki Shikibu B. "University of Oxford Text Archive." [OTA] *The Tale of Genji* [Electronic Resource]. Trans. Edward G. Seidensticker. N.p., n.d. Web. 16 Apr. 2017. <http://ota.ox.ac.uk/desc/2245>.
- Murakami, Janel R. Goodman. "Metonymy in The Tale of Genji: An Analysis of Translation Strategies." *Arizona Working Papers in SLA and Teaching* 20 (2013): 55-75.
- Moretti, Franco. "Network Theory, Plot Analysis." *Distant Reading*. London: Verso, 2015. N. pag. Print.