

**Regional and audience bias
in the Digha and Majjhima Nikaya**

Contents

1. Introduction.....	1
2. Research agenda.....	1
3. Data.....	3
4. Method.....	5
5. Related work.....	6
6. Experiment design.....	7
7. Results and discussion.....	8
8. Conclusion.....	10
9. References.....	11

1. Introduction

To define Digital Humanities may be to pose the *Big Questions* such as, “[W]hat were, are and will be the humanities in Western, and now world, culture?” (Gardiner & Musto 2015, 2) Another way is to acknowledge that people like to use the term widely, but usually disagree about what it means (Gardiner & Musto 2015, 4). Painting in broad strokes the outlines of the Digital Humanities territory, then, the term may be used to refer to humanities work with digital tools, humanities work on digital sources, classical humanities work disseminated by digital means, any combination thereof, or it may be redundant, implying either that humanists work with digital tools and sources anyway, or that the digital has transformed the field entirely (Gardiner & Musto 2015, 2ff.). Consequently, whether, and if so, in what ways the Digital Humanities are different from traditional humanities is just as much in question.

Similar to the debate about the externality of social facts between Durkheim and Tarde at the outset of last century (Greve 2015), adoption of any one definition can be inferred to draw inspiration from personal and career interests as much as a search for knowledge. Intrinsically, pinning the abstraction of a scientific field designation down has philosophical value at best. The term is an arbitrary social construction, however its definition comes with social and material implications. Based on one’s position in the web of academia, one will use any particular definition to gatekeep the “community” to the advantage of one’s preferred peers and validate one’s claims to prestige and grant money.

Leaving aside any of these factors, this project falls into the category of traditional humanities work with digital tools, more specifically the field of Buddhist studies.

2. Research agenda

The Pali canon is a collection of suttas (didactic discourses), monastic rules, poems, inspired utterances and more, orally transmitted for several centuries before being written down in the 1st century BCE (Britannica n.d.b). If and to what extent it is indeed a historical account of the lives and teaching activities of Gotama and his disciples is the subject of ongoing research and discussion. While the canon has clearly hagiographical elements, and shows the usual signs of editing and doctrinal tampering evident in the sacred texts of many religions, I will assume for the sake of argument that the parts examined in this paper are historical, i.e. a more or less reliable rendering of the original Buddhist sangha’s discourses, originating from the historical person of Siddhattha Gotama (see also Sujato and Brahmali’s (2015) argument in favor the Early Buddhist Texts’ authenticity¹).

1 The EBTs, although overlapping with large parts of the Pali canon, are not identical with it. While some parts of the canon are considered later additions, parallel versions of texts appearing in the canon exist in different languages, most importantly Chinese, and show some differences ranging from minor to substantial (Allon 2022). The comparison of Pali and Chinese texts has received much scholarly attention in recent years, and has inspired a reexamination of religious practice (see f.ex. Anālayo 2020a; 2020b).

The Sutta Pitaka (“basket of discourse”, i.e. the division of the Pali canon containing discourses) is organized in five Nikayas (collections) (Britannica n.d.c). Of these five, I have chosen the Digha and Majjhima Nikaya (DN and MN; collections of long and medium-length discourses) owing to their structure: each sutta contained in these collections is a clearly delineated, closed narrative, giving historical and geographical details of the context in which the discourse was given. In contrast, the Samyutta and Anguttara Nikaya are organized in ways that would necessitate extensive manual annotation, while the Khuddaka Nikaya contains mainly verse.

In this paper, I want to find out whether and to what extent the topics of discourses vary based on audience and country:

The Buddha and his disciples have been known to give different discourses depending on their audience, tailoring the contents to their interests, talents, and spiritual development, but generally reserving more advanced doctrinal content and meditative instruction for monastics. This discrepancy is acknowledged in MN143, a retelling of a break with convention: a devoted lay follower of the Buddha lies on his deathbed, and Sariputta and Ananda, two of the foremost monastic disciples, visit him to provide advanced meditative instructions:

[The layman:] “[...] [F]or a long time I have paid homage to the Buddha and the esteemed mendicants. Yet I have never before heard such a Dhamma talk.”

[Sariputta:] “Householder, it does not occur to us to teach such a Dhamma talk to white-clothed laypeople [devout followers of the Buddha who have made vows but are not monastics]. Rather, we teach like this to those gone forth [i.e. monastics].”

“Well then, Honorable Sāriputta, let it occur to you to teach such a Dhamma talk to white-clothed laypeople as well! There are gentlemen with little dust in their eyes. They’re in decline because they haven’t heard the teaching. There will be those who understand the teaching!”

(Sujato n.d.b)

As this break with convention and the layman’s pleas suggest, lay audiences up to that point were usually initiated in a limited number of doctrines. Whether that changed, and whether any variation in theme-audience combination is indeed reflected quantitatively in the canon, however, is not a given, and thus deserves some attention.

Less obviously, perhaps, one might ask whether discourse topics varied in different countries. Northern India in the Buddha’s time was divided into 16 great countries which often engaged in warfare (Sujato & Brahmali 2015, 15, 21f.). The Buddha did not closely align with any of them politically, however, he had numerous patrons in the upper echelons of their respective societies, including the royal family of Kosala (Dictionary of Pali proper names n.d.b) and the king of Magadha (Dictionary of Pali proper names n.d.a), the two largest countries at the time (the latter of which would go on to conquer all others after the Buddha’s death to form an empire (Sujato & Brahmali 2015, 21f.)). Even though the Buddha is portrayed as a transcendently

independent and uninterested party in the canon, his religion's striking success in spreading and acquiring monasteries may not exclusively stem from patrons' inspired faith, but from a political sensibility which could be reflected in the texts.

3. Data

I am using annotated and cleaned TSV datasets created by Charith Wijewardena (2022) from the English translations of the DN and MN by Bhikkhu Sujato hosted at SuttaCentral. In addition to organizing the text into small segments and larger sections, it provides IDs for each, the collection the sutta appears in, and titles and numbers of individual suttas.

Furthermore, I have added columns for the exact location given for each sutta and the country it belonged to. Even though suttas in the DN and MN start with a paragraph describing in stock phrases where they took place, making an automated approach viable in theory, variations in phrasing and spelling, as well as several instances of elaborate changes of locality before the actual discourse suggested a manual approach, noting down place names and comparing with a map of 500 BCE India, which proved feasible at under 200 suttas.

Another column indicating the type of audience was created semi-manually starting from a dataset created by michaelh-sc (2023) through Named-Entity Recognition using the Stanza toolkit by Stanford NLP. I have divided the audience into the categories monastic (ordained Buddhists), ascetic (serious followers of different religions), Brahmin (upper-class laypeople²), householder (non-Brahmin laypeople), royal, and supernatural (roughly equivalent to gods, nature spirits, and demons). This required a few judgment calls with suttas that describe several audiences being present at once, people appearing as humans first and being reborn as gods in the course of the narrative, or those which feature nested stories and/or longer stretches relayed by an omniscient narrator. With some difficulty, I have decided against an N/A tag, treating narration and speaking parts of non-Buddhists as part of the same didactic unit, counting even the words of the tempter-trickster figure Mara (who may be thought of as a mix of Goethe's Mephistopheles and Loki from Norse mythology) as (an antagonistic) part of teaching sequences. This tagging system is necessarily imperfect.

Finally, I added a per-line word count. At about 493,000 words, the contents of the two Nikayas just so supersede the epic length of Lord of the Rings, but fall short of War and Peace (wordcounter.net 2016). Creating plots for the newly minted country and audience variables, however, reveals the first major problem facing this analysis: despite the Nikayas' impressive scope in pure verbosity, they do not match the novels' adventurous diversity of location and characters.

2 Brahmanism was a religion in its own right, descending from the Vedic religion, and Brahmins were its hereditary priest class (Britannica n.d.a). Perhaps because of their religion and caste's safe entrenchment in the social order, however, many Brahmin characters in the canon are remarkably comfortable with paying respect to the Buddha and having friendly debates.

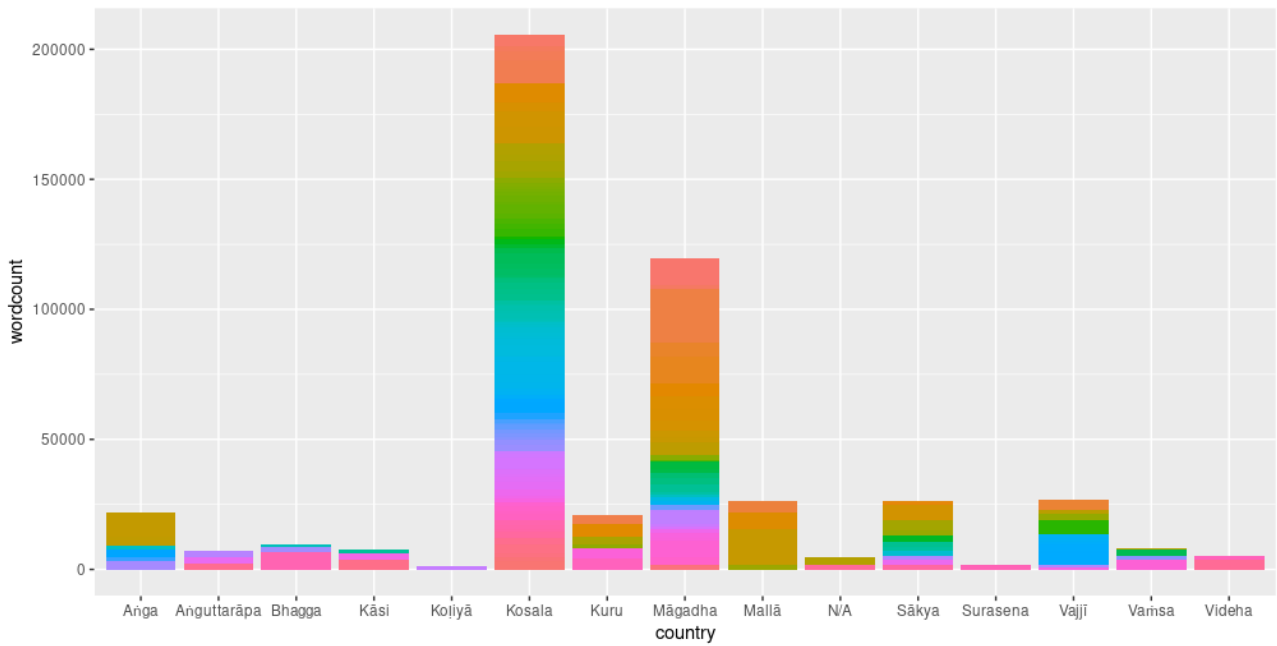


Figure 1: word count per country, colors indicating suttas

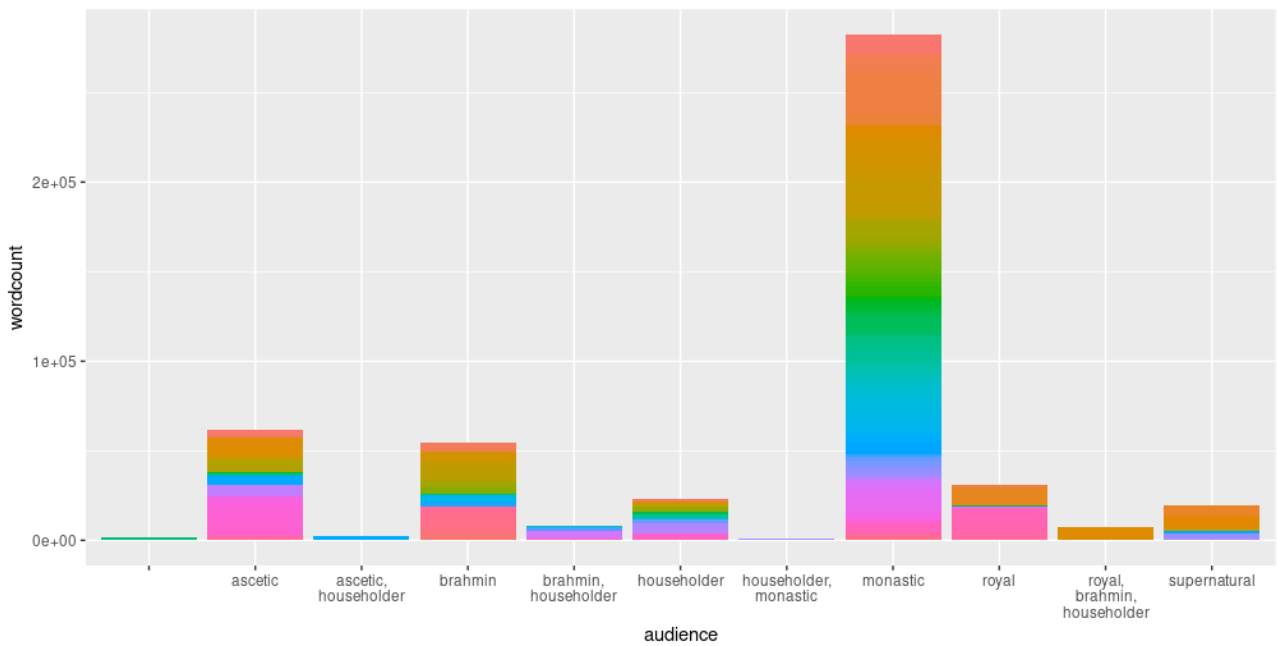


Figure 2: word count per audience, colors indicating suttas

Rather, they focus on Kosala and Magadha to the point where no other country makes it beyond novella length, and the majority of content is addressed at a monastic audience. While ascetics and Brahmins amass a respectable number of lectures, the Buddha and his disciples, according to the DN and MN, spent about as much time talking to common folk as they did conversing with kings and gods. This lopsidedness raises another question: Assuming the historicity of individual suttas, as indicated above, what about the historicity of their selection? Was Buddhism a religion of the rich and powerful while giving lip service to the primacy

of the existential condition³ over socio-economic strata? If it was not, how did this focus on the upper classes come about? And if the canon does not reliably represent the teaching activity of the original sangha, what does the insight gained from an analysis of the canon pertain to instead?

The bars' colors in the diagram reveal yet another, more mundane problem. Each hue stands for one sutta featuring a tag. As any one sutta tends to have a narrow thematic focus, and some bars are problematically non-rainbow-colored, there simply is insufficient data to answer the research question for a number of tags (Anguttarapa, Bhagga, Kasi, Koliya, Surasena, Vamsa, and Videha as well as combined audience tags and, arguably, royal and supernatural audiences).

All of these problems combine to call into question my approach to this project. The data may be unsuited to answer the research question, or a different methodological approach might be called for. A more thorough examination of the data before committing to this project would no doubt have revealed this, prompting either a widening of the corpus to include the rest of the Sutta Pitaka, or the adoption of a more modest research question.

4. Method

Topic analysis is a computational method under the umbrella of “distant reading” techniques. The term was coined by Franco Moretti in opposition to the “close reading” focus popular in the humanities around the 2000s, and is cited in the innovation of computational methods “that take into account more abstract, often statistical properties of text” (Elwert 2021, 165f.), thus “provid[ing] a suite of algorithms to discover hidden thematic structure in large collections of texts” (Blei 2012) in the form of topic modeling, situated within the field of probabilistic modeling.

The simplest (and most popular) topic model is LDA (Latent Dirichlet Allocation). It identifies clusters of terms that often occur in close proximity to each other. These topics are then correlated with each document of a corpus, determining a percentage to which the topic is present in the document, and each word is assigned a topic (ibid.).

3 According to the Buddha, not even the most exalted beings are immune to the effects of kamma and rebirth, going up and down the cosmic ladder indefinitely unless they achieve final liberation. In the famous mountain simile, he remarks:

“Suppose there were vast mountains
of solid rock touching the sky
drawing in from all sides
and crushing the four quarters.

So too old age and death
advance upon all living creatures -
aristocrats, brahmins, merchants,
workers, outcastes, and scavengers.
They spare nothing.
They crush all beneath them.”
(Sujato n.d.c)

In preparation for the modeling process with LDA, a text corpus is usually tokenized (i.e. removing punctuation and capitalization) and stemmed (i.e. replacing morphological variants of a word with its stem), and “stopwords” (a, the, and, etc.) to be ignored by the algorithm are defined (Brett 2012).

Blei asserts that topic modeling operates on the assumption that a text corpus is permeated by a hidden layer of meaning – hidden, it is implied, to a human (close) reader, who may possess the means to identify face-value meaning and any number of hidden meanings in individual texts, but lacks the raw computing power to identify hidden meaning evident over a large range of texts using distant reading. Ideally, however, the model would have to “understand”, for lack of a better word, the face-value meaning of texts in the corpus. Text using ciphers, metaphor, or esoteric language filled with allusions may be too complicated for an LDA model. Imagine, for example, three texts that deal with the movie *The Wizard of Gore*. The first is a standard review, using common terms such as director, actor, audience, etc. The second is a poststructuralist *critique* dedicated to the exclusive use of sophisticated language like *auteur*, *thespian*, *connoisseurs of the silver screen*, etc., perhaps even upgrading the movie’s title to *The Thaumaturge of Viscera*. The third is a satirical morality fable featuring a rabbit, a troupe of hedgehogs, and a penguin colony, implied to be the movie’s director, actors, and audience. An LDA model will fail to find the common topic of movies evident in the first text on a lexical basis, in the second text on a different lexical basis, and in the third as an allegory, since the texts, despite addressing the same thing, use different words and language structure. Failing to find this obvious thematic connection, a topic model might still point the way to some hidden connection between them, but its utility is much more doubtful than it would be with texts using roughly the same language.

5. Related work

Topic modeling has been used in numerous creative ways in religious studies. Three examples:

Building on previous humanities work arguing against the assertion that early Chinese textual sources display no distinction between mind and body, Slingerland et al. (2017) report several studies using computational methods (word collocation, hierarchical clustering, and topic modeling) on a corpus of texts spanning roughly 1800 years which yield further evidence for mind-body dualism in Chinese religious and philosophical writings.

Stine et al. (2020) compare two topic models created from corpora of the Christianity and Buddhist subreddits to identify structural rather than lexical similarities between discussed topics, finding, for example, a similarity between the abortion debate in the Christian subreddit and discussions about the ethics of meat consumption in the Buddhist subreddit (a connection which may sound unexpected at first, but is entirely sensible since the structure of pro-life and animal rights arguments exhibit marked similarities).

Chandra & Ranjan (2022) compare topic models of Upanishad and Bhagavad Gita translations created with the deep learning model BERT to track similarities between the two, testing the common knowledge claim that the latter summarizes messages of the former. The claim is validated by their finding of large thematic overlap between the two corpora.

6. Experiment design

Preprocessing the corpus, I employed the `corpus_tokens` function of the `Quanteda` package for R, which handled tokenization and, with the addition of a lemma dictionary and a stopword list, stemming and stopword removal. I then generated and added collocations to the corpus. From the corpus, I created a document term matrix, excluding words occurring in less than 1% of documents, as well as the common audience terms "mendicant", "brahmin", "brahman", "ascetic", "king", "god", "prince".

Calculating the LDA model (using the `topicmodels` package), I started out with 15 topics, and reduced them gradually, checking for interpretability, until I hit the point of least nonsense at 7. As the discussion of results below will show, even these are hard to interpret. I have considered two culprits: Excluding terms featured in several topics led to a further deterioration of interpretability. Changing document length was more likely to yield results, as the sections can vary considerably in the original dataset, especially with omitted repetitions common in translations of the canon. Experimentally reorganizing about a tenth of the corpus into thematic sections (dataset included in the repository), however, yielded even more nonsensical topics. Notably, one of them dealt entirely with the behavior of deer. This leads us back to the *Wizard of Gore* thought experiment from Section 4.

The Pali canon is structured in peculiar ways. Partly, this is due to its history, having been compiled from chants which were optimized for rote learning and oral transmission, and handed down in this way for several centuries before being committed to writing (Thanissaro 2013, ix). Optimization for oral transmission includes a musical form featuring frequent repetition with or without minor variation, often in a verse-chorus pattern, and extensive use of identical stock phrases found throughout the suttas.

Another part of its peculiar structure is to be attributed to the Buddha's teaching style. A large chunk of the canon's raw word count is occupied by similes. One example for this is MN25 (Sujato n.d.d), which gives a particularly long simile about deer who show various degrees of cleverness in dealing with a trapper's bait. These are then explained to represent meditators and the ways they deal with the bait of sensual stimulation laid out by Mara (the trickster, see above) to keep them from reaching final liberation. The model, being unable to link the explanation to the simile, treats them as separate lexical clusters, in effect making any simile into its own potential topic. Every time this happens, the signal to noise ratio gets worse - and it happens often. In addition, the canon makes extensive use of stock phrases as coded reference for doctrinal concepts: an arahant (fully awakened being) is often hailed as one of the perfected, their mind (or just they themselves) called a thoroughbred (horse); someone who enters jhana (states of deep meditative absorption) dwells secluded from greed and distress with reference to the world, etc.

So, to rephrase an example for a bad research question: Can this corpus (even) be analyzed (in a way that yields valuable insight) using this method? Sophisticated deep learning models may be more successful, but (getting ahead of the discussion below) a simple LDA model is likely unsuited to penetrate the canon's coded language.

7. Results and discussion

A repository containing all relevant files can be found at https://github.com/latukika/dn_mn_bias

Some of the topics are more comprehensible than others. First, the good:

buddha time sir sit speak side hear stay reply ananda comprises the most common terms in an opening paragraph like the following:

“So I have **heard**. At one **time** the **Buddha** was **staying** near Sāvattḥī in the Eastern Monastery, the stilt longhouse of Migāra’s mother.

And then Sakka, lord of gods, went up to the Buddha, bowed, stood to one **side**, and **said** to him:”

(Sujato n.d.c)

Ananda, who is the Buddha’s personal attendant and, despite being an accomplished meditator himself, often plays audience stand-in by appearing endearingly sensitive and slow on the uptake, gets relegated to this topic, as he is often present in these narrative paragraphs.

Curiously, the model has also generated another topic for narrative paragraphs, suppose eat day food time water person make life body, which have some overlap with opening paragraphs, but tend to occur later in a sutta and often deal with meals.

reborn perception dimension consciousness realize death form end sentient kind obviously has something to do with rebirth, but rather than including kamma and ethics, as one would expect, it includes more basic doctrinal content on consciousness. This may be hidden structure or simply an aberration. As there is no topic clearly relating to ethics (neither with a 7 topic model nor with higher numbers), the latter is more likely.

mind give desire understand meditate body arise pain pleasure may be related to a meditative understanding of how the workings of the mind give rise to pain and pleasure. suffer feel thing cessation understand realize body kind cease condition seems to show large thematic overlap.

teach mind train absorption remain noble thing disciple teacher place may deal with meditative aspects discussed around the formulaic jhana passages depicting how a meditator, based on certain conditions being satisfied, enters the enumerated states of absorption.

good give live people view teach lie life homelessness thing sir may be interpreted as talking about the difference of lay life (butchered by stemming to become the lie_life 2-gram) and mendicant life (represented by the code word homelessness), and the relationship between the two, where lay people are taught the dhamma by monastics and support them materially.

These are merely interpretations of the top 10 signal words. Looking into documents categorized into any one topic with 90% affinity reveals some serious problems:

1. Despite organizing the textual data so that one section is treated as a single document, the returned documents are single segments instead. Whether that means that the topic model was created on the basis of these far too short segments as well, I have no idea.

2. Segments are often categorized in ways that make no intuitive sense, as they would fit better into a different topic, such as parts of opening paragraphs outside the opening paragraph topic, or some of the paragraphs most clearly related to meditative absorption in the rebirth topic.
 3. Topics are expectably infiltrated by unrelated passages too diverse to fall into a unified dump topic, such as an enumeration of torture methods found in the absorption topic.
 4. Things that could very well form their own topic get categorized into others, such as formulaic closing paragraphs appearing in meditation-related topics.
 5. With doctrinal topics, despite their signal words having a clear connection to themes evident in the canon on close reading, very few documents that actually relate to this topic are displayed.
- This all amounts to the insight that while it does paint in very broad strokes some interpretable larger tendencies in the corpus, this model is so inaccurate as to be meaningless as a basis for analysis, and I don't know how to fix it. Topic distributions per country and audience are thus meaningless as well.

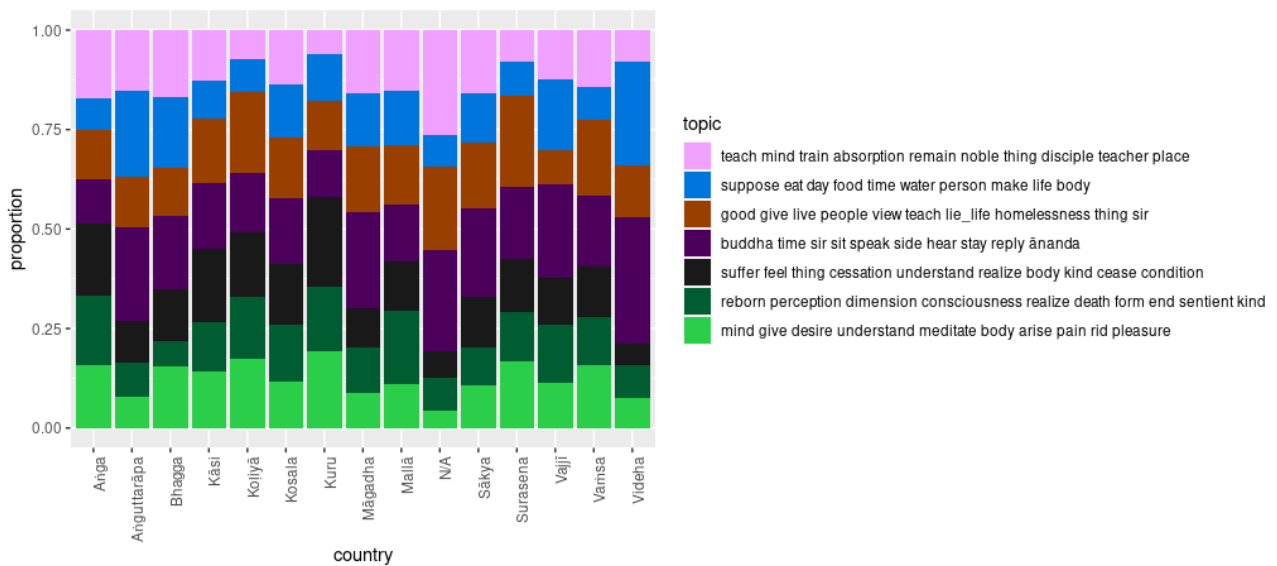


Figure 3: topic distribution per country

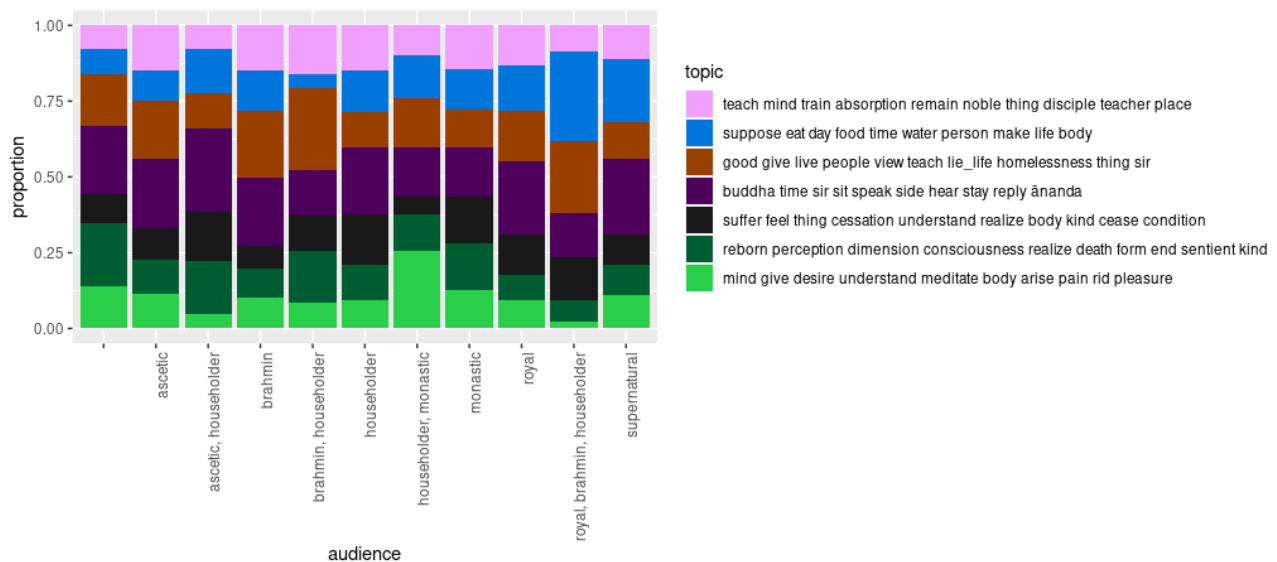


Figure 4: topic distribution per audience

Unreliable as the topic model may be, going over the corpus several times for the purposes of tagging and organizing into thematic paragraphs has yielded some insight into the questions.

1. As the two major Buddhist monasteries of the canon are located in Kosala and Magadha, the discourses held there will naturally differ from other countries, being mostly addressed at the monastic sangha. So not only does the disparity in data hamper research into the original question, the question itself (topic bias by country) is unsuited to address the underlying suspicion of political leanings.
2. Of the discourses addressed at Brahmins and householders, many deal wholly or in part with advanced doctrinal topics and meditative instruction, while comparatively few deal with kamma and ethics. It may be that the canon was simply compiled with an emphasis on these advanced discourses. It also displays some preoccupation with refuting wrong view and advertising the mendicant sangha's virtue and spiritual prowess in the presence of Brahmins, householders, and ascetics of other sects, which usually requires some dazzling doctrine or descriptions of deep meditative states. Most of these refutations end with lay people being converted and ascetics attaining arahantship, some over the course of two or three suttas. Thus, there seems to be a thematic difference depending on audiences, but centered on refutation of wrong view vs. monastic matters rather than shallow vs. deep doctrine, as I originally suspected based on MN143.

8. Conclusion

Setting out to find regional and audience bias in the Digha and Majjhima Nikaya, analysis of the data revealed a sizable imbalance in favor of major monastery locations and monastic audiences which makes topic model analysis on the basis of this particular corpus ill-suited for answering this particular research question as there is not enough data for many of the tags.

Going ahead nevertheless revealed more fundamental problems. LDA modeling may be unable to penetrate the highly coded language of the canon, requiring more sophisticated models. The actual model created from the corpus did not even reliably identify topics which could conceivably be identified using LDA.

Reviewing statistical metrics of the corpus suggests that the question of regional bias does not address any suspicion of political bias on the part of the sangha, as it is distorted by the overrepresentation of major monastery sites, while close reading done prior to this project and during manual tagging and reorganization work leads me to hypothesize that the variation of topics between lay and monastic audiences does not quantitatively represent the disparity in doctrinal depth alluded to in MN143.

9. References

- Allon, Mark. 2022. „Early Buddhist Texts: Their Composition and Transmission“. *Journal of Indian Philosophy* 50 (4): 523–56. <https://doi.org/10.1007/s10781-021-09499-6>.
- Anālayo, Bhikkhu. 2020a. *Mindfulness of Breathing*. Windhorse Publications.
- . 2020b. *Mitgefühl und Leerheit in der früh-buddhistischen Meditation*. Übersetzt von Hedwig Kren und Robert Grosch. Berlin: edition steinrich.
- Blei, David M. 2012. „Topic Modeling and Digital Humanities“. *Journal of Digital Humanities* 2 (1). Accessed March 7, 2023. <https://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/>.
- Brett, Megan R. 2012. „Topic Modeling: a Basic Introduction“. *Journal of Digital Humanities* 2 (1). Accessed March 7, 2023. <https://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/>.
- Britannica, The Editors of Encyclopaedia. n.d.a „Brahmanism“. In *Encyclopedia Britannica*. Accessed March 7, 2023. <https://www.britannica.com/topic/Brahmanism>.
- . n.d.b „Pali canon“. In *Encyclopedia Britannica*. Accessed March 7, 2023. <https://www.britannica.com/topic/Tipitaka>.
- . n.d.c „Sutta Pitaka“. In *Encyclopedia Britannica*. Accessed March 7, 2023. <https://www.britannica.com/topic/Sutta-Pitaka>.
- Chandra, Rohitash, und Mukul Ranjan. 2022. „Artificial Intelligence for Topic Modelling in Hindu Philosophy: Mapping Themes between the Upanishads and the Bhagavad Gita“. Edited by Fu Lee Wang. *PLOS ONE* 17 (9): e0273476. <https://doi.org/10.1371/journal.pone.0273476>.
- Dictionary of Pali proper names. n.d.a „Definitions for bimbisāra“. Accessed March 7, 2023. <https://suttacentral.net/define/bimbis%C4%81ra>.
- . n.d.b „Definitions for pasenadi“. Accessed March 7, 2023. <https://suttacentral.net/define/pasenadi>.
- Elwert, Frederik. 2021. „Computational Text Analysis“. In *The Routledge Handbook of Research Methods in the Study of Religion*, edited by Steven Engler and Michael Stausberg, 2nd Ed., 164–79. London: Routledge. <https://doi.org/10.4324/9781003222491-12>.

- Gardiner, Eileen, und Ronald G. Musto. 2015. *The Digital Humanities: A Primer for Students and Scholars*. 1st Edition. Cambridge University Press. <https://doi.org/10.1017/CBO9781139003865>.
- Greve, Jens. 2015. „Durkheim und Tarde: Holismus, Individualismus oder dritte Wege?“ In *Reduktiver Individualismus*, edited by Jens Greve, 29–42. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-06557-7_2.
- michaelh-sc. 2023. „Sutta metadata“. Accessed March 7, 2023. <https://github.com/michaelh-sc/suttamap>.
- Slingerland, Edward, Ryan Nichols, Kristoffer Neilbo, and Carson Logan. 2017. „The Distant Reading of Religious Texts: A “Big Data” Approach to Mind-Body Concepts in Early China“. *Journal of the American Academy of Religion* 85 (4): 985–1016. <https://doi.org/10.1093/jaarel/lfw090>.
- Stine, Zachary K, James E Deitrick, and Nitin Agarwal. 2020. „Comparative Religion, Topic Models, and Conceptualization: Towards the Characterization of Structural Relationships between Online Religious Discourses“. *CHR 2020: Workshop on Computational Humanities Research*.
- Sujato, Bhikkhu. n.d.a „MN37“. Accessed March 7, 2023. <https://suttacentral.net/mn37/en/sujato?layout=plain&reference=none¬es=asterisk&highlight=false&script=latin>.
- . n.d.b „MN143“. Accessed March 7, 2023. <https://suttacentral.net/mn143/en/sujato?layout=plain&reference=none¬es=asterisk&highlight=false&script=latin>.
- . n.d.c „SN3.25“. Accessed March 7, 2023. <https://suttacentral.net/sn3.25/en/sujato?layout=plain&reference=none¬es=asterisk&highlight=false&script=latin>.
- . n.d.d „MN25“. Accessed March 7, 2023. <https://suttacentral.net/mn25/en/sujato?layout=plain&reference=none¬es=asterisk&highlight=false&script=latin>.
- Sujato, Bhikkhu, and Bhikkhu Brahmali. 2015. *The Authenticity of the Early Buddhist Texts*. <https://ocbs.org/the-authenticity-of-the-early-buddhist-texts-2/>.
- Thanissaro, Bhikkhu. 2013. *Wings to Awakening: An Anthology from the Pali Canon*. 7th Edition. Barre: Dhamma Dana Publications.
- Wijewardena, Charith. 2022. „sutta-science“. Accessed March 7, 2023. https://github.com/chaz23/sutta-science/tree/main/data/sutta-translations/dataset_2.
- wordcounter.net. 2016. „How Many Words Is War and Peace?“ Accessed March 7, 2023. https://wordcounter.net/blog/2016/10/28/102640_how-many-words-is-war-and-peace.html.