

A Choice of Relationship-Revealing Variants for a Cladistic Analysis of Old Norse texts: Some Methodological Considerations

Katarzyna Anna Kapitan

Department of Nordic Studies and Linguistics, University of Copenhagen

kak@hum.ku.dk

Keywords: stemmatology, manuscript studies, manuscripts, cladistic analysis, Old Norse

Abstract

The research presented in this article centers on the methodological field of computer-assisted stemmatics, specifically the application of the tools and methods originating from phylogenetics to answer questions of textual criticism. Given the well-known problem of stemmatics that the shape of the stemma changes depending on the readings selected for analysis, surprisingly little discussion has been devoted to the definition of a relationship-revealing reading among the practitioners of computer-assisted methods. This article discusses some of the controversies regarding the methodological principles within the field of computer-assisted stemmatics (new stemmatics and cladistic textual criticism), with the main focus on a choice of relationship-revealing readings. This research takes an experimental approach towards different methodological principles, and tests them using PHYLIP (the Phylogeny Inference Package, version 3.695). The experiments aim to assess the influence of different types of variants on the results of a cladistic analysis. The experiments are based on readings collected from the oldest part of the manuscript tradition of an Icelandic saga, *Hrómundar saga Gripssonar*. The results of the experiments suggest that the cladistic method can be employed in traditional textual research, but the results achieved through this process are highly dependent on the type of variation included in the input file: the shape of the unrooted tree of relationships changes depending whether it was built on major or on minor variants.

Introduction

The application of computer-assisted methods, originating from phylogenetics, to answer questions of textual criticism has been recognized in academic discourse as a powerful tool for revealing the filiation of manuscripts (recently in Old Norse studies: Hall & Parsons, 2013; Zeevaert et al., 2013). Surprisingly, not much discussion within the field of "New Stemmatics" has been devoted to the definition of a relationship-revealing reading, and there is disagreement among practitioners of computer assisted-methods regarding the fundamental question: What type of textual variation can, or should, be used for text-genealogical analysis? Salemans (1996) suggested a strictly systematized classification of text-genealogically informative variants, while Robinson (1996, p. 75) opted for basing the analysis on all substantive variants of different readings. This paper takes an experimental approach towards this problem, and presents the results of applying a phylogenetic analysis to the oldest part of the tradition of an Icelandic saga, *Hrómundar saga Gripssonar* (HsG). The discussion is based on experiments conducted employing a package of programs for inferring phylogenies, PHYLIP, developed by Felsenstein (version 3.695, 2013). The aim of the experiments was to assess the influence of different types of variants on the results of cladistic analysis.

Theoretical framework

Close similarities between the theoretical assumptions of cladistics and stemmatics have been noted by Platnick and Cameron (1977, pp. 384–385), who pointed out "that cladistic analysis is a general comparative method applicable to all studies of historical interrelationships based on actual ancestor-descendant sequences," including stemmatics and historical linguistics. The same idea has been recently revived by Howe et al. (2004), who discussed the main similarities between evolutionary biology and stemmatics, emphasizing the similarity of the challenges both disciplines face, namely contamination in manuscript traditions and the evolutionary processes occurring in DNA-sequences, such as recombination, transposition, and homoplasy (cf. O'Hara & Robinson, 1993; Robinson & O'Hara, 1996).

In recent years, the use of a cladistic analysis in stemmatological research has become highly popular, and has led to the appearance of two terms, "cladistic textual criticism," and "New Stemmatics." Even though both terms refer to computer-assisted stemmatics based on cladistic analysis, the principles governing the data collection are different. "New Stemmatics," according to the definition published on *The Textual Scholarship* webpage (Robinson & Bordalejo, 2010b), aims at obtaining, as far as possible, a comprehensive overview of the relationships between witnesses through an analysis based on all available data using quantitative tools, typically computer-assisted analysis. In contrast, "cladistic textual criticism," a term introduced by Salemans (1987), argues for a careful selection of variants to be processed by a computer, and

emphasizes that only very few textual differences can be considered genealogically informative, and hence used for the analysis.

In my view, it is justifiable to state that all philologists, regardless of their background, would agree that not all modifications of a text (innovations) can be considered genealogically informative. There is, however, no consensus regarding which variants (readings, errors) can be considered as such, and can be used to build the stemma (chain, tree) of the manuscript tradition. The well-known approach, practiced by neo-Lachmannians, including to some extent also Salemans, follows the rule that exclusively “non-polygenetic significant errors” can be considered text-genealogically informative, and can thus be used to build a stemma (cf. Trovato 2014, p. 55, p. 110). In this case, the judgment of a polygenetic reading remains in the philologist's individual domain, and gives philologists an opportunity to make their critical judgment based on their own preferences and experiences. This subjectivity, as suggested recently by Andrews (2016), might pose some challenges to text-critical research, and bring under scrutiny the quality of the decisions made by philologists.

Conversely, the dominant view within "New Stemmatics" is that orthography, punctuation, and formal presentation are types of variation that do not carry genealogy-revealing information (cf. Bordalejo 2015, p. 566; Robinson 2016, p. 639); accordingly, only substantive variants can be used to build a stemma. This suggestion follows Greg's (1950) distinction between substantives, which are likely to have been copied from witness to witness, and accidentals, which might be particular to a scribe (cf. Robinson 1996, p. 75). However, Greg's considerations were more of an editorial than a stemmatic nature and, even though he observed that "the distribution of substantive variants generally agrees with the genetic relation of the texts" (Greg 1950, p. 22), he did not comment on how accurate this "general agreement" is, or how useful substantives are to build a stemma. On the contrary, in the earlier stages of his career, Greg was a zealous aficionado of building stemmas using exclusively type-two variants, which he defined as genealogically significant variants that divide a given tradition into exactly two groups, where each variant occurs in at least two text versions (Greg, 1927, pp. 22–23). Due to the lack of more detailed discussion within “New Stemmatics” to define genealogically informative readings, the most reasonable way to reveal some of the principles applied in computer-assisted analysis is to consult lists of variants used in previous scholarship. The complete lists of variants of *The Wife of Bath's Prologue*, *The General Prologue*, *The Miller's Tale*, *The Nun's Priest's Tale*, and *Sólarljóð* are all available online on *The Textual Scholarship* webpage (Robinson & Bordalejo, 2010a); considering the scope of this paper, I have decided to consult Robinson's (2004) list of variants for Old Norse-Icelandic poem *Sólarljóð*.

The readings listed for *Sólarljóð*, and therefore those considered genealogically informative, include, for example, the sentence: *helgir englar komu or himni ofan ok tóku sál hans* "holy angels came from heaven above and took his soul" (nos. 117-125). In this sentence, variation appears in a verb tense (*komu* - *koma*, no. 119), omission or use of a different preposition (*or* - *af* - *frá*, no. 120), the number of a noun (*himni* - *himnum*, no. 121), and inversion of a determiner and head noun (*sál hans* - *hans sál*, no. 125). Other examples include omission and variation in conjunctions (*því* - *því að*, no. 797), prepositions (*fyrir* - *frá*, no. 847), adverbs (*brot* - *í burt* - *burt*, no. 853), and word order (*menn ég sá þá* - *menn sá ég þar*, nos. 1129-1131), as well as obvious errors (*fuglar* [birds] as "fulgar" [*sic*], no. 991). Even though many of the listed variants can be considered minor from a traditional point of view, it has been suggested by Hall and Parsons (2013, § 39) that "an accumulation of minor variants all pointing in the same direction can become a powerful argument for a particular manuscript filiation" and as Robinson (2016, p. 649) emphasized, the results achieved by a phylogenetic analysis are reliable "because our analysis does not rest on only these one or two variants ('indicative' as they might be), but on patterns within the whole mass of variation."

Research questions

Given, firstly, the dichotomy in approaching genealogically informative variants, and secondly, the well-known problem of stemmatology that the shape of the stemma changes depending on the readings selected for the analysis, there is a need to evaluate the effectiveness of both approaches in the context of computer-assisted stemmatology. Obvious questions arise, such as how to select readings that carry relationship-revealing information, and how to use them in computer-assisted analysis. Should we base stemmas on neo-Lachmannian non-polygenetic significant errors, and build the input file exclusively on *loci critici*, or should we register all types of substantives, as new stemmaticists suggest? To my knowledge there has been no research evaluating how these two categories influence the results of computer-assisted analysis. Instead, scholars have been creating computer-assisted stemmas and comparing them with existing, traditionally-built stemmas, in order to evaluate how accurate computer-generated stemmas can be. It is hoped that the innovative approach of the experiments presented in this paper will shed new light on this problem.

Methodology

Choice of manuscripts

The experiments presented in this paper are a case study of the oldest witnesses of HsG, listed in Table 1. These are the manuscripts on which previous scholars, mainly Kölbing (1876) and Andrews (1911), but also to some extent Rafn (1829, pp. xii–xiii), based their arguments regarding the saga's origin and transmission, but arriving at contradictory conclusions. An

exception is B4859, which was dismissed by both Andrews and Kölbing as worthless, and L222, which was most likely unknown to them: both manuscripts are included in my analysis based on chronological criteria.

Table 1: Oldest manuscripts of HsG, by shelf mark.

Siglum	Shelf mark	Date	Scribe
A193	AM 193 e fol.	1690-1697	Ásgeir Jónsson
A345	AM 345 4to	1695	Jón Þórðarson
A587	AM 587 b 4to	1686-1688	Ásgeir Jónsson
A601	AM 601 b 4to	1650-1689	Jón Eggertsson
B4859	BL Add. 4859	1695	Jón Þórðarson
L222	Lbs 222 fol.	1695	Jón Þórðarson
P67	Papp. Fol. nr 67	1687	Jón Eggertsson
T1768	Thott 1768 4to	1686-1688	Ásgeir Jónsson

This choice of witnesses serves the purpose of placing the results of the computer-assisted analysis in the context of the existing classifications of HsG manuscripts, by Kölbing and Andrews, who based their judgment on traditional genealogical methods. Kölbing (1876, p. 181) excluded the possibility that any of the witnesses examined could be the *codex optimus* of the existing tradition, or even an ancestor of the other manuscripts examined, as presented in Figure 1. Andrews (1911, p. 531) rejected the idea that all the manuscripts represent independent branches; instead he suggested that A601 preserves an original text of the saga, and is an ancestor of the entire tradition, as presented in Figure 2. Andrews did not discuss the relationships between other manuscripts, but his stemma seemed to group other witnesses into two branches: on one side he placed manuscripts in Ásgeir Jónsson's hand, A587, A193, and a third unnamed node, which certainly represents T1768, to which he refers in the text; on the other side he placed P67 in Jón Eggertsson's hand, and A345 in Jón Þórðarson's hand.

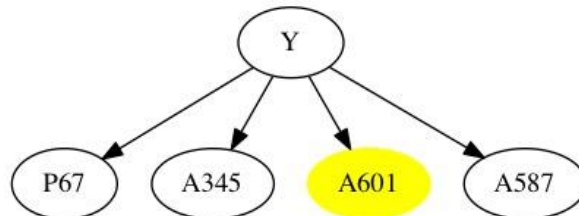


Figure 1: Stemma based on Kölbing 1876, p. 182.

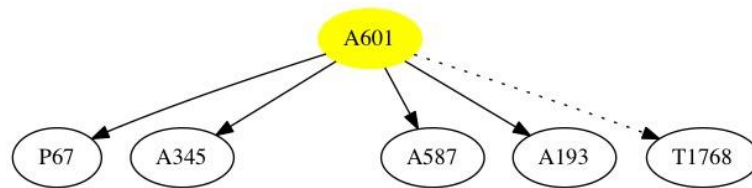


Figure 2: Stemma based on Andrews 1911, p. 531. The dotted line represents a connection between A601 and a manuscript that was missing a label in the original stemma.

Andrews's stemma is certainly incorrect regarding the relationships between the manuscripts in Ásgeir Jónsson's hand (A193, A587, T1768). A193 and A587 are almost identical and must be closely related, and A193 is most likely a descendant of A587. Kölbing's stemma includes only four manuscripts and might be a good hypothesis for the relationships between these manuscripts, but whether A601 and P67 are independent witnesses is not obvious. Since there is insufficient space here for a detailed discussion of relationships between the manuscripts based on extra-textual evidence, neither to present this author's own stemma, nor make an argument in favor of it, this paper focuses on comparing the trees of relationships based on different criteria. Additionally, the computer-assisted stemmas are put into the context of the two traditionally achieved stemmas presented above, testing whether any of them can be obtained by computer-assisted analysis based on different data sets.

Cladistic analysis

The experiments presented in this article were inspired by Hall's (2013) experiments with the phylogenetic analysis conducted on small samples from *Konráðs saga keisarasonar*, as well as experiments conducted on chapter 86 of *Brennu-Njáls saga* — during the stemmatology workshop organized by Hall and Zeevaert at the Arnarnagnæan Summer School in Manuscript Studies (Zeevaert et al., 2013). Following Hall's example, instead of conducting the experiments with the help of PAUP* (Phylogenetic Analysis Using Parsimony), which currently seems to be the most popular software in the field of stemmatics, I have chosen a free open-source package of programs for inferring phylogenies, PHYLIP, specifically the general parsimony program PARS and tree-plotting program DRAWTREE and CONSENSE (Felsenstein, 2013).

Complete transcriptions of the texts were prepared in plain text format, and collated in a spreadsheet. The texts were transcribed on a simplified diplomatic level, with abbreviations expanded in round brackets (), unclear readings marked in square brackets [], deletions likely to be by the scribe between slashes //, and scribal additions in insertion marks ˇ. No variation of graphemes or orthography has been represented and, for the sake of practicality, modern Icelandic letterforms have been employed (but not modern Icelandic orthography). Variation in the transcriptions posed no difficulty for the research because PARS requires manual encoding of the variants: this means that it is the scholar's subjective decision which readings will be

considered genealogically informative variants and which not. In practice, within the spreadsheet the numeric values are assigned manually to each character and then converted into the PARS input files: therefore, normalization of the transcriptions is not crucial. Where it was impossible to determine which variant appears in a particular witness, for example due to its illegibility, the character was encoded with a question mark to indicate uncertainty.

The complete transcriptions comprised approximately 3600 words per witness (e.g. 3587 words in A587, and 3643 words in L222) and were collated in 690 characters (columns), which correspond to places of possible variation, averaging 5.2 words per character. This number may give rise to some controversies, since scholars usually tend to have characters as small as possible: Salemans (1996, p. 15), for example, preferred a collation where each word is considered an independent place of variation. However, the collation should imitate the manuscript's copying process. It is rather unlikely that any professional scribe would copy any given text word by word, especially a text in the vernacular: copying phrase by phrase seems to be more probable. In the situation where more than one place of variation appears within one character, this character is divided into as many smaller characters as necessary; this excludes possibility that some changes arose independently from each other on different stages of the saga's transmission and accidentally ended up in one character. For example, a reading *Hrómundur spyr hver nú vill ganga í hauginn* "Hrómundur asks who now wants to go into the mound", which was first intended as one character, was divided into three characters based on the registered variation (e.g. *hver - hvor, nú vill - vill, ganga í - ganga inn í*). The characters, which do not contain any places of actual variation, remained in their initial form and can be fairly long, reaching up to eight words, such as *Setti hann þá klær sínar á hnakka Hrómundi* "He put then his claws on Hrómundur's neck".

Choice of relationship-revealing variants

There is no golden rule for revealing significant errors regardless of genre, language, and scribal tradition. Therefore, as Trovato (2014, p. 115) suggested, philologists must distinguish significant errors from noise in each individual case. Van Mulken (1993, p. 25) drew the same conclusion, stating that "it is the corpus which should dictate a typology of variants with respect to the kinship-revealing character." The typology she used to distinguish and assess the use of a number of variant types in the *Perceval* tradition is vital for the development of stemmatic methodology. Van Mulken (1993, p. 36-40) distinguished variants with low and high relationship-revealing power. The variants with low relationship-revealing power include interpolation or omission of verses, variants affecting the possessive or determinative quality of a word, numbers, changes in a narrative point of view, as well as extra-textual variants (e.g. historiated

initials and lombards). The variants with high relationship-revealing power include interchange of verses, or of rhyming constituents, changes in word order, variants concerning the aspect, time or mood of verbs, and important semantic changes. Salemans (1996, p. 4) also developed a typology of relationship-revealing variants and established rules of text-critical research; he suggested that "only very few textual differences can serve as genealogical, relationship-revealing elements." He defined four basic text-genealogical rules, which included a definition of a genealogical variant, place of variation, type-two variants, and the necessity of presenting variants in the apparatus. Within the first rule, Salemans discussed the types of variants that cannot be used for stemmatic analysis (essentially polygenetic variants): these include synonymous, regional, inflectional and historical parallelism.

The present paper draws extensively on the implications of Salemans's (1996) and van Mulken's (1993) methodology, which served as the basis for developing my preliminary typology of genealogically informative variants in the HsG early tradition. It must be emphasized, by contrast, that both aforementioned classifications were developed for poetic texts, not for prose such as HsG. Taking into consideration the stylistic differences which might influence the copying process, the criteria applied to the saga's tradition therefore required some adjustments. For the purpose of this paper I divided all the variants appearing in the oldest witnesses of the saga into two main categories: major and minor. Major variants are readings with a high likelihood of having been copied from the exemplar, which are mainly lexical, such as:

- omission, addition, or replacement of nouns, e.g. *grátt og sítt hár, skegg* "grey and long hair, beard" - *grátt og sítt skegg* "grey and long beard", and *stokkar* "timber stakes" - *steinar* "stones", *herskip* "a warship" - *skip* "a ship";
- omission, addition, or replacement of verbs, e.g. *batt* "bound" - *bar* "carried", including synonyms, e.g. *datt* - *féll* "fell", but excluding verbs introducing speech;
- omission, addition or replacement of adjectives, including synonyms, e.g. *myrkt* - *dimt* "dark";
- omission, addition, or replacement of entire phrases, e.g. *heldur enn ræna kotkarla* "rather than rob a cottager" - omitted;
- clear errors, e.g. *kü* for *ský* "sky", and *Svylöð* for *Gunnlöð* (personal name).

Minor variants are readings with a high probability of occurring independently from the exemplar, and might be scribe-individual, and thus often polygenetic, such as:

- number of nouns, e.g. *búkum* (dat. pl.) "human corpses" - *búki* (dat. sg.);
- tense of verbs, e.g. *komu* (imperf.) "they came" - *koma* (pres.);
- definiteness of nouns and adjectives, e.g. *líf* "a life" - *lífið* "the life";
- position, addition, omission, or replacement of prepositions, adverbs, and conjunctions, e.g. *og spyr eftir* "and asks after" - *og spyr* "and asks";

- linguistic variation, e.g. *hvor* - *hver* "who", *gera* - *gjöra* "to do";
- word order, e.g. *sá hét Greipur* "this [man] was called Greipur" - *er Greipur hét* "who was Greipur called", and *maður frægur* "famous man" - *frægur maður* "man famous"
- numbers, e.g. *xví undir* "sixteen wounds" - *xíiv undir* "fourteen wounds".

Additionally, I use Greg's definition of type-two variants, as introduced above, and build on this concept by including in one experiment quasi-type-two variants, which are minor variants that are hypothetically not genealogically significant, and which divide the tradition into two groups of readings with at least two witnesses representing each group.

For each experiment, different types of variation were used to build unrooted trees of relationships; therefore the number of parsimony-informative characters changes from one experiment to another. The criteria for selecting the variants are described in detail in the next section. The environment in which experiments were conducted, such as the number of characters analyzed, order of manuscripts in the input file and the settings in PARS, DRAWTREE, and CONSENSE, remained unchanged.

Experiments and results

Experiment 1

The input file for the first experiment was based on all types of variants, both major and minor, thus 255 characters were considered parsimony-informative: around 37% of all characters in the data set. The vast majority of variants included in this experiment were minor variants. Due to limitations of space, the full list of 255 variants is not included in this paper but the full data set containing collated transcriptions and PARS input files can be obtained from the author on request.

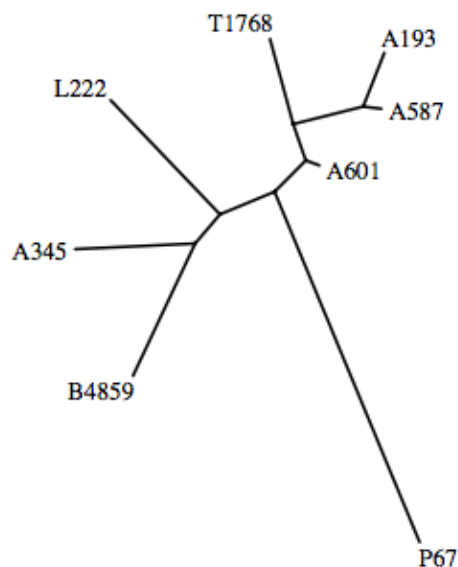


Figure 3: An unrooted tree of relationships: Experiment 1 (all variants included).

In the result of the PARS analysis, one most-parsimonious tree was found, which can be presented in Newick notation with branch lengths removed and spaces introduced for the reader's convenience: (((B4859, A345), L222), ((T1768, (A193, A587)), A601), P67). The results were plotted into the program DRAWTREE in order to visualize an unrooted tree of relationships, as presented in Figure 3. The unrooted tree presents the relationships between the manuscripts, including the branch lengths. The manuscripts in Jón Þórðarson's hand (A345, B4859, L222) are placed relatively close to each other, while the manuscripts in Ásgeir Jónsson's hand (A193, A587, T1768) are grouped together with A601. The relationship between Jón Eggertsson's manuscripts (P67 and A601) was not recognized in such a straightforward way.

Experiment 2

In the second experiment, only 73 potentially significant variants or major variants were considered parsimony-informative: around 11% of all the characters analyzed. In this experiment, three possible trees of relationships were found, which can be represented as follows:

- a. ((A345, (B4859, L222)), (T1768, (A193, A587)), A601, P67);
- b. ((A345, (B4859, L222)), (T1768, ((A193, A587), A601)), P67);
- c. ((A345, (B4859, L222)), ((T1768, (A193, A587)), A601), P67);

All three trees have the same groupings of A193 and A587 (green) and of Jón Þórðarson's manuscripts (red). Given the PARS results – three possible trees of relationships – the PARS output file was plotted first to the CONSENSE program, in order to obtain a consensus tree, applying a strict consensus method. Next, the results of CONSENSE were plotted to DRAWTREE to visualize the results, as presented in Figure 4.

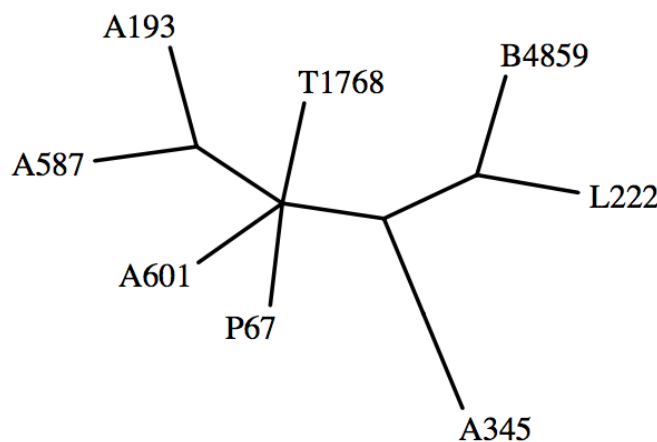


Figure 4: A strict consensus tree: Experiment 2 (note that branch lengths in the consensus tree are not relevant)

As shown in Figure 4, Jón Þórðarson's manuscripts (A345, B4859, L222) are derived from a common ancestor, while the two manuscripts in Ásgeir Jónsson's hand (A193 and A587) are presented as siblings; also P67, A601, and T1768 can be interpreted as siblings.

Experiment 3

In the third experiment, the data set was based on quasi-type-two variants and contained 68 parsimony-informative characters: around 10% of all characters in the data set.

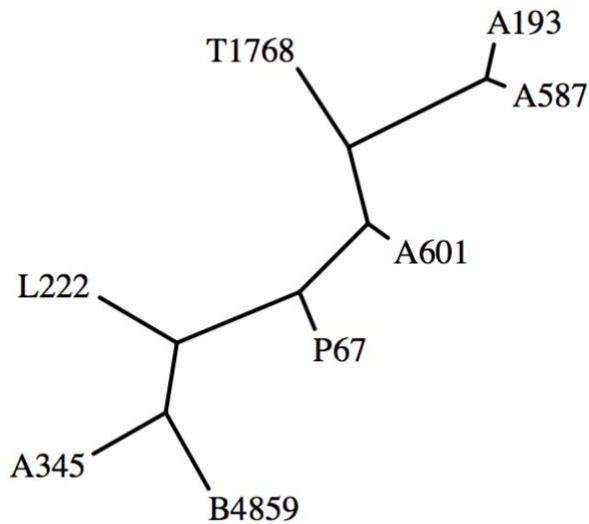


Figure 5: An unrooted tree of relationships: Experiment 3 (all quasi-type-two variants included)

As a result (Figure 5), one most-parsimonious tree was found: (((B4859, A345), L222), ((T1768, (A193, A587)), A601), P67). The results allow us to identify two groups of manuscripts: Ásgeir Jónsson's manuscripts (A193, A587, T1768) and Jón Þórðarson's group (A345, B4859, L222). The relationship between Jón Eggertsson's manuscripts is inconclusive, and A601 is grouped together with Ásgeir Jónsson's group.

Experiment 4

This experiment was based on type-two variants, which are listed in Table 2; variants are collated against P67. The symbol \checkmark indicates that the particular witness contains the same reading as P67, while when the readings disagree with P67 they are given in the table.

Table 2: Distribution of major type-two variants used in Experiment 4

	P67	A601	A587	A193	T1768	L222	A345	B4859
1.	Sá kongur réði fyrir Görðum	add. `í Danmörk/´	add. í Danmörk	add. í Danmörk	✓	✓	✓	✓
2.	engin herskip	✓	✓	✓	✓	herskip] skip	herskip] skip	herskip] skip
3.	og ræna drauga fé	✓	✓	✓	✓	om. fé	✓	om. fé
4.	heldur enn ræna kotkarla	✓	om.	om.	✓	✓	✓	✓
5.	Hrómundur þakkar karli fregn þessa að grjót og	✓	✓	✓	✓	fregn þessa] frá söguna	fregn þessa] frá söguna	fregn þessa] frá sögu
6.	stokkar gengu upp,	✓	stokkar] steinar	stokkar] steinar	✓	✓	✓	✓
7.	og víst ertu hraustur	✓	add. maður	add. maður	✓	✓	✓	✓
8.	Gunnlöð	✓	✓	✓	✓	Svílöð	Svílöð	Svílöð
9.	og er þeir voru á leið komnir	✓	✓	✓	✓	leið] veg	leið] veg	leið] veg
10.	koma af landi svört sky mér þótti	✓	✓	✓	✓	sky] kü	sky] kü	sky] kü
11.	jarnhringur settur	✓	✓	✓	✓	settur] sleginn	✓	settur] sleginn
12.	í `jördina´ upp að hjöltum	í /völlinn/ upp að hjöltum	í upp að hjöltum	í upp að hjöltum	í upp að hjöltum	í völlinn upp að hjöltum	í völlinn upp að hjöltum	í völlinn að hjöltum
13.	svo sverðið sókk að hjöltum	sókk] hljóp	sókk] hljóp	sókk] hljóp	sókk] hljóp	sókk] hljóp	om.	om.
14.	om.	ofan /í völlinn/	ofan	ofan	ofan	ofan í völlinn	om.	om.
15.	karl sagði, að hann	✓	karl] hann	karl] hann	✓	✓	✓	✓
16.	og að líðnum fjórum dögum	✓	fjórum] sex	fjórum] sex	fjórum] sex	✓	✓	✓

As presented in Table 2, this experiment is based mainly on type-two variants, but there are some exceptions from this rule, for example the definite form of the noun *saga*, against its indefinite form (no. 4) is not considered a variant; only reading *frá sögu* against *fregn þessa* is encoded as text-genealogically informative. Additionally, variants in the light grey rows (nos. 12-14) are type-four variants, which need some explanations. In no. 12, the omission of the preposition *upp* was not considered parsimony-informative; only the omission of *völlinn* is considered a major variant – because it creates a sentence without an object – thus creating a type-two variant in the input file. At the same time, it is open to discussion whether this reading can be considered as genealogically informative, since P67 has a supralinear addition of *jördina*, while A601 has *völlinn* deleted, so the scribe copying A601 could either restore the deleted

reading or follow the deletion, and create a sentence without an object, as in case of A587 and A193. In no. 13, P67 is the only manuscript that reads *sóck* for *hljóp*, while A345 and B4859 omit the entire reading; the omission of the entire phrase should be considered a major difference and this is why it was encoded in the input file. No. 14 could be considered a minor variant, since the omission of the preposition does not meet the criteria for major variants. However, the readings were included in the analysis as a type-four variant, because both readings 12 and 14 seem to be somehow related to each other, perhaps as a result of an error involving the omission of the word *völlinn*.

The variants in the dark grey rows (nos. 15-16) can be polygenetic, but are included in the analysis: no 16 because in some of the witnesses the numbers were spelled out, while other witnesses had roman numerals, so even though there is a high possibility that the roman numerals will be incorrectly copied, the written-out forms are more likely to remain unchanged; no 15 because the use of a personal pronoun for a noun did not fit directly with my definition of a minor variant, even though it might be polygenetic. Taking these methodological concerns into consideration, I have split the experiment into four subtests:

- Test 4a: based on 14 parsimony-informative characters (nos. 1-14)
- Test 4b: based on 16 characters (nos. 1-16)
- Test 4c: based on 13 characters (nos. 1-11, and 15-16)
- Test 4d: based on 11 characters (nos. 1-11)

As the result of Test 4a, three possible trees of relationships were found, which were identical in their shapes to the results of Experiment 2. The visualization of the tree is not reproduced here since it is identical with the one presented in Figure 4 above.

As the results of Test 4b, one most-parsimonious tree was found (Figure 6), which has a promising distinction between Ásgeir Jónsson's manuscripts (A193, A587, T1768), Jón Þórðarson's manuscripts (L222, B4859, A345), and Jón Eggertsson's manuscripts (P67, A601).

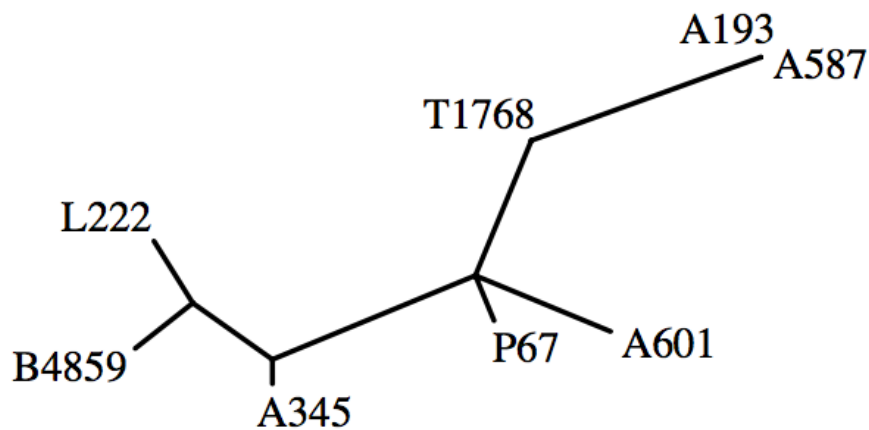


Figure 6: An unrooted tree of relationships: Experiment 4b (16 major type-two variants).

As the result of Test 4c, one most-parsimonious tree was found (Figure 7). In this unrooted tree P67 and A601 (Jón Eggertsson's manuscripts) are identical, A193 and A587 (Ásgeir Jónsson's manuscripts) are also identical and are related to T1768, while B4859 is no longer a sibling of L222, but is now presented as its descendant, which is a descendant of A345 (Jón Þórðarson's group).

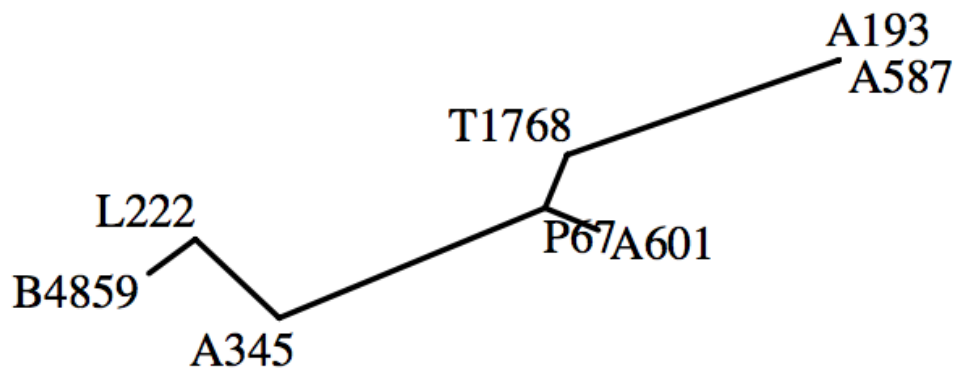


Figure 7: An unrooted tree of relationships: Experiment 4c (13 major, type-two variants).

As the result of Test 4d, one most-parsimonious tree was found (Figure 8). The distinction between P67, A601 and T1768 has disappeared, while other manuscripts have stayed in the same position relative to each other.

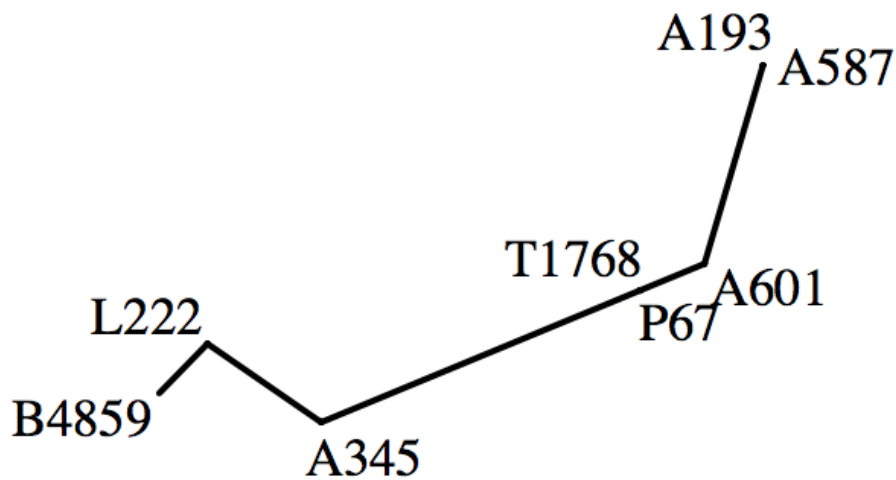


Figure 8: An unrooted tree of relationships based on 11 major, type-two variants: Experiment 4d.

Additional experiment 2'

Inspired by the significant differences in the results of Experiment 4, I decided to apply similar rules to the data set from Experiment 2, and include readings 15-16 (from Table 2), into the data set; thus the input file for Experiment 2' contained 75 parsimony-informative characters (11%).

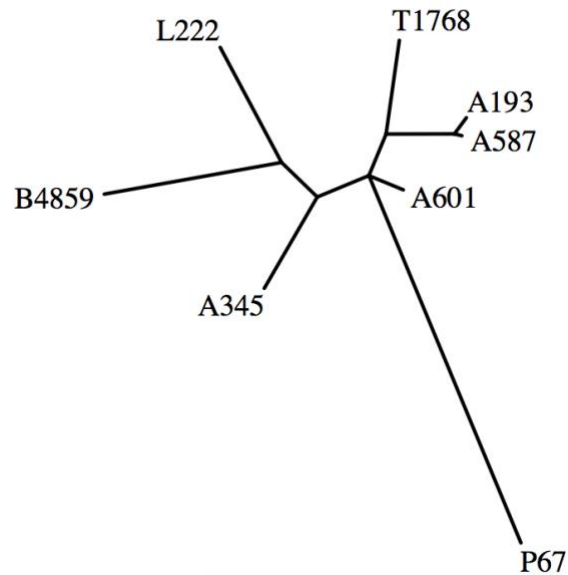


Figure 9: An unrooted tree of relationships: Experiment 2' (73 major variants, variants 1-16 from Table 2 included).

Excluding branch lengths, the trees obtained in Experiment 2' and Experiment 4b are identical, and in Newick notation can be represented as ((A345, (B4859, L222)), (T1768, (A193, A587)), A601, P67).

Discussion

The experiments presented in the previous section tested the effectiveness of the assorted types of variants for cladistic analysis. Not surprisingly, the results show that the shape of unrooted trees of relationships changes depending on the criteria applied while selecting the potentially genealogically informative readings. As presented by tests within Experiment 4, but also 2 and 2', the analysis is very sensitive to any change in the data set, namely that the presence or absence of two or three variants in the input file resulted in different results, varying from three possible trees of relationships to one most-parsimonious tree.

Generally, it can be observed that, regardless of which variants were included in the input file, the general grouping of the manuscripts corresponds to the expected grouping by scribe. In all experiments, manuscripts in Jón Þórðarson's hand (A345, B4859, L222) were grouped as derived from a common ancestor, as were all manuscripts in Ásgeir Jónsson's hand (A193, A587, T1768). However, the relationship between manuscripts in Jón Eggertsson's hand (A601, P67) was not obvious, and it must be emphasized that one of them, A601, had previously been considered a *codex optimus*, so its position is especially interesting from the perspective of the HsG tradition. The tree resulting from Experiment 1 corresponds closely to the tree from Experiment 3, while the consensus tree achieved in Experiment 2' corresponds to the tree from Experiment 4b. So, the data sets based on major variants deliver similar results, while the tests based on all types of variants result in a different grouping. The change can be observed in Jón Þórðarson's group: in Experiments 1 and 3, based on all variants, A345 can be interpreted as a sibling of B4859, and their parent as a sibling of L222, while in Experiments 2' and 4b, based on major variants, L222 can be interpreted as a sibling of B4859 and their parent as a sibling of A345.

Even though Salemans, following Greg, claimed that only type-two variants can be used to build a stemma, these experiments show that the shape of the tree is more influenced by the type of variation included in the analysis than by the selection of exclusively type-two variants, as in Experiment 1 and 2, 3 and 4. The tree resulting from Experiment 2', which is based on all major variants, seems to represent a better hypothesis of the relationships between the witnesses within the tradition under scrutiny than the one in Experiment 4b. This is certainly a result of including readings which are unique to single witnesses (*lectio singularis* or type-one variants), since some of them might play an important role as separative errors. I did not test how type-four variation influences the results of the analysis, but they were also included in the input files of Experiments 2 and 2'.

The similarities between trees based on minor variants and trees based on major variants require further explanation. The number of parsimony-informative characters included in each experiment was as follows: Experiment 1: 255 (all variants), 2: 73 (major variants), 3: 68 (quasi-type-two variants), 4: a - 14, b - 16, c - 13, d - 11 (major type-two variants), and 2': 75 (major variants). This shows that minor variants, which are likely to be polygenetic or individual to a scribe, established the vast majority of the variation within the HsG tradition. 71% of the characters in Experiment 1 can be considered to be minor variants; similarly 79% in Experiment 3. In both cases there is a remarkable imbalance between major and minor variants in favor of the latter; thus the results achieved in Experiments 1 and 3 are largely based on minor variants, which overwrite any potentially genealogical information carried by major variants.

After a close examination of the distribution of selected minor variants, as presented in Table 3, one might get the impression that there are some patterns which perhaps reflect the stylistic preferences of the scribe. For example, the use of the historical present (no. 16) seems to be a stylistic choice by Ásgeir Jónsson, while neither Jón Þórðarson nor Jón Eggertsson employ this form in this particular example. In some instances, however, other scribes also use the historical present, for example Jón Þórðarson in L222 and Jón Eggertsson in both P67 and A601 (no. 15). Moreover, the historical present is very common, effectively the default form used in Icelandic saga narrative: so it cannot be considered as a distinctive feature of Ásgeir Jónsson's manuscripts. The use of the singular instead of plural genitive of *ferð* "journey" (no. 1) seems again to be an innovation by Ásgeir Jónsson, who additionally in T1768 changed the tense of a verb from *bjóst* (imperf. passive) to *býst* (pres. passive).

Other variants, which I refer to as "linguistic variants" (nos. 3-12), are not consistently distributed between the scribes. Even though some of them give this impression, such as Ásgeir Jónsson's choice of the definite form of the noun *sverðið* "the sword" (no. 3), which is common to all of his manuscripts, it is rather difficult to defend as being not polygenetic. Similarly, it seems as though Ásgeir Jónsson preferred the unrounded form *gera* instead of *gjöra* "to do" (nos. 5-6), and words starting *hver-* instead of *hvor-* (nos. 9-10); elsewhere in the saga, however, he used *hvornin*, for example in "spurðu menn þa hvorninn þræinn oc hann hefþo skilip" (A587, 5r:15-16), and *hvorn* in "Hrongsvidur matti kiosa hvorn `dag´ mann fyrir sverdsins odde" (T1768, 261v:13-14).

Word order seems to be even more accidental, as in example 17, where T1768 agrees with L222, and in example 2, where T1768 agrees with A345, but it is impossible to prove any relationship between these manuscripts. The distribution of conjunctions, pronouns, and prepositions also seems to be absolutely fortuitous, as shown in examples 12-14.

Table 3: Distribution of selected minor (quasi type-two variants).

	Jón Eggertsson		Ásgeir Jónsson			Jón Þórðarson		
	P67	A601	A587	A193	T1768	L222	A345	B4859
1.	hann bjóst til ferða	ferða] ferðar	ferða] ferðar	ferða] ferðar	bjóst] býst, ferða] ferðar	✓	✓	✓
2.	í hólmgöngu m þú vanst með Mistilteini	✓	✓	✓	þú vanst í hólmgöngu með Mistilteini	✓	þú vanst í hólmgön gu með Mistilteini	✓
3.	hvar sverð hangir	sverð] sverðið	sverð] sverðið	sverð] sverðið	sverð] sverðið	✓	✓	✓
4.	þeir gjörðu svo	✓	✓	gjörðu] gerðu	✓	✓	gjörðu] gerðu	gjörðu] gerðu
5.	og svo var gjört	gjört] gert	gjört] gert	gjört] gert	✓	✓	gjört] gert	gjört] gert
6.	og gjörðist þar	✓	gjörðist] gerðist	gjörðist] gerðist	✓	✓	✓	gjörðist] gerðist
7.	engin jarn	✓	engin] engi	engin] engi	✓	engin] engi	engin] engi	engin] engi
8.	hvorn dag	✓	hvorn] hvern	hvorn] hvern	✓	✓	✓	hvorn] hvern
9.	hvor sá væri	hvor] hver	hvor] hver	hvor] hver	hvor] hver	✓	✓	✓
10.	hvort er nafn þitt	✓	hvort] hvert	hvort] hvert	hvort] hvert	hvort] hvert	✓	✓
11.	spurðu menn þá hvornin	✓	✓	hvornin] hvernin	hvornin] hvernin	✓	✓	hvornin] hvernin
12.	sagði Blindur	✓	✓	add. að	✓	✓	add. að	✓
13.	og kastaði niður	og] enn	og] enn	og] enn	✓	og] enn	og] enn	og] enn
14.	þó sér væri niður slept	✓	✓	✓	✓	þó] að	þó] að	þó] að
15.	hann tekur sér kylfu	✓	✓	✓	✓	✓	tekur] tók	tekur] tók
16.	og þegar þeir komu	✓	komu] koma	komu] koma	komu] koma	✓	✓	✓
17.	hafa með göldrum gjört ísin	✓	✓	✓	með göldrum hafa gjört ísin	með göldrum hafa gjört ísin	gert hafa ísin með göldrum	✓

Even if the minor variants can be regarded as individual to a scribe, the grouping we achieved using them is based on some sort of "scribal signal" and not on genealogically significant information. In the case of the oldest witnesses of HsG, the general grouping was surprisingly accurate, most likely because there were only three scribes who might have been somewhat consistent in their stylistic choices, but it does not mean that the relationship between the manuscripts was correctly established from a genealogical point of view. If we consider the minor variants as polygenetic, assuming that any scribe at any point in time could make the same change (for example in the word order or use of adverbs), then the results obtained with the use of minor variants must be considered inconclusive.

The distribution of the major variants does not deliver definite results either. As shown in Experiment 4, the shape of the tree changes depending on which variants are considered major and whether we include 11, 13, 14, or 16 parsimony-informative variants in the input file. Moreover, the results obtained from the data set based on 11 variants do not allow us to determine the relationship between A601, P67, and T1768, since they are identical in the matter of major type-two variants. The only reading that differentiates them in the input file is the reading "sá konungur réði fyrir Görðum," for which A601 reads "sá konugur réði fyrir /Görðum/ `i Danmörk´." It should be noted that in A601 "Görðum" is crossed out, and "í Danmörk" is a supralinear addition, which was later also crossed out: thus T1768, A587, and A193 are the only witnesses that preserve the fully written out form "sá konungur réði fyrir Görðum í Danmörk." In this case, one has to refer to variants classified in this paper as minor variants in order to reveal the relationships between these manuscripts, which brings into question the purpose of variant classification.

Taking the external evidence into consideration, or rather material aspects of the manuscripts A601 and P67, the tree obtained in experiments 4b and 2' seems to represent a better hypothesis about the relationship between these manuscripts than the others and, together with the consensus tree of Experiments 2 and 4a, these trees are close to the outcome of Kölbing's hypothesis. However, it should not be forgotten that P67 and A601 are both in Jón Eggertsson's hand, and one can be a copy of the other.

There is no space here for an extensive discussion of the material features of these manuscripts, but it is important to mention some of them. P67 is a large manuscript in folio with the Old Norse-Icelandic text written only on the verso sides of the leaves, while the recto sides are left blank, presumably to leave room for a Latin or Swedish translation. This is typical of manuscripts from

the late seventeenth and early eighteenth centuries, for example, Papp. fol. nr 73, dated to 1738, preserves *Gríms saga loðinkinna*, *Ketils saga hængs* and *Örvar-Odds saga*, with the Old Norse-Icelandic text on the verso side and a Swedish translation on the recto side; Papp. fol. nr 90, dated to 1683-1720, preserves *Hálfðanar saga Eysteinsonar* written in two columns, Old Norse-Icelandic on the left-hand side and Swedish on the right-hand side; similarly Papp. fol. nr 88, dated to 1683-1691, preserves *Göngu-Hrólf's saga* written in two columns with Old Norse and Swedish side by side. P67 does not employ many abbreviations and the text is not very dense, which might suggest a representative function that this manuscript was originally meant to have. A601 is a quarto manuscript with a denser text, containing many abbreviations and corrections in the scribe's own hand. A601 might therefore be considered as a draft, or a fast copy made while the scribe had access to the exemplar for a limited period of time, perhaps during Jón Eggertsson's incarceration in Copenhagen (cf. Jucknies, 2009). Extensive textual differences between A601 and P67 may suggest that P67 is an intentionally revised version of the text in A601, and thus useless from a text-genealogical point of view.

This brings us to some final considerations of the purpose of the stemma. If the stemma plays only the role of visualizing a general network of relationships between manuscripts, and not as a tool for application of majority rule, as it does in traditional textual criticism, then perhaps the results achieved in the experiments based on all types of variants are sufficient to fulfill this role and the classification of variants is not necessary. It is especially relevant in the light of Experiment 4d, which did not allow us to draw any conclusions about the filiation of some of the manuscripts. Moreover, if one is inclined to build a stemma based exclusively on 11 variants, then it seems more reasonable to draw the stemma by hand rather than to go through the entire process of complete transcriptions, collations, and computer processing. Also, if the results achieved by a computer-assisted analysis are only of a preliminary nature and always require manual adjustments, as recently suggested by Buzzoni et al. (2016, pp. 652, 665), and a stemma is always a hypothesis and simplification of the manuscript tradition, then perhaps it is not crucial whether we take the results from Experiments 1 or 4b as a point of departure.

Conclusion

The results of my experiments suggest that the cladistic method can be employed in traditional textual research, and a careful selection of variants can improve the results, but it does not guarantee conclusive results. As shown by the results of Experiment 4 and 2', the input file can be based only on traditionally selected major variants (or major type-two variants) and the results still have some sign of manuscript filiation, so the data set can be built exclusively on *loci critici* characters. Computer-assisted methods improve the efficiency of a textual analysis by delivering a rough hypothesis of the relationships between manuscripts relatively quickly, but the results

achieved through this process are of a preliminary nature and require further detailed investigation to construct the stemma.

The presence of minor variants in the case of HsG did not disturb the general grouping, but the relationships between the selected manuscripts were different from the ones based on major variants only. It has to be emphasized that we were dealing with only three scribes in this case study. Even if the minor variants are scribe-specific and if one can distinguish Ásgeir Jónsson's group from Jón Þórðarson's group based on their "scribal signal", it does not mean that these minor variants are genealogically informative because descendants of these manuscripts might not preserve these features. This is a vital question which requires further investigation, and is a subject of my current on-going research on the complete manuscript tradition of HsG.

REFERENCES

Manuscripts by repository

Stofnun Árna Magnússonar í íslenskum fræðum, Reykjavík:

- AM 193 e fol.
- AM 345 4to
- AM 587 b 4to
- AM 601 b 4to

Det Kongelige Bibliotek, Copenhagen:

- Thott 1768 4to

Landsbókasafn Íslands og Háskólabókasafn, Reykjavík:

- Lbs 222 fol.

Kungliga biblioteket, Stockholm:

- Isl. papp. fol. nr 67
- Isl. papp. fol. nr 73
- Isl. papp. fol. nr 88
- Isl. papp. fol. nr 90

British Library, London:

- BL Add. 4859

Secondary literature

Andrews, A. L. (1911). Studies in the fornaldarsögur Norðurlanda. *Modern Philology*, 8, 527–544.

Andrews, T. L. (2016). Analysis of Variation Significance in Artificial Traditions Using Stemmaweb. *Digital Scholarship in the Humanities*, 31(3), 523–539.

- Bordalejo, B. (2015). The Genealogy of Texts: Manuscript Traditions and Textual Traditions. *Digital Scholarship in the Humanities*, 31(3), 563–577. Retrieved from <https://doi.org/http://dx.doi.org/10.1093/llc/fqv038>
- Buzzoni, M., Burgio, E., Modena, M., & Simion, S. (2016). Open versus closed recensions (Pasquali): Pros and cons of some methods for computer-assisted stemmatology. *Digital Scholarship in the Humanities*, 31(3), 652. Retrieved from <https://doi.org/10.1093/llc/fqw014>
- Felsenstein, J. (2013). PHYLIP (Phylogeny Inference Package) (Version 3.695). Seattle: Department of Genome Sciences. Retrieved from <http://evolution.genetics.washington.edu/phylip/doc/main.html>
- Greg, W. W. (1927). *The calculus of variants, an essay on textual criticism*. Oxford: Clarendon Press.
- Greg, W. W. (1950). The Rationale of Copy-Text. *Studies in Bibliography*, 3, 19–36.
- Hall, A., & Parsons, K. (2013). Making stemmas with small samples, and digital approaches to publishing them: testing the stemma of Konráðs saga keisarasonar. *Digital Medievalist*, 9.
- Howe, C., Barbrook, A., Mooney, L., & Robinson, P. (2004). Parallels Between Stemmatology and Phylogenetics. In *Studies in Stemmatology II* (pp. 3–11). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Jucknies, R. (2009). *Der Horizont eines Schreibers, Jón Eggertsson (1643-89) und seine Handschriften*. Frankfurt am Main: Peter Lang.
- Kölbing, E. (1876). *Beiträge zur Vergleichenden Geschichte der Romantischen Poesie und Prosa des Mittelalters*. Breslau: Verlag von Wilhelm Koebner.
- O'Hara, R. J., & Robinson, P. M. W. (1993). Computer-assisted methods of stemmatic analysis. *Occasional Papers of the Canterbury Tales Project*, 1, 53–74.
- Platnick, N. I., & Cameron, H. D. (1977). Cladistic Methods in Textual, Linguistic, and Phylogenetic Analysis. *Systematic Zoology*, 26(4), 380–385. Retrieved from <https://doi.org/10.2307/2412794>
- Rafn, C. C. (1829). *Fornaldarsögur Norðrlanda*. Kaupmannahöfn.
- Robinson, P. (1996). Computer-Assisted Stemmatic Analysis and “Best-Text” Historical Editing. In *Studies in Stemmatology* (pp. 71–104). Amsterdam: John Benjamins Publishing Company.
- Robinson, P. (2004). The Old Norse Sólurljóð. NEXUS file created from “NexSeg.xml” Monday September 13 16:07:41 2004. Retrieved from <http://www.textualscholarship.org/newstemmatics/data/Sol.nex>
- Robinson, P. (2016). Four rules for the application of phylogenetics in the analysis of textual traditions. *Digital Scholarship in the Humanities*, 31(3), 637. Retrieved from <https://doi.org/10.1093/llc/fqv065>

- Robinson, P., & Bordalejo, B. (2010a). The New Stemmatics: Data. Retrieved from <http://www.textualscholarship.org/newstemmatics/data/index.html>
- Robinson, P., & Bordalejo, B. (2010b). *What is "The New Stemmatics"?* Retrieved from <http://www.textualscholarship.org/newstemmatics/index.html>
- Robinson, P., & O'Hara, R. J. (1996). Cladistic analysis of an Old Norse manuscript tradition. *Research in Humanities Computing*, 4, 115–137.
- Salemans, B. (1987). Van Lachmann tot Hennig: cladistische tekstkritiek. *Gramma*, 11, 191–224.
- Salemans, B. (1996). Cladistics or the Resurrection of the Method of Lachmann: On Building the Stemma of Yvain. In *Studies in Stemmatology* (Pieter van Reenen, Margot van Mulken, Janet Dyk). Amsterdam: John Benjamins Publishing Company.
- Trovato, P. (2014). *Everything you always wanted to know about Lachmann's method: a non-standard handbook of genealogical textual criticism in the age of post-structuralism, cladistic, and copy-text*. Padova: Libreriauniversitaria.it edizioni.
- van Mulken, M. (1993). *The manuscript tradition of the Perceval of Chrétien de Troyes. A stemmatological and dialectological approach*. University of Amsterdam, Amsterdam.
- Zeevaert, L., Hall, A., Kapitan, K. A., et. al. (2013). A new stemma of Njáls saga -a working paper. Retrieved from [https://www.academia.edu/7317515/A New Stemma of Njáls saga](https://www.academia.edu/7317515/A_New_Stemma_of_Njáls_saga)

ACKNOWLEDGMENTS

This article reflects on the content of the paper presented during the International Digital Humanities Symposium held in Vaxjö, Sweden 7-8 November 2016, organized by Koraljka Golub, Marcelo Milrad, and Tamara Laketic. I would like to thank the organizers for allowing me to present my methodological considerations during the symposium and participants of the conference for some stimulating discussions. The content of this article exceeds the subject of my original presentation and it is based on the work I have conducted as a part of my PhD fellowship at the University of Copenhagen (2015-2018). I would like to thank the supervisors of my PhD project Matthew James Driscoll and Annette Lassen for their feedback, as well as my colleagues Sheryl McDonald Werronen, Tarrin Wills, and Seán Douglas Vrieland, who read drafts of this paper and helped to improve its style.