

Computer-Assisted Analysis of Manas Narratives: Demonstrations and Directions

James PLUMTREE¹

Abstract

This paper shows how computer-assisted text analysis can greatly aid the field of Manas Studies. Though general patterns can be noted through traditional means of analysis, quantifying textual variants sidesteps the use of vague general impressions by providing statistical data that can permit more specific and nuanced insights. In addition to illuminating individual performances and performers via comparison, this method can further research into the topic including attribution, and measuring the impact of variables. The paper then proposes possible research into specific elements to deal with an ever-growing corpus, and stresses the potential findings and fresh perspectives from using large-scale computer-assisted corpus analysis.

Keywords: Manas Epos, Digital Humanities, quantitative analysis, 'Ever-Increasing Juk'

Manas Anlatılarının Bilgisayar Destekli Analizi: Gösteriler ve Yol Tarifleri

Öz

Bu makale, bilgisayar destekli metin analizinin Manas Çalıřmaları alanına nasıl çok büyük bir katkıda bulunabileceğini göstermektedir. Her ne kadar genel örüntüler ve kalıplar geleneksel analiz araçlarıyla çalıřılabilir de, metinsel varyantların nicelleřtirilmesi daha özgün ve detaylı kavrayıřları mümkün kılacak istatistiksel veriler sağlayarak müphem genel kanaatlerin ötesine geçmemizi sağlar. Bu yöntem, bireysel icralara ve icracılara karşılařtırılmalı bir şekilde ışık tutarken, ilişkilendirme ve deęiřkenlerin etkisinin ölçülmesi de dâhil olmak üzere konuyla ilgili daha fazla arařtırma için bir zemin sunabilir. Kısacası bu makale, sürekli büyüyen bir literatürle başa çıkmak üzere belirli unsurlar üzerine olası arařtırmalar önermekte ve büyük ölçekli bilgisayar destekli literatür analizinden elde edilebilecek potansiyel bulguları ve yeni bakıř açılarını vurgulamaktadır.

Anahtar Kelimeler: Manas Destanı, Dijital Beřeri Bilimler, Niceliksel Analiz, 'Sürekli Artan Juk'

Atıf İin / Please Cite As:

Plumtree, J. (2021). Computer-Assisted Analysis of Manas Narratives: Demonstrations and Directions. *Manas Sosyal Arařtırmalar Dergisi*, 10(3), 1506-1515.

Geliř Tarihi / Received Date: 27.07.2020

Kabul Tarihi / Accepted Date: 19.05.2021

¹ Dr. - American University of Central Asia, plumtree_j@auca.kg

 ORCID: 0000-0001-5546-7749

Introduction

This research paper shows how computer-assisted analysis of Manas narratives can aid scholarly appreciation and understanding of the tradition. In addition to the unmeasurable number of oral performances that went, and continue to go, unrecorded, these Kyrgyz-language texts exist in a sizable number of variants, numerous amounts of media, and even in a large number of scripts (Arabic, Cyrillic, Latin, and Chinese). It would be an understatement to say that the stemma of the tradition is complex. With the aid of computer-assisted analysis, the breadth and depth of this material can be re-examined, categorised, and assessed.

To give a sense of the variety of the material, a brief outline follows. In this work, the word ‘performer’ refers to someone who constructs a text in performance or authors a narrative. The corpus includes versions transcribed by foreign scholars from oral performance in the mid-nineteenth to early twentieth century (Radloff, 1885; Almásy, 1911-1912; Hatto, 1977; 1990; Botoyarov, 1996), variants copied by local Kyrgyz literati near the turn of the twentieth century (Prior, 2002; Öztürk, 2015), recorded on early sound cylinders (Prior, 2006), manuscripts written by Soviet scholars and *manaschis* – performers of Manas narratives – themselves (Musaev, 1995; National Academy of Sciences, 2016; Plumtree, 2021a), chapbooks containing select narratives (Sarybekov, 1994; Heide, 2015), a compilation of multiple versions into a ‘harmonized’ complete narrative (Yunusaliev, 1958-1960), entire narratives by individual *manaschis* (Orozbekov, 1978-1982; Karalaev, 1984-1991; Mamay, 1984-1994), oral performances in post-Independence Kyrgyzstan (Reichl, 2016; AKYN, 2018), to publications prepared by the *manaschis* themselves (Bakchiev, 2011; 2012). Material also exists from surviving traditions in the Kyrgyz diaspora in the Pamirs (Dor, 1982) and in Xinjiang (Reichl, 2019; Jacquesson, 2020). Choosing what to study shapes the conclusions: in addition to individual texts and eras, one could also study individual performers (Jumaturdu, 2016) or ‘schools’ (Kydyrbaeva, 1984; Reichl, 1992). Atop of these issues, there is the issue of size: a ‘complete’ version of the life of Manas by Orozbekov totalled 180,378 lines; a life of Manas and his descendents by Karalaev amounted to 500,566 lines; a recently discovered version by Jüsüp Mamay is reputedly around 570,000 lines (Musaev, 1995; Liu, 2014; Jumaturdu, 2016). Careful assessment of even just one of these versions would constitute an enormous amount of labour; attempting a comparison would be even greater; determining whether the variant was the best, or the norm, would be speculative.

The substantial amount of material collected over a notable time period is of great significance, but such a wealth and variety makes it difficult to assess. Selections of the corpus can illuminate both aspects of Kyrgyz history and culture – such as the early formation of Kyrgyz national identity (Prior, 2002; Duishembieva, 2015) and the impact of external elements on Kyrgyz attitudes to their traditions (Plumtree, 2021a) – and subjects like orality, textualization, and epic. The sheer abundance has led to excessive generalization (and frequently incorrect hyperbole and misunderstanding) about the tradition. To help picture the issue, imagine an immense *juk* (жук), the pile of fabrics (blankets, quilts, pillows, and such), in a yurt. Each time you see the *juk*, the pile gets larger. If a visitor wanted to see a quilt or blanket, you would happily either take the one off the top (the most recent), or the most treasured (the one said to be the best). Now, the visitor can see the immense wealth due to the ever-increasing number of textiles, but if he or she was to ask whether each blanket or quilt was the same size, design, or had the same stitch-pattern, you would balk at the effort required to examine each work. Moreso, the ever-increasing collection would make it difficult for you to even observe each work and make a close assessment. The contents of the *juk* are parallel to the ever-increasing number of Manas variants: how can one speak of a tradition, or what is typical, when only a handful of known variants are examined? (And here it must be stressed the focus is on the recorded and/or published variants that are an anomaly in the tradition.) Though the most recent or the most treasured is briefly pleasing, and the size of the ever-increasing number of items impressive, how do the two relate? This is additionally problematic with Manas narratives given the changing attitudes, appreciation, and even sense of what is a Manas ‘text’ given the seismic social, political, and cultural changes to which the tradition has been subject (Prior, 2000; Plumtree, 2021a). To further the *juk* comparison: the type of quilts that are available to you to add to the pile are likely different to those that were available to your grandmother. Your taste in quilt design might also differ from your grandmother, and your perception of what your grandmother’s taste in quilt design could similarly differ from what her taste actually was. As with the *juk* and the Manas variants, up close comparison consequently will say more about the viewer than about the selection and the collection. With each new variant, the ‘Ever-Increasing Juk’ becomes more difficult to comment upon.

The potential value of computer-assisted analysis was shown in a pilot Digital Humanities project at the Analyzing Kyrgyz Narratives (AKYN) Research Group based at the American University of Central Asia (AUCA). In 2017-2018, wishing to explore how technology can aid the study and conservation of Kyrgyz oral poetry concerned with the legendary hero Manas and his descendents, six recordings were made of two prominent contemporary *manaschis*. Three performances narrating the birth of the titular hero were recorded on separate occasions at AUCA from Talaantaaly Bakchiev (b. 1971), and, likewise, from Doolot Sydykov (b. 1983). These, respectively **T₁**, **T₂**, **T₃**, **D₁**, **D₂**, **D₃**, were later transcribed and are in the process of being made available online (AKYN, 2018). Preliminary investigations of this material wanted to determine whether analysis of these transcripts could confirm the Parry-Lord theory of oral-formulaic composition (Parry, 1971; Lord, 2003) was applicable for Kyrgyz epic poetry. This hypothesis was confirmed by the use of simple plagiarism-spotting programmes and software designed to recognize repetitions in texts. This revealed the composition methods of the two *manaschis* while highlighting the relationship between contemporary post-Soviet performers and printed variants of their Soviet predecessors (Plumtree, 2019; Plumtree & Jeenbekova, 2020). These preliminary studies not only indicated that computer-assisted analyses are possible, but also revealed the potential rewards of such enterprises.

This research article shows the applicability of another method of computer-assisted analysis, and discusses further potentially fruitful Digital Humanities projects that would promote, and provoke, fresh scholarly approaches in researching intangible cultural heritage and examining a living oral tradition. The method presented below is a practical demonstration incorporating two open-source applications, Voyant Tools and Python (using Jupyter Notebook). These programmes calculate word frequencies, vocabulary density, and type-token ratios – information can be visualised and used for comparative purpose, to prompt further questions, and to potentially attribute a text to an individual performer. Since AKYN had three recordings each from two *manaschis*, the question arose as to whether computer-assisted analysis of these texts could be used to stylometrically determine the ‘author’ of the variant. The positive result of this research establishes the potential of such methods. Consequently, the final section of this research paper suggests possible approaches to address the ‘Ever-Increasing Juk’ of Manas variants. By using computer-assisted analysis to examine specific details across the corpus, it becomes possible to map and manage the huge understudied and underappreciated corpus of Manas texts. These suggestions, which would previously be quick to dismiss as speculative and wishful, are becoming increasingly possible to realize with technology and reliable, digitalized, editions of the variants. The intention behind showing these methods is to promote, and provoke, fresh scholarly approaches in researching intangible cultural heritage and examining a living oral tradition.

Method

Watching Bakchiev and Sydykov perform, the question arose as to what extent these performances differed and how this could be shown. Though Bakchiev appeared more concise and more specific in his vocabulary, and Sydykov gave the impression of being expansive and using a larger vocabulary, such thoughts were merely speculative; a more exact means of comparison was necessary. Later, seeing the six transcriptions, the hypothetical question arose as to how one would in the future confirm which text was created by which *manaschi*. This likewise needed to be resolved without resorting to personal judgments or vague generalizations. A computational quantitative approach, which could efficiently provide information that would almost be impossible to obtain manually, was required. With audiovisual recordings made by Kamila Baimuratova in the AUCA recording studios, transcriptions could be later produced by Alymkan Jeenbekova (AKYN, 2018).

Such information can be obtained by using Voyant Tools (Rockwell, & Sinclair, 2016a; 2016b). This online text analysis resource calculates the number of unique words, the average number of words per sentence, and the vocabulary density. The latter is the type-token ratio (TTR): the ‘types’, the number of unique word forms, divided by the ‘token’, the total number of words. The lower the density, the less frequent are new words. For example, ‘To be or not to be’ has four types, six tokens, and therefore has a vocabulary density of 0.667; ‘To be or not to be that is the question’, has eight types, ten tokens, and therefore has a higher vocabulary density of 0.8. These statistics can then be used for comparative purposes. Two significant elements needed to be taken into consideration. Firstly, sentence length was of little use, since punctuation was provided by the transcriber and not the oral performer. Secondly, the length of the performances differed. To address this, a fixed sample size was chosen for comparative purposes: the first hundred lines (FHL) of each variant. This would help indicate whether not taking into account lengthy when comparing different sized texts could skewer the result.

Further analysis can be provided by using another piece of open-source software: a Python programme run via Jupyter Notebook. This builds upon the previous problem of text length by measuring the type-token ratio within selected fixed parameters (such as every 200 words), and plotting them onto a graph for visual comparison. Hypothetically, if the computer-analysis is capable of assisting ‘authorship’ of the six AKYN transcriptions, there should be two distinct groups on the graph: one of Bakchiev’s performances, the other Sydykov’s.

Findings

Use of Voyant Tools for the six AKYN transcriptions resulted in the statistical calculations presented in Table 1. The transcriptions of Bakchiev’s performances (**T₁**, **T₂**, **T₃**) are set aside those of Doolot Sydykov (**D₁**, **D₂**, **D₃**). (This table also corrects some of the data previously misprinted in (Plumtree, 2021b).)

Table 1. Comparison of Transcriptions of Performances by Talantaaly Bakchiev and Doolot Sydykov

	Recording Length	Number of Lines	Total Words	Unique Word Forms	Vocab. Density	Average Words per Sentence
T₁*	36 m 48 s	973	3,275 (3,364)	1,037 (1,075)	0.317 (0.320)	32.4 (30.6)
T₂	35 m 59 s	1,016	3,315	1,143	0.345	35.6
T₃	69 m 39 s	1,851	6,284	1,671	0.277	37.4
D₁	56 m 23 s	1,760	5,512	1,860	0.337	51.5
D₂	70 m 15 s	2,157	6,947	2,001	0.288	128.6
D₃	85 m 46 s	2,716	8,765	2,379	0.271	486.9

* (numbers in brackets includes Bakchiev’s spoken introduction in the calculation)

These numbers appear to give weight to the speculative hypothesis mentioned earlier in the ‘Method’ section. Though Sydykov’s variants have the greater number of words and the highest number of unique forms, one of Bakchiev’s transcriptions (**T₂**) has the highest vocabulary density. Such results, it must be stressed (also mentioned in ‘Method’ above), are dependent on the length of the original text, and on the quality of the transcription. The far right column on the table above, ‘Average words per sentence’, makes this issue apparent: the statistic is not reflective of the oral performer (who, obviously, does not use punctuation in performance), but the consistency (or workload) of the transcriber. The information that Voyant Tools provides however can, with certain qualifications, allow for more concrete claims to be made about performances and performers.

The concern with Table 1 that, while they compare different versions of the same part of the Manas epos, each variant is of a different length, can be addressed as follows. The greater the length of a text, the more likely the vocabulary density will decrease. One way to accommodate this is to select a particular sample size. Comparing overall totals for texts with the results for the first hundred lines (FHL) of each variant, it becomes apparent that not taking into account length when comparing different sizes of text can skewer the result. As Table 2 shows, all of Sydykov’s variants (**D₁**, **D₂**, **D₃**) have a higher vocabulary density than Bakchiev’s variants (**T₁**, **T₂**, **T₃**).

Table 2. Comparison of Transcriptions of Talantaaly Bakchiev and Doolot Sydykov Including FHL

	Total Lines	Total Words	Total Unique Word Forms	Total Vocab. Density	Words in FHL	Unique Words in FHL	Vocab. Density of FHL
T₁*	973	3,275 (3,364)	1,037 (1,075)	0.317 (0.320)	340	200	0.588
T₂	1,016	3,315	1,143	0.345	332	218	0.657
T₃	1,854	6,248	1,671	0.267	328	204	0.622
D₁	1,760	5,512	1,860	0.337	331	228	0.685
D₂	2,157	6,947	2,001	0.288	330	223	0.676
D₃	2,716	8,765	2,379	0.271	342	226	0.661

* (note **T₁** excludes the spoken introduction in the calculation)

Another tool to clarify this preliminary finding is textual analysis done by Python via Jupyter Notebook. This measured the type-token ratio within selectable fixed parameters. The result is Figure 1. As the graph shows, when plotting the six texts in this manner, two distinct groupings are visible. Promisingly, these correspond with the performer. In addition to showing distinctions between the two performers, this chart could, in a hypothetical case, be used to identify the performer of one of these variants.

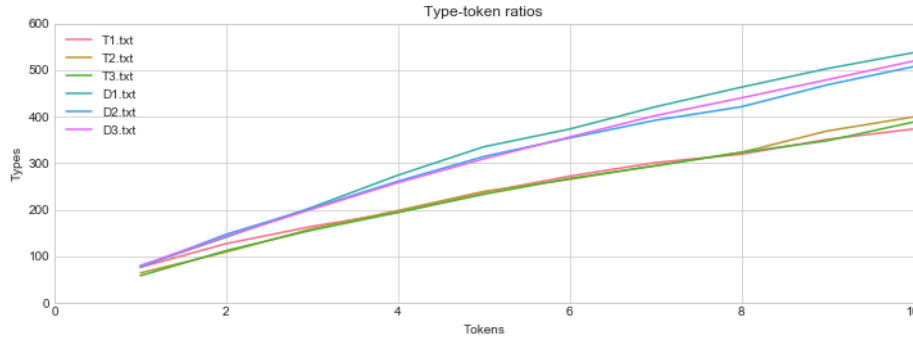


Figure 1. *Type-Token Ratio Chart*

This brief demonstration shows that computer-assisted textual analysis can be productively used on transcripts of oral poetry. Voyant Tools helps identify particular traits in a performance and a performer that, in turn, can be visually aided by Python. As this controlled experiment has shown, this method can potentially be used to identify the performer of a variant.

Discussion, Conclusion, and Recommendations

Before discussing and recommending potential Digital Humanities projects regarding the entire Manas corpus, a short comment is necessary regarding the potential uses of the method shown above. The process can be used for a variety of purposes, and on a larger scale. For instance, it can be used to compare prominent Soviet-era *manaschi* (Orozbekov, 1978-1982; Karalaev, 1984-1991) to determine their key characteristics. (Comparison of Soviet editions with recent editions, such as Orozbekov, 2010, could similarly reveal differences in editorial choices and, potentially, the context of the edition). In addition to calculating the vocabulary density of individual performers, comparisons using such data could also establish whether such features are shared between performers from the same background (birthplace, social class, education, literacy level), lineage (those who taught, or influenced, the performer), and context (the era in which the performer performs). It could also be used to attribute currently anonymous texts to an individual performer, determine which sections of amalgamated versions belong to which performer, and even, perhaps, assess the ability of the transcriber and method of the transcription.

To give a sense of the possible avenues for future research regarding historical variants, the following planned examples are provided. For orally-created texts, it would be possible to test the hypothesis (Hatto, 1990) that Wilhelm Radloff transcribed Manas material in 1862 from four different performers. For written texts, one could for instance check attributions of material to Musa Chaghatay uulu (Prior, 2013; 2019). It might also be possible to attribute sections of the 'harmonized' version (Yunusaliev, 1958-1960), created with the desire for a single overarching version for study, to specific performers (and revisers). One recent study (Plumtree, 2021b) compares the oral performances of a contemporary *manaschi* (AKYN, 2018) with those the performer prepared for publication (Bakchiev, 2011; 2012). Though such a study is statistical, such research will illuminate how the *manaschi* prepares the narrative for a different medium. Furthermore, tentative computerized comparison of post-performance transcriptions of contemporary performances and early sound recordings (AKYN, 2018; Prior, 2006) with those transcribed by Radloff (Hatto, 1990) indicated a significant difference in type-token ratio between texts transcribed during performance and those transcribed from recordings. This provokes the question whether it would be possible to investigate the impact of the transcribed upon the performance and, potentially, comparison of the abilities of the transcriber and the script of transcription – e. g. Valikhanov's Arabic script transcription and Radloff's adapted Cyrillic script (Hatto, 1977; 1990). Further study of these elements would add to understanding of the issue of textology (Hatto, 2000; Ready, 2015).

To give a sense of the possible avenues for future research regarding contemporary performance, attention can be directed towards the variance in the statistics of computer-calculated transcripts. Careful

control of variables within future recordings could reveal the impact of external elements on the text constructed in performance. Variables such as time constraints and audience interaction could potentially be measured. As study of popular performers might indicate which linguistic elements (if any) are appreciated by particular audiences, study of how a *manaschi* constructs his or her text in performance for audiences from different regions, different social classes, and in different contexts has the potential to reveal sociolinguistic aspects of presenting the epos.

Regarding the entire textual corpus and the *juk* question raised in the introduction, one potential approach is computational. A usable database of textual variants would allow the material to be examined in different ways, and not merely the stylistic features examined above. This does not reject the traditional means of textual scholarship, since with this approach, like the method demonstrated in the previous section, the more accurate the data the more precise the result. Rather, such research compliments the traditional methods. Digital Humanities opens up the possibility for ‘distant reading’, an established method of analysing a vast array of texts at a scale far greater than human concentration, effort, and time, can allow (Moretti, 2013; Underwood, 2017; 2019). Computer-assisted analysis of a greater range of materials than a mere selection has the potential for highlighting patterns and elements that have otherwise gone unnoticed either from lack of perspective, time, or difficulty. The work of the Stanford Literary Labs (Moretti, 2007; 2017) shows such projects are achievable. Such studies can assist small details such as cliché and its relationship to literary judgment (Cranenburgh, 2018). The ‘ever-increasing *juk*’, the corpus of Kyrgyz-language material relating to *Manas*, however would be undoubtedly unique in being one of the first non-Western languages to have such a corpus open to study, and its greater range of text available (oral, later-transcribed oral, written, printed, etc). Every known variant could be included. With contextual data (when available) provided – such as performer and his/her background, audience, date and place of performance, media in which the variant exists, main topic and narrative features – each user of the corpus was the potential to explore all the material or to narrow their focus to particular elements. Investigation could fruitfully reveal these characteristics and highlight their significance on the form and content of the narrative. Such studies, such as establishing periodization, would be of value both for Kyrgyz history, culture, and linguistics, and of comparative importance for understanding of orality, literacy, intangible heritage, and cultural memory.

Below are potential projects that can examine the detail to show the bigger picture. References indicate where the topic has been previously discussed, and scholars currently working on the material have been included.

- A lemmatical concordance of the entire corpus, created and sortable by machine, would be a resource that could be used to investigate the stability and flux of the Kyrgyz language, and examine an oral performance responding to writing and print.
- Plotting the episode of the narrative (e. g. ‘Birth of *Manas*’, ‘Great Campaign’) against the date of the performance would give an indication of the popularity of the episodes. This could establish the fashions (and focus) of each era, and subsequently assist with the periodization of *Manas* (Hatto, 1990; 1997; Prior 2002).
- Storytelling elements – speeches, flashbacks, dreams, and such – could be marked and measured to see how they correspond to the period and medium in which they were recorded. This could also reveal how some of the narratives diverged, developed, disappeared, or collated through time (Prior, 2002).
- Examining temporal features (such as use of the past tense, flashbacks, premonitions) may establish different conceptions of time. This focus may also help establish contemporary perceptions of when the events were said to take place, and how the audiences were meant to regard the narratives (e.g. didactic and practical entertainment, nostalgic and escapist, mythical story, narrative history) (Prior, 2002; Plumtree, 2021a).
- Tagging the names of the characters from the first version to the latest would allow us to determine whether the *dramatis personae* became fixed and at what moment, and see which eras (and which performers) focused on which. For instance, the greater number of narratives concerned with *Manas*’s son around the start of the twentieth century suggests a different attitude and context to epos when compared with the mid-Nineteenth century versions (Hatto, 1997; Prior, 2002). Graphing such details would similarly clarify the periodization of the storytelling tradition, and allow for greater understanding of external historical factors upon the epic.

- In a similar manner, the enemies of Manas could be charted, establishing audience attitudes towards the Kalmaks, the Chinese, and the Kazakhs (Kara, 2010). Such elements, incorporating the Manas materials since 1856, could be visually mapped to highlight, perhaps, contemporaneous concerns and disputes (and thus revealing a popular comment on historical relations.). (A work concerned with ‘orus’ – Russians – is in preparation.)
- Likewise, references to animals, both named and species, can be plotted. This may reflect a change in attitude to human-animal relations as audiences change from semi-nomadic to settled.
- Locations in the epic can be mapped. An AUCA research project, led by Aijamal Sarybaeva and Daniyar Karabaev, is currently working with the Radloff and Orozbekov variants. Examination of this data will illustrate alterations of geographic perspectives among the performer and his or her audience. The realistic itineraries of the semi-nomadic mid-nineteenth century Kyrgyz, reflecting their *mentalité* (Hatto, 1992; Prior, 1998), appear to change into the more escapist and fanciful routes of the turn of the century as a result of increased settlement and a decline in nomadism (Prior, 2002; 2009), then more ideological geographies that attempt to mimic the past in the early Soviet period (Plumtree, 2021a). The noted relationship between post-Independence ‘sacred’ geography and current political boundaries (Artman, 2016) may also be present in contemporary performances of Manas narratives. In addition to charting the expanding horizons of the world and the relationship between the real and the mythical, a larger corpus analysis would also assist identification of anomalies of individual performances (or performers) (Hatto, 1997).
- Searching for specific words could show their popularity and usage. In addition to revealing fashions, this could be applied to specific terminology. Religious phrases, and words of foreign origin (e.g. Persian, Arabic, Russian – for starting work on this see (Abdyrahmanov et al, 2018)), could be charted to see the importance of these elements for audiences, and measure their influence. Such a study could also, potentially, be used to analyse the impact of colonialism and trade, and the degree these influences impacted upon language.
- Stylistic and functional features like word patterns and epithets could be traced from variant to variant (Hatto, 1989; 1990), or even *n*-grams. This could establish which performer was familiar with which predecessor or contemporary. A stemma of the story could therefore be produced. This could also be plotted geographically, to see the route of such narratives, and to measure if and how the lineage and ‘schools’ of the oral period were disrupted by the availability of printed variants. (Alymkan Jeenbekova is currently working on dynasties and schools of Manas performers.)

Such studies could go further and potentially establish the effect upon the tradition of geographical boundaries both natural – comparing variants from the Pamirs (Dor, 1982) to versions from the territory of modern Kyrgyzstan – and political – seeing the degree of deviation and shared features when including variants from Xinjiang (Reichl, 2019; Jacquesson, 2020). These approaches allow us to see both the specifics and the span at the same time, a perspective with a clarity and thoroughness previously unobtainable. In doing this, it will become possible to tell the epic story of the epic of Manas, with greater detail in how it continued as well as changed as it went from an oral tradition transcribed by foreign scholars to a manuscript culture to a multi-volume work to the present state.

This paper has shown the value of computer-assisted analysis of Manas narratives. The statistical methods demonstrated above regarding the transcriptions of performances by Talantaaly Bakchiev and Doolot Sydykov provide an easier, and more substantiated, means to make comparisons between variants and performers. Such methods can be used to further investigate the impact of variables upon the in-performance text. Computer-assisted analysis can also be used to chart specific elements of the narratives in order to provide a larger perspective grounded on data. This is research that uses narrative texts to potentially reveal the social, linguistic, and cultural history of the Manas narratives to graph, map, and tell the ever-continuing epic of the epic.

Ethical Declaration

This study, “*Computer-Assisted Analysis of Manas Narratives: Demonstrations and Directions*”, followed the standard scientific, ethical, and citation rules. The data is presented as it is, free of tampering, and the findings

discussed here have not been submitted elsewhere. The research was supported by an AMICAL Consortium programme enabled by a grant from the Andrew W. Mellon Foundation, a fellowship for Digital Scholarship sponsored by Elsevier at the Scaliger Institute (Leiden), and an AUCA Faculty Research Grant.

Acknowledgements

Thanks are given to Talantaaly Bakchiev and Doolot Sydykov, without whom this research would not exist; to Alymkan Jeenbekova, Anguelina Popova, and Kamen Bonov in the AKYN Research project, for their contributions to this study and for checking the findings; to Kamila Baimuratova for making the audiovisual recordings; Ben Companjen and Peter Verhaar at the Centre for Digital Scholarship at Leiden University Library for sharing the methodology, fruitful conversations, and confirming what is possible; to Marie Hakenberg and Srdjan Uljevic for covering my classes at AUCA during the Leiden fellowship; to Walid Azizi for discussing Arabic and Persian vocabulary in Manas variants; to Elena Popa and Kelly Hydrick for comments on earlier drafts, and to the anonymous reviewers for their constructive comments.

References

- Abdyrahmanov, T., Karataeva, S., & Dyushembiev, S. (2018). *"Manas" Eposundagy özdöshtüpülgön sözdör*. Bishkek: kitep.kg.
- AKYN Research Group. (2018). 'AKYN Project'. Access address: <http://akynproject.auca.kg/en/>.
- Almásy, G. v. (1911-12). Der Abschied des Helden Manas von seinem Sohne Sémety: aus dem karakirgisischen Epos "Manasın kısası". *Keleti Szemle*, 12, 216-223.
- Artman, V. M. (2016). *The State and the Sacred: Memory, Theology, and Identity in Kyrgyzstan* (Doctoral Thesis). University of Kansas, Kansas.
- Bakchiev, T. (2011). *Manastyn Ashy*. Bishkek: Turar.
- Bakchiev, T. (2012). *Almambettin Jomogu*. Bishkek: Kut-ber.
- Botoyarov, K. (1992). *Köönörgüs muras. "Manastyn" "Kökötöydün ashy" epizodunun Ch. Valibanov jazdyryp algan varianty*. Bishkek: Aybek.
- Bruley, J. (2019). *L'épopée de Manas; étude historique, patrimoniale et ethnographique* (Doctoral Thesis). Université de Lille, Lille.
- Cranenburgh, A. v. (2018). Cliché Expressions in Literary and Genre Novels. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLJL-2018)* (pp. 34-43). Sante Fe: Association for Computational Linguistics.
- Dor, R. (1982). Un fragment pamirien de Manas. *Central Asiatic Journal*, 26, 1-55.
- Duishembieva, J. (2015). *Visions of Community: Literary Culture and Social Change among the Northern Kyrgyz, 1865-1924* (Doctoral Thesis). University of Washington, Washington.
- Hatto, A. T. (Eds.). (1977). *The Memorial Feast for Kökötöy-Khan (Kökötöydün aşı): A Kirghiz Epic Poem*. Oxford: Oxford University Press.
- Hatto, A. T. (1989). Epithets in Kirghiz Epic Poetry 1856-1869. In J. B. Hainsworth (Eds.), *Traditions of Heroic and Epic Poetry*, vol. ii, *Characteristics and Techniques* (pp. 71-93). London: Modern Humanities Research Association.
- Hatto, A. T. (Eds.). (1990). *The Manas of Wilhelm Radloff*. Wiesbaden: Otto Harrassowitz.
- Hatto, A. T. (1992). Die Marschrouten in der älteren kirgisischen Heldenepik. *Fragen*, 5, 331-341.
- Hatto, A. T. (1997). Tradition and Change in the Kirghiz Manas. In S. Akiner & N. Sims-Williams (Eds.), *Languages and Scripts of Central Asia* (pp. 99-105). London: School of Oriental and African Studies.
- Hatto, A. T. (2000). Textology and epic texts from Siberia and beyond. In L. Honko (Eds.), *Textualization of Oral Epics* (pp. 129-160). Berlin/New York: de Gruyter. doi: 10.1515/9783110825848.129.
- Heide, N. v. d. (2015). *Spirited Performance: The Manas Epic and Society in Kyrgyzstan*. Bremen: EHV Academicpress GmbH.
- Jacquesson, S. (2020). Claiming heritage: the Manas epic between China and Kyrgyzstan. *Central Asian Survey*. doi: 10.1080/02634937.2020.1765739.
- Jumaturdu, A. (2016). A Comparative Study of Performers of the Manas Epic. *Journal of American Folklore*, 129, 288-296. doi: 10.5406/jamerfolk.129.513.0288.
- Kara, D. S. (2010). Kalmak: The Enemy in Kazak and Kirghiz Epic Songs. *Acta Orientalia Academiae Scientiarum Hungaricae*, 63(2), 167-178. doi: 10.1556/AOrient.63.2010.2.3.
- Karalaev, S. (1984-1991). *Manas, Semety, Seytek*. Frunze: Kyrgyzstan.
- Kydyrbaeva, R. Z. (1984). *Skazitel'skoe masterstvo manaschi*. Frunze: Ilim.
- Liu, H. (2014). 57 wan han "manasi" shou chao ben bei fa xian yuan ben yi bei shao hui?. *ChinaNews* July 7 2014. Access address: <http://www.chinanews.com/cul/2014/07-07/6357127.shtml/>.
- Lord, A. B. (2003). *The Singer of Tales*. Second edition: S. Mitchell and G. Nagy (Eds.). Cambridge, MA/London: Harvard University Press.
- Mamay, J. (1984-1994). *Manas*. 18 vols. Urumqi: People's Publishing House.
- Moretti, F. (2007). *Graphs, Maps, Trees: Abstract Models for Literary History*. London/New York: Verso.
- Moretti, F. (2013). *Distant Reading*. London: Verso.

- Moretti, F. (Eds.). (2017). *Canon/Archive: Studies in Quantitative Formalism from the Stanford Literary Lab*. New York: n+1 Foundation.
- Musaev, S. (1995). The Epos “Manas”. In S. Aliev, R. Sarypbekov, & K. Matiev (Eds.), *Encyclopaedical Phenomenon of Epos “Manas”* (pp. 276-294). Bishkek: Glavnaya redakciya Kyrgyzskoy Enciklopedii.
- National Academy of Sciences. (2016). National Academy of Sciences Online Manuscript Catalogue. Access address: <https://manuscript.lib.kg/>.
- Orozbekov, S. (1978-1982). *Manas*. Frunze: Kyrgyzstan.
- Orozbekov, S. (2010). *Manas: Kyrgyz elinin baatyrdyk eposu*. S. Musaev & A. Akmataliev (Eds). Bishkek: Han-Tengir.
- Öztürk, G. (2015). Müşterek Doğu Türkçesi İmlasıyla Yazılmış Arap Harfli Metinlerin Çağdaş Lehçelere Uyarlanarak Okunması Meselesi ve Kırgız Akin Maldıbay Borzu uulu nun Kısxsax i Semetey Bu Turur Örneği. *Marmara Türkîyat Araştırmaları Dergisi*, 11, 61-71.
- Parry, A. (Eds.). (1971). *The Making of Homeric Verse: The Collected Papers of Milman Parry*. Oxford: Clarendon Press.
- Plumtree, J. (2019). A Kyrgyz Singer Of Tales: Formulas in Three Performances of the *Birth of Manas* by Talantaaly Bakchiev. *Doklady Nacional'noy akademii nauk Kyrgyzskoy Respubliki*, 1, 125-133.
- Plumtree, J. (2021a). A Telling Tradition: Preliminary Comments on the Epic of Manas, 1856- 2018. In S. C. Thomson (Eds.), *Medieval Stories and Storytelling: Multimedia and Multi-Temporal Perspectives* (pp. 239-301). Turnhout: Brepols. doi: 10.1484/M.MNT-EB.5.121610.
- Plumtree, J. (2021b). A Contemporary Manaschi in Oral Performance and in Print. *Alatoo Academic Studies*, 1, 238-244. doi: 10.17015/aas.2021.211.29.
- Plumtree, J. & Jeenbekova, A. (2020). Replicating the *Birth of Manas*: Composition Methods of Doolot Sydykov and Talantaaly Bakchiev. *Izvestiya Nacional'noy akademii nauk Kyrgyzskoy Respubliki*, 3, 175-187.
- Prior, D. (1998). Bok-murun’s Itinerary Ridden: Report on an Expedition Through Kirghiz Epic Geography. *Central Asiatic Journal*, 42, 238-282.
- Prior, D. (2000). *Patron, Party, Patrimony: Notes on the Cultural History of the Kirghiz Epic Tradition*. Bloomington: Indiana University Research Institute for Inner Asia Studies.
- Prior, D. (2002). *The Twilight Age of the Kirghiz Epic Tradition* (Doctoral Thesis). Indiana University, Department of Central Eurasian Studies, Indiana.
- Prior, D. (Eds.). (2006). *The Semetey of Kenje Kara: A Kirghiz Epic Performance on Phonograph with a Musical Score and a Compact Disc of the Phonograph*. With the assistance of Ishembi Obolbekov. Wiesbaden: Harrassowitz Verlag.
- Prior, D. (2009). Travels of Mount Qāf: From Legend to 42° 0' to N 79° E. *Oriente Moderno*, 89, 425-444. doi: 10.1163/22138617-08902015.
- Prior, D. (Eds.). (2013). *The Šabdan Baatır Codex: Epic and the Writing of Northern Kirghiz History: Edition, Translation and Interpretations, with a Facsimile of the Unique Manuscript*. Leiden/Boston: Brill.
- Prior, D. (2019). A Qırghız verse narrative of rebellion and exile by Musa Chaghatay uulu. In A. Chokobaeva, C. Drieu, & A. Morrison (Eds.), *The Central Asian Revolt of 1916: A Collapsing Empire in the Age of War and Revolution* (pp. 308-326). Manchester: Manchester University Press. doi: 10.7765/9781526129437.00022.
- [Radloff, W]. Radlovym, V. V. (1885). *Obraztsy narodnoj literatury severnykh tjurkskich plemen*, vol. 5, *narechie dikokamennykh Kirgizov*. St Petersburg: Commissionäre der Kaiserlichen Akademie der Wissenschaften.
- Ready, J. L. The Textualization of the Homeric Epic by Means of Dictation. *Transactions of the American Philological Association*, 145, 1-75. doi: 10.1353/apa.2015.0005.
- Reichl, K. (1992). *Turkic Oral Epic Poetry*. New York and London: Garland, 1992.
- Reichl, K. (2016). Oral Epics into the Twenty-First Century: The Case of the Kyrgyz Epic *Manas*. *Journal of American Folklore*, 129, 327-344. doi: 10.5406/jamerfolk.129.513.0327.
- Reichl, K. (2019). Oral Epics Along the Silk Road: The Turkic Traditions of Xinjiang. *CHINOPERL: Journal of Chinese Oral and Performing Literature*, 38, 45-63. doi: 10.1080/01937774.2019.1633161.
- Sarypbekov, R. (Eds.). (1994). *Manas: Kökötöydün ashy. Tynybek: Semetey baatyrdan bir bölüm epos*. Bishkek: Ala-Too.
- Sinclair, S. & Rockwell, G. (2016a). *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Cambridge, MA / London: MIT Press.
- Sinclair, S. & Rockwell, G. (2016b). Voyant Tools. Access address: <http://voyant-tools.org/>.
- Underwood, T. (2017). A Genealogy of Distant Reading. *Digital Humanities Quarterly*, 11. 2. Access address: <http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html>.
- Underwood, T. (2019). *Distant Horizons: Digital Evidence and Literary Change*. Chicago and London: University of Chicago Press.
- Yunushaliev, B. M. (Eds.). (1958-1960). *Manas, Semetey, Seytek*. Frunze: Kyrgyzmambas.

TÜRKÇE GENİŐ ÖZET

Makale, vezinli Manas anlatılarının bilgisayar destekli analizi üzerine olup iki bölümden oluřmaktadır. Çağdař Manas icralarını arařtıran pratik bir örneęi sergileyen ilk bölüm, Kırgız sözlü destan Őiirinin bilgisayar destekli analizinin yapılabilirlięini göstermektedir. İkinci bölüm, bu alanda ne gibi potansiyel arařtırmaların gerçekleştirilebileceęinin bir listesini sunmaktadır.

İlk bölüm, Kırgız dili metinlerinin bilgisayar destekli analizinin yazarın/kaynaęın belirlenmesine yardımcı olmak için kullanılabilirlięini göstermektedir. Bu, Manas destanı için yöntemin bilinen ilk kullanımıdır. 2017-2018 yıllarında iki Manasçının Manas icralarının görsel-iřitsel kayıtları yapılmıřtır. Her bir icra, anlatının aynı bölümünün, yani Manas'ın doęumunun, icra edildięi üç ayrı seferde kaydedildi. Bu Manas icralarına tanıklık ederken, icra edenler ve icralar arasındaki linguistik farklılıkları ölçmenin mümkün olup olmadıęı fikri ortaya çıktı. Bu fikri tatbik etmek üzere, Talantaaly Bakchiev'in icrası (sırasıyla **T₁**, **T₂** ve **T₃** olarak kodlanmıřtır) ve Doolot Sydykov'un icrasından (sırasıyla **D₁**, **D₂** ve **D₃** olarak kodlanmıřtır) transkripsiyonlar yapılmıřtır. Açık eriřimli çevrim içi yazılım Voyant Tools kullanılarak, hem tek bir icracının farklı icraları arasında hem de iki farklı icracı arasında bilinen ilk nicel linguistik karřılařtırma yapılmıřtır. (Bu itibarla, Orozbekov ve Karaliev gibi icracıların temel özelliklerini tanımlamak ve karřılařtırmalı çalıřmalar için de bu yöntemin kullanılabilirlięini varsayabiliriz). İstatistiksel veriler sunan bu yöntem, bir icracıyı ve icrayı deęerlendirirken kiřisel önyargılara veya varsayımlara dayanmaktan daha bilimsel bir deęerlendirme zemini saęlar.

Birinci bölüm, analizin nasıl daha kesin hale getirilebileceęine dair bir kanıtlama ile devam etmektedir. İcranın uzunluęundaki deęiřimin analiz sonucunu etkileme ihtimalinin farkında olarak, bu çalıřma, TTR (type-token ratio) analizinin (metni sabit uzunluklu parçalara ayıran, orijinal kelimelerin sayısını toplam kelime sayısı ile karřılařtıran analiz tipinin) bu olası sorunu nasıl çözebileceęini göstermektedir. Bu karřılařtırma, Sydykov'un Bakchiev'den daha yüksek bir kelime yoęunluęuna (yani, sabit uzunluklu fragman başına daha benzersiz kelimelere) sahip olduęunu ortaya koydu. Bu veriler, bir metnin "kaynaęını/yazarını" belirlemek için kullanılabilir; zira istatistiklerin gösterdięi gibi her sanatçının bir "imza" stili vardır. Transkripsiyonları Jupyter notebook ile çalıřtırılan bir Python programına girdiğimizde (makalede de görüleceęi üzere) bir grafik oluřturur. Bu grafiklerde iki ayrı grup hat görülebilir. Bu iki grup iki icracıya karřılık gelir. Buradan çıkarılabilecek sonuç, bu analiz yönteminin bireysel icracıların ve icraların özelliklerini belirlemek için kullanılabilir olduęuyla sınırlı deęildir. Bir hipotez olarak řu söylenebilir ki, bu yöntem bu tür metinlerin 'kaynaęını/yazarını' belirlemek (ve keza anonim metinleri tanımlamak) için de kullanılabilir. Daha fazla görsel-iřitsel kayıt yapılırsa, bu yöntem, hassas bir şekilde uygulanmak kaydıyla, dıř etkilerin (izleyici veya zaman kısıtlamaları gibi) bir icracının icra sırasındaki doęaçlaması üzerinde hangi etkilerde bulunabileceęini deęerlendirmek için de kullanılabilir.

İkinci bölüm, Manas'la ilgili gittikçe büyüyen literatürü anlamak ve incelemek üzere çeřitli yollar önermektedir. Karřılařtırırken, 'juk' (bir yurttaki battaniye yıęını) yapılı; yani incelenen ya da en fazla kıymet biçilenler sadece en yeniler ve dolayısıyla en üstte olanlardır. Oysaki böyle bir odaklanma dięer materyali görünmez kılar. Bu sorundan kaçınmak için, potansiyel yeni yaklařımları mümkün kılan biliřimsel yöntemler önerilmektedir. Kullanılan kelimeler için bir veri tabanı oluřturmak ve belirli öęeleri (oluřturulma tarihi, tür, hikâye anlatımı özellikleri, zamansal öęeler, karakterler, düşmanlar, konular, lakaplar) imlemek, anlatıların deęerlendiriliřini, dönemlendiriliřini ve geliřimini haritalamaya yardımcı olurken aynı zamanda deęiřikliklerin ve sürekliliklerin nedenlerini anlama yönelik yeni arařtırmaları da mümkün kılar. Literatürü bu şekilde incelemek; Manas destanının linguistik ve toplumsal bağlamına yeni içgörüler saęlayacaktır, artan okuryazarlıęın ve sömürgecilik gibi faktörlerin içerik üzerindeki etkisini ortaya çıkarabilir ve hatta modern Kırgızistan dıřındaki destan varyantları da analize dâhil edildiğinde mevcut sınırların destan üzerindeki etkisi incelenebilir olur. Gelecekteki çalıřmalar için önerilen bu alanlar, önümüzdeki on yıl için arařtırma gündemini belirleme potansiyeline sahiptir.