One of the most important decisions that needs to be made before conducting a topic modeling analysis is to decide an optimal number of topics. Most researchers suggest experimenting with different numbers to evaluate the quality of topic outputs. On one hand, analyzing larger data sets with a low number of topics can result it topics that are too general, missing out on the granularity of the data. On the other hand, using too many topics will generate outputs with topics too rare and non-representative. For the current analysis, I experimented with 10, 30, 50 and 100 topics. The outputs generated using 10 and 30 topics were more interesting than those of 50 and 100. For this reason, the below analysis, will include excerpts from using 10 and 30 topics. The first table shows most representative and unique terms across 9 top topics. Then I included terms that uniquely appeared in each topic in the Terms column. Lastly, I recorded the subjective topic titles for each set of terms in the third column.

|  |  |  |
| --- | --- | --- |
| **Topics** | **Terms** | **Topic title** |
| Topic 1 | Женский - female  Население - population  Девушка - young woman  Положение - status  Муж - husband  Мужчина - man  Oтношение - relationship | женщина в обществе  women’s role in the community |
| Topic 2 | Население - population  Муж - husband  Ребёнок - child  Мужчина - man  Положение - relationship | роль женщины в семье  the role of women in the family |
| Topic 3 | Женский - female  Ребёнок - child  Oбласть - province | Мать  mother |
| Topic 4 | Область - province  Ребёнок - child  Революция - revolution | предстоящие перемены  upcoming changes |
| Topic 5 | Население - population  Семья - family  Ребёнок - child | Демография  demographics |
| Topic 6 | Край – kray (bigger than province)  Экономический - economic  Tруд - labor | женский труд в экономике  female labor in the economy |
| Topic 7 | Хозяйство - farm  Борьба - fight | Борьба с предрассудками  Fights against prejudices |
| Topic 8 | Oбраз – portrayal  Женский - female | представление женщинах  Portrayal of women |
| Topic 9 | Bласть - power  Oтношение - relationship | Oтношение c властью  relationship to power |
| Topic 10 | Not representative |  |

The most common topics according to this table are women’s role in community and family. The appearance of these topics was an expected outcome of topic modeling, sequence of their appearance in the results remind of the organizational structure of the brochures. Close reading of the texts using the term “status”, revels peculiar findings.

Finding 1. “**Положение** черкешенки до Октябрьской революции. Вся жизнь черкешенки до революции была обставлена рядом стеснительных обычаев, часто совершенно варварских по характеру.”

Translation: “The **position** of a Circassian woman before the October Revolution. The entire life a Circassian woman before the revolution was surrounded by a myriad of embarrassing customs, often completely barbaric in nature.” – “Cherkeshenka” brochure, p. 14

Finding 2. “Целый ряд стеснительных правил закреплял это **положение**. «Женщина не должна громко петь в присутствии посторонних мужчин. Молодые женщины должны говорить мало, вполголоса, не смеяться громко”.

Translation: “A whole series of embarrassing customs reinforce this **position**. “A woman should not sing loudly in the presence of strange men. Young brides should speak little, in a low voice, and not laugh loudly.” “ЯКУТКА” brochure, p. 23

Looking at the context in which the word “position” is used in each case, reveals a different circumstance. Often in the first two chapters, it is used to illustrate the oppressiveness of women’s position, using specific examples drawn from traditional practices. Adjectives associated with those traditions, such as “barbaric” and “embarrassing”, clearly transmit the judgment of the authors. This finding reveals the motivation of the ethnographers who authored the brochure and the civilizing mission of the Soviet regime which was aimed at freeing women of the East from their pre-revolutionary misery.

In some instances, the term “position” describes the status and credibility of other local authorities, like religious ulama or traditional healers. In rare instances, the authors highlight a positive pre-revolutionary status of women, like in Votak brochure shown in Example 3. Nevertheless, the except from Votak brochure is an excellent support for Chernyaeva’s assertion that authors exaggerated and at time fabricated some characteristics of the “ethnic backwardness” to dramatize the pre- and post-revolutionary impact of the regime.

Example 3: “Не в пример женщинам соседних народностей, вотячка и некоторых отношениях пользовалась и до революции большой свободой…Но, на ряду с этим, в ее быте есть много отрицательного...”

Translation: “Unlike women of neighboring ethnicities, the Votyak woman enjoyed great freedoms even before the revolution... But, along with this, there were many negative aspects in her everyday life…” “Votachka” brochure, p. 17

Topic 3 demonstrates that one of the main roles of women within their respective societies was baring children. Comparatively, Topic 6 suggests signals to discussions about the economy of labor, which is an ambiguous theme. To understand the meaning of it, I close read the texts and concluded that two types of economies have been described in the brochures: one, the primary income generating occupation within each community, and second, the economic dependency of women on their husbands.

What are the main topics that are shared among brochures? Do they match to the results of the close reading?

How are the Muslim women, who represent 13 out of 30 ethnic groups accounted for in the brochures, described in comparison to their non-Muslim counterparts? How does the topic modeling demonstrate these discrepancies? Are there any discrepancies that came up in the topic modeling outputs?

What decisions are made to determine which stopwords to keep and which to remove in the pre-processing? How do these decisions impact the topic modeling output?

* Considering that topic modeling is designed to analyze representative or thematic samples of large data sets, what the benefits of utilizing it with a small corpus?
* A finding is that this didn’t work to locate the religious minorities I hope to find
* It is talking about talking about the Muslim women in the same way as everyone else, so they are not racialized
* How do these women get propagandized in other in another corpus – future work – next step and a possible inclination