Haukur Þorgeirsson*

How similar are Heimskringla and Egils saga? An application of Burrows' delta to Icelandic texts

Abstract: Recent methodological and technological developments greatly facilitate the use of stylometry for authorship attribution. Burrows' delta method, proposed in 2002, has been shown to yield good results for a variety of corpora in different languages. The present article demonstrates that this method is highly effective in analysing 19th century Icelandic fiction. The method is then applied to the classical question of the stylistic affinity between two 13th century texts: *Heimskringla* and *Egils saga*. *Heimskringla* proves to be more similar to *Egils saga* than it is to a variety of contemporary texts, including other kings' sagas. This supports the theory that the two texts have the same author.

Introduction

The authorship of *Egils saga* has long been a popular topic of discussion among scholars of Old Norse literature.[1] In 2002, the saga was published as a part of Snorri Sturluson's collected works with an introduction stating that "most scholars" accept the attribution to Snorri (Vésteinn Ólason 2002, lxiv). While this view is certainly widespread, skeptical and dissenting voices are not difficult to find. Guðrún Nordal (2002) has argued against attributing the preserved text to Snorri and Margaret Cormack (2001) finds it difficult to reconcile certain differences between *Egils saga* and *Heimskringla* with a common authorship. Ármann Jakobsson describes the authorship of the saga as "spurious" and calls for "further weighing of the evidence" (Ármann Jakobsson 2002, 146). Jonna Louis-Jensen (2009, 2013) rejects Snorri's authorship and argues that *Egils saga* has a more archaic style than authentic Snorri texts.

---

*Corresponding author: Haukur Þorgeirsson, Stofnun Árna Magnússonar, Árnagarði við Suðurgötu, 101 Reykjavík, Iceland, e-mail: haukurth@hi.is

A major component of the case for Snorri's authorship is based on a comparison of vocabulary and style between *Egils saga* and *Heimskringla*. The most energetic proponent of this approach was Peter Hallberg. In a series of publications (principally Hallberg 1962, 1963, 1965, 1968) he argued that the commonalities in the use of rare words and certain stylistic features formed a compelling case for common authorship. A later study by Ralph West used computer-aided analysis to conclude that "Snorri did indeed write Egils saga" (West 1980, 191).

The kind of study Hallberg and his successors have engaged in is referred to as *stylometry* or *non-traditional authorship analysis*. Since the advent of the digital age this discipline has developed rapidly. The major studies on *Heimskringla* and *Egils saga* were conducted before 1980. Since then, there has been a great deal of technological and methodological progress in the field so it is high time for a fresh examination of the evidence.[2]

Methodology

In the early days of stylometry research, scholars were hampered by the lack of a standard or accepted methodology. This is evident in Hallberg's studies, where *ad hoc* methods tend to be developed for each new attribution problem, for the most part without any particular theoretical justification. Another problem typical of early stylometry is the lack of a control corpus. A control corpus is a collection of texts by known authors which can be used to verify that a proposed stylometry method is actually effective at authorship attribution. In the works of Hallberg and West, only very limited steps are taken to verify that the methods used are empirically effective.

By mentioning these limitations, I am by no means claiming that Hallberg's contributions to stylistic research are worthless. On the contrary, I think any reader of Hallberg will come to appreciate both his keen eye for stylistic details and his almost superhuman patience for collecting data. Nevertheless, there is clearly a lot of room for further work.

In the present study I apply a tried and tested stylometry method to the problem of *Egils saga* and *Heimskringla*. In the last decade, a method proposed by Burrows (2002) has been shown to be effective for a variety of corpora in different

---

2 As this present article was in proofs an article on saga authors and stylometry by Jón Karl Helgason, Sigurður Ingibergur Björnsson and Steingrímur Páll Kárason appeared in *Skírnir* 2017. Their methods are different but their results further cement the connection between Heimskringla and Egils saga.

languages and gained widespread acceptance (Fotis et al., 2015; Eder and Rybicki, 2013). It is particularly valuable and convenient that implementations of Burrows' method, with several variations, are freely available.

I am not aware of any previous testing of Burrows' method for Icelandic texts so it will be valuable to work through a control corpus. This will also give me the opportunity to explain the working of the method.

In the following I use Hoover's (2005–2015) implementation of Burrow's method. The method is, for the most part, intuitive and does not require a heavily technical background to understand. The basic idea is that each author has a characteristic word frequency pattern. By comparing word frequency in two different texts we can establish *delta*, a statistical measure of the difference between the texts. When a pair of texts has a low delta this can, in the right context, indicate that they have the same author.

Control corpus experiment – 19th century fiction

For a control corpus we need a collection of substantial texts by known authors working in the same time period. The texts should be similar in kind, not e. g. a mix of fiction, personal letters and spiritual literature. Since most Old Icelandic texts are anonymous, there is little prospect of an Old Icelandic control corpus. Instead I turn to modern Icelandic texts, specifically fiction from the period 1850–1920, where a number of out-of-copyright texts are available in digital form from *Netútgáfan*. My *primary sample* consists of works by five authors:

Table 1: Primary corpus.

Halla by Jón Trausti (1873–1918)	53,410 words
Piltur og stúlka by Jón Thoroddsen (1818–1868)	56,909 words
Upp við fossa by Þorgils gjallandi (1851–1915)	58,546 words
Brynjólfur biskup Sveinsson by Torfhildur Hólm (1845–1918)	75,082 words
Five short stories (*Grímur kaupmaður deyr, Hans Vöggur, Kærleiksheimilið, Skjóni* and *Uppreistin á Brekku*) by Gestur Pálsson (1852–1891)	24,521 words

I have obtained digital copies of those texts from *Netútgáfan* and removed some material extraneous to the prose fiction under examination: chapter headings, poetry quotes and non-narrative prologues and epilogues.

To analyze the primary sample we compile tables showing the number of occurrences of each word form in each text. Various tools exist for this purpose,

I have used Linux command line utilities but applications with graphical user interfaces are also available. The results for *Halla*, in abbreviated form, are as follows:

Table 2: Occurrences of word forms in *Halla*.

og	2697
að	2398
hann	1360
var	1186
í	1056
… 8896 lines omitted …	
örvæntingin	1
öryggis	1
öskuþreifandigrenjandiblindviðrishríð	1
öxl	1
öxlina	1

When we have tabulated word form occurrences for each text in the primary sample, we can assemble a table with frequencies. For example, the lexeme *og* occurs 2697 times in *Halla*, which has a total of 53,410 words – thus there are about 505 occurrences per 10,000 words.

Table 3: Frequency per 10,000 words in primary corpus.

	JTr	JTh	Þg	TH	GP	StDev
og	505	498	585	469	513	43.2
að	449	496	400	381	495	52.9
í	198	201	194	210	216	9.1
á	173	208	175	193	224	21.8
það	193	208	173	124	175	31.7

Table 3 shows the frequency of the five most common words in the primary corpus. The table also shows the standard deviation for each lexeme, a measure of the variation in the data. The frequency of the lexeme *að* varies substantially more in our texts than that of the lexeme *í*, thus the former has a much higher standard deviation.

Before proceeding further we may wonder if we should exclude some words from the comparison. Research on English texts has shown that results can sometimes be improved by removing pronouns. This is helpful to attenuate the differ-

ence between first and third person narratives and the effect of the gender of the main characters. I have experimented with removing pronouns from the analysis but this did not have a substantial effect so I have included them in the results published here.

Another issue is the presence of words primarily or exclusively found in one of the primary texts. Table 4 shows some examples:

Table 4: A selection of words only frequent in one primary text.

	JTr	JTh	Þg	TH	GP
biskups	0	0	0	15	0
anna	0	0	0	0	34
skálholti	0	0	0	8	0
grímur	0	0	0	0	23
frændi	0	0.5	0	4	0

Words which are frequent in one text but rare or non-existent in the others are typically names or other very context-specific words. It is moderately helpful to remove these from consideration. In the following analysis I have, following Hoover (2005–2015), removed words where one of the primary texts is responsible for more than 70 % of the occurrences.

A final question which needs to be answered is how many words are to be included in the analysis. Burrows' original demonstration used 150 words but subsequent research has shown that using more words improves accuracy, up to a point (Hoover, 2005–2015). The ideal number of words depends on the length of the texts. For the following analysis I have chosen to use the 1000 most frequent words, which is a reasonable compromise for texts of varying length. For good measure I also include runs with the 500 and 2000 most frequent words.

Test corpus A – works by the five authors

After this preparatory work we can get down to the business of evaluating sample texts. I use the following novels, novellas and short stories by the authors from the primary corpus. I am in no way cherry-picking texts, my corpus consists of all the texts available to me:

Table 5: Test corpus 1.

Leysing by Jón Trausti	112,864 words
Maður og kona by Jón Thoroddsen	93,293 words
Anna frá Stóruborg by Jón Trausti	55,222 words
Gamalt og nýtt by Þorgils gjallandi	27,119 words
Hækkandi stjarna by Jón Trausti	23,170 words
Veislan á Grund by Jón Trausti	22,565 words
Borgir by Jón Trausti	21,368 words
Seingróin sár by Þorgils gjallandi	18,307 words
Snæfríðar þáttur by Þorgils gjallandi	12,662 words
Vordraumur by Gestur Pálsson	9,160 words
Aftanskin by Þorgils gjallandi	7,641 words
Söngva-Borga by Jón Trausti	7,408 words
Þjóðólfsþáttur by Þorgils gjallandi	6,624 words
Á fjörunni by Jón Trausti	5,709 words
Gísli húsmaður by Þorgils gjallandi	5,113 words
Tvær systur by Jón Trausti	4,935 words
Týndu hringarnir by Torfhildur Hólm	4,424 words
Strandið á Kolli by Jón Trausti	3,884 words
Friðrik áttundi by Jón Trausti	3,790 words
Í minni hluta by Þorgils gjallandi	3,414 words
Við sólhvörf by Þorgils gjallandi	2,704 words
Ef Guð lofar by Þorgils gjallandi	2,557 words
Frá Grími á Stöðli by Þorgils gjallandi	1,996 words
Brestur by Þorgils gjallandi	1,492 words
Bernskuminning by Þorgils gjallandi	1,357 words
Einar Andrésson by Þorgils gjallandi	1,248 words
Ósjálfræði by Þorgils gjallandi	1,055 words
Vetrarblótið að Gaulum by Þorgils gjallandi	961 words

We now proceed to tabulate the frequency of word forms in each sample text and compare the result with the samples in the primary set. The results for the top five words are as follows in a comparison between *Leysing* and the Jón Trausti primary sample (*Halla*):

Table 6: Z-scores for a comparison between *Leysing* and *Halla*.

	JTr	Leysing	Difference	StDev	Z-score
og	505	398	107	43.2	2.5
að	449	414	35	52.9	0.7
í	198	238	46	9.1	4.4
á	173	209	40	21.8	1.7
það	193	143	36	31.7	1.6

We obtain a z-score, also known as a standard score, by dividing the difference in frequency with the standard deviation. To finally obtain a Burrows' delta score we add up the z-score of all 1000 words. When we do this we get the following results for *Leysing*:

Table 7: Delta scores for *Leysing*.

Author	Delta
Jón Trausti	925
Þorgils gjallandi	1152
Gestur	1174
Torfhildur	1222
Thoroddsen	1296

We see that delta is lowest between *Leysing* and Jón Trausti (*Halla*), which is consistent with the fact that Jón Trausti is the actual author of the text. The method, then, has scored a success but we may be interested in some measure of confidence in a given result. We see from the table that the difference between the best and the second best match is 227 while the difference between scores 2 and 5 is merely 144. The best match is clearly separated from the rest of the pack, which is consistent with high confidence in the result. As a convenient metric we can use the percentage increase in delta between the best and the second best match (i. e. the difference between the best two scores divided by the best score). In this case there is an increase of 25 %.

We can now show the results for all the test texts:

Table 8: A delta test for 28 texts of varying length (1000 most frequent words).

Text	Actual author	Lowest delta	Difference	Wordcount
Maður og kona	Jón Thoroddsen	Jón Thoroddsen	36.6 %	93,293
Gamalt og nýtt	Þorgils gjallandi	Þorgils gjallandi	27.3 %	27,119
Leysing	Jón Trausti	Jón Trausti	25.6 %	112,864
Seingróin sár	Þorgils gjallandi	Þorgils gjallandi	22.4 %	18,307
Snæfríðar þáttur	Þorgils gjallandi	Þorgils gjallandi	13.1 %	12,662
Gísli húsmaður	Þorgils gjallandi	Þorgils gjallandi	12.2 %	5,113
Aftanskin	Þorgils gjallandi	Þorgils gjallandi	10.9 %	7,641
Tvær systur	Jón Trausti	Jón Trausti	10.2 %	4,935
Anna frá Stóruborg	Jón Trausti	Jón Trausti	9.8 %	55,222
Borgir	Jón Trausti	Jón Trausti	9.4 %	21,368

Table 8: (Continued)

Text	Actual author	Lowest delta	Difference	Wordcount
Á fjörunni	Jón Trausti	Jón Trausti	9.4 %	5,709
Vordraumur	Gestur Pálsson	Gestur Pálsson	5.7 %	9,160
Þjóðólfsþáttur	Þorgils gjallandi	Þorgils gjallandi	5.6 %	6,624
Við sólhvörf	Þorgils gjallandi	Þorgils gjallandi	5.3 %	2,704
Veislan á Grund	Jón Trausti	Jón Trausti	4.3 %	22,565
Í minni hluta	Þorgils gjallandi	Þorgils gjallandi	3.4 %	3,414
Ef Guð lofar	Þorgils gjallandi	Þorgils gjallandi	3.2 %	2,557
Bernskuminning	Þorgils gjallandi	Þorgils gjallandi	2.6 %	1,357
Friðrik áttundi	Jón Trausti	Jón Trausti	2.6 %	3,790
Frá Grími á Stöðli	Þorgils gjallandi	Þorgils gjallandi	2.5 %	1,996
Hækkandi stjarna	Jón Trausti	Jón Trausti	2.0 %	23,170
Söngva-Borga	Jón Trausti	Jón Trausti	1.4 %	7,408
Vetrarblótið	Þorgils gjallandi	Þorgils gjallandi	1.1 %	961
Ósjálfræði	Þorgils gjallandi	Þorgils gjallandi	1.0 %	1,055
Brestur	Þorgils gjallandi	Þorgils gjallandi	1.0 %	1,492
Einar Andrésson	*Þorgils gjallandi*	*Jón Trausti*	0.8 %	1,248
Strandið á Kolli	Jón Trausti	Jón Trausti	0.6 %	3,884
Týndu hringarnir	Torfhildur Hólm	Torfhildur Hólm	0.4 %	4,424

The delta test identifies the correct author in 27 cases out of 28. Considering that many of the texts are quite short, this is an astonishingly successful run. In eight cases the algorithm identifies the correct author only by the skin of its teeth, with a difference score of 2.5 % or less. This suggests that we have been lucky and indeed it turns out that a small change in parameters yields less accurate results. If we run the test with the 500 most frequent words we get four erroneous attributions:

Table 9: Errors from a delta test of 500 MFW.

Text	Actual author	Lowest delta	Difference	Wordcount
Söngva-Borga	Jón Trausti	Þorgils gjallandi	1.0 %	7,408
Hækkandi stjarna	Jón Trausti	Torfhildur Hólm	0.7 %	23,170
Veislan á Grund	Jón Trausti	Torfhildur Hólm	0.5 %	22,565
Strandið á Kolli	Jón Trausti	Þorgils gjallandi	0.2 %	3,884

The average size of the incorrectly attributed texts is 14,257 words, this is consistent with the principle that longer texts typically benefit from taking more words into account. If we run the test with the 2000 most frequent words we get nine errors:

Table 10: Errors from a delta test of 2000 MFW.

Text	Actual author	Lowest delta	Difference	Wordcount
Frá Grími á Stöðli	Þorgils gjallandi	Gestur Pálsson	1.8 %	1,996
Brestur	Þorgils gjallandi	Jón Trausti	1.6 %	1,492
Vetrarblótið	Þorgils gjallandi	Jón Trausti	1.4 %	961
Strandið á Kolli	Jón Trausti	Gestur Pálsson	1.0 %	3,884
Einar Andrésson	Þorgils gjallandi	Jón Trausti	0.7 %	1,248
Ósjálfræði	Þorgils gjallandi	Gestur Pálsson	0.7 %	1,055
Bernskuminning	Þorgils gjallandi	Jón Trausti	0.4 %	1,357
Týndu hringarnir	Torfhildur Hólm	Jón Thoroddsen	0.3 %	4,424
Ef Guð lofar	Þorgils gjallandi	Jón Trausti	0.1 %	2,557

In this case the short texts run into difficulties, the average text has a length of 2,108 words. Consistent with previous research, short texts are most profitably analyzed with a relatively low number of words.

It is important to note that we do not get any erroneous attributions with a high difference score. All attributions with a score of 2 % or more are correct in the data examined so far.

Test corpus B – works by other authors

In the previous section we were in the happy position of dealing with texts which we knew to be by one of the authors in the primary set. A more difficult situation, and one more realistic in the context of our ultimate goals here, is when the text whose authorship is in question may not be by any of the authors whose stylistic fingerprint we are comparing with. To test the method against this possibility I have compared the primary corpus from the previous section with a test corpus consisting of prose by other authors. Since the supply of available texts is limited, this corpus includes some translations. The texts are as follows (all from *Netútgáfan* except *Umhverfis jörðina á 80 dögum* and *Fanginn í Zenda*, which were obtained from *Rafbókavefurinn*):

Table 11: Test corpus 2, seven objects by other authors.

Ævintýri og sögur by H. C. Andersen, translated by Steingrímur Thorsteinsson	136,827 words
Umhverfis jörðina á 80 dögum by Jules Verne (translator not listed)	59,876 words

Table 11: (Continued)

Fanginn í Zenda by Anthony Hope, translated by Stefán Björnsson (1876–1942)	58,861 words
Four short stories (*Brúðardraugurinn, Írafells-Móri, Ferðasaga* and *Þórðar saga Geirmundarsonar*) by Benedikt Gröndal	27,080 words
Björn í Gerðum by Jónas Jónasson	8,296 words
Brennivínshatturinn by Hannes Hafstein	5,155 words
Two short stories (*Grasaferð* and *Þegar drottningin á Englandi fór í orlof sitt*) by Jónas Hallgrímsson	4,915 words

The results of the delta test for those texts is as follows:

Table 12: Results for test corpus 2.

Text	Lowest delta	Difference	MFW	Word count
Stories by Gröndal	Jón Thoroddsen	0.1 %	500	27,080
Stories by Gröndal	Jón Thoroddsen	1.1 %	1000	27,080
Stories by Gröndal	Jón Thoroddsen	5.0 %	2000	27,080
Brennivínshatturinn	Þorgils gjallandi	0.3 %	500	5,155
Brennivínshatturinn	Gestur Pálsson	2.4 %	1000	5,155
Brennivínshatturinn	Gestur Pálsson	4.1 %	2000	5,155
Umhverfis jörðina	Jón Trausti	0.8 %	500	59,876
Umhverfis jörðina	Jón Trausti	1.5 %	1000	59,876
Umhverfis jörðina	Jón Trausti	4.0 %	2000	59,876
Fanginn í Zenda	Torfhildur Hólm	0.7 %	500	58,861
Fanginn í Zenda	Torfhildur Hólm	5.6 %	1000	58,861
Fanginn í Zenda	Torfhildur Hólm	3.1 %	2000	58,861
Stories by JH	Þorgils gjallandi	0.6 %	500	4,915
Stories by JH	Jón Thoroddsen	3.5 %	1000	4,915
Stories by JH	Jón Thoroddsen	2.9 %	2000	4,915
Ævintýri og sögur	Þorgils gjallandi	0.4 %	500	136,827
Ævintýri og sögur	Þorgils gjallandi	1.3 %	1000	136,827
Ævintýri og sögur	Jón Thoroddsen	0.5 %	2000	136,827
Björn í Gerðum	Þorgils gjallandi	4.3 %	2000	8,296
Björn í Gerðum	Jón Thoroddsen	0.7 %	2000	8,296
Björn í Gerðum	Jón Thoroddsen	0.0 %	2000	8,296

The difference scores are in the range 0.0 %–5.6 % and the average is 2.0 %. This is a bit higher than the results for the erroneously attributed texts in test corpus 1. But compared with the correctly attributed texts, this is not very high. Of the 27 correctly attributed texts in Table 8, 12 have a difference score higher than 5.6 %. This includes all four texts longer than 25,000 words.

Heimskringla and the major sagas of Icelanders

The preceding sections have allowed me to showcase Burrows' method and to establish that it is effective in identifying the authors of texts in Icelandic. I now turn to the main object of inquiry, the putative stylistic connection between *Heimskringla* and *Egils saga*.

In the control corpus, I used texts available in digital form and only exerted minimum effort to normalize their presentation. But when dealing with Old Icelandic texts, more care is needed. It would, for example, be absurd to compare editions with modern Icelandic orthography directly to editions with normalized Old Norse orthography. Since most digital texts available to me use modern orthography, I have decided to use this as a standard. More specifically, I have followed the conventions used at *Netútgáfan* for the Sagas of Icelanders but striven for more consistency.

Even in a framework of normalized spelling, editors may choose to retain some archaic forms or to follow the main manuscript on its choice between variant word forms. But since such variation is not likely to stem from the original author, I have sought to normalize it away. Some representative examples of variation which I have done away with are as follows:

> öngvir / engvir
> mart / margt
> öngvan / engvan / engan
> aldregi / aldri / aldrei
> yðart / yðvart
> gjafir / gjafar
> erendi / örendi / erindi
> þessari / þessi
> hvernug / hvernig
> hvetvetna / hotvetna / hvaðvetna
> þeira / þeirra
> myrgin / morgin / morgun
> orrusta / orusta
> durum / dyrum
> hlupu / hljópu
> nakkvað / nökkuð / nokkuð

As in the previous test, I have removed words where one primary text is responsible for more than 70 % of occurrences. I have also manually removed all remaining proper names.

We can now move on to defining a primary corpus. For the first study the question to be examined is whether one of the large sagas of Icelanders is significantly

more similar than the others to *Heimskringla*. This is the same question Hallberg originally studied. The primary corpus is as follows:

Table 13: Primary corpus, the sagas of Icelanders.

Njáls saga	98,926 words
Egils saga	62,109 words
Grettis saga	61,146 words
Laxdæla saga	57,496 words
Eyrbyggja saga	38,062 words

The work to be tested is *Heimskringla* which I will handle in two parts. There is good evidence that *Óláfs saga helga* ("Heimskringla A") was written separately from the rest ("Heimskringla B"). Indeed, Jonna Louis-Jensen (2009, 2013; see also 1997) argues that the two parts have different authors.

Table 14: Test corpus, the two parts of *Heimskringla*.

Heimskringla A	92,779 words
Heimskringla B	137,044 words

We can now compare our test corpus with our primary corpus. For the 1000 most frequent words, the results are as follows:

Table 15: Burrows' delta for both parts of *Heimskringla*, 1000 MFW.

Heimskringla A	Delta	Heimskringla B	Delta
Egils saga	968	Egils saga	1013
Eyrbyggja saga	1243	Eyrbyggja saga	1232
Grettis saga	1285	Grettis saga	1329
Laxdæla saga	1309	Laxdæla saga	1423
Njáls saga	1343	Njáls saga	1464

For both parts of *Heimskringla*, *Egils saga* is the most similar text, by a healthy margin. It matters little whether the analysis uses the 500, 1000 or 2000 most frequent words:

Table 16: The two parts of *Heimskringla* compared to the five sagas of Icelanders.

Text	Lowest delta	Difference	Most frequent words
Heimskringla A	Egils saga	34.8 %	500
Heimskringla B	Egils saga	24.2 %	500
Heimskringla A	Egils saga	28.4 %	1000
Heimskringla B	Egils saga	21.6 %	1000
Heimskringla A	Egils saga	32.5 %	2000
Heimskringla B	Egils saga	22.4 %	2000

These results are consistent with the theory that *Egils saga* and (both parts of) *Heimskringla* have the same author. But this is not the only conceivable explanation. One alternative possibility is that Egils saga has stylistic affinity with the kings' sagas in general rather than *Heimskringla* in particular. To check for this effect, comparison should be made with more kings' sagas.

Another alternative is that the similarity between *Egils saga* and *Heimskringla* is a consequence of their being composed around the same time while the other major sagas of Icelanders might be younger (Elín Bára Magnúsdóttir 2015, 274). In Vésteinn Ólason's categorization of the sagas of Icelanders, the sagas are divided into three chronological groups. *Egils saga* belongs to the oldest group (1200–1280) while *Grettis saga* is placed in the youngest group (1300–1450) and *Eyrbyggja saga*, *Njáls saga* and *Laxdæla saga* are placed in the middle group (1240–1310) (Vésteinn Ólason 1993, 42). To test against this idea, comparison should be made with works believed to be close in age to *Egils saga* and *Heimskringla*.

Heimskringla and other historical texts

For a more challenging test of the connection between *Egils saga* and *Heimskringla* I have prepared a second primary corpus to address the concerns raised in the previous section. In this case we are using a variety of historical texts for comparison:

Table 17: Primary corpus, a variety of historical texts.

Egils saga	62,109 words
Íslendinga saga	101,228 words
Jómsvíkinga saga	38,660 words
Knýtlinga saga	48,343 words
Morkinskinna (a sample)	30,039 words

Jómsvíkinga saga may have been composed in the early 13th century. It is a *sui generis* text with affinity to the kings' sagas (Finlay, 2014). *Knýtlinga saga* is one of the kings' sagas – seemingly composed in conscious imitation of *Heimskringla*. There are reasons to suppose that it was composed by Óláfr Þórðarson (d. 1259), Snorri Sturluson's nephew (Bjarni Guðnason 1982, clxxxix–clxxxiv).[3] *Íslendinga saga* deals with events in Iceland in the 13th century. It is attributed to Sturla Þórðarson (1214–1284), another nephew of Snorri Sturluson. *Morkinskinna* is one of the kings' sagas, believed to have been composed shortly before *Heimskringla*. It is one of *Heimskringla's* sources.

Jómsvíkinga saga was obtained from *Netútgáfan*, *Knýtlinga saga* was obtained from the *Heimskringla* text collection, *Íslendinga saga* was specially provided by the Árni Magnússon Institute from *Íslenskt textasafn*, the *Morkinskinna* text was typed up from the *Íslenzk fornrit* edition (Ármann Jakobsson and Þórður Ingi Guðjónsson 2002) – I limited myself to the first 30,000 words of the text preserved in the oldest manuscript (GKS 1009 fol.). As in the previous case, I took pains to normalize all the texts to the same standard.

The results of a test with the 1000 most frequent words yields the following results:

Table 18: Burrows' delta for both parts of *Heimskringla*, 1000 MFW.

Heimskringla A	Delta	Heimskringla B	Delta
Egils saga	803	Egils saga	873
Knýtlinga saga	936	Knýtlinga saga	897
Íslendinga saga	1138	Íslendinga saga	1066
Morkinskinna	1147	Jómsvíkinga saga	1248
Jómsvíkinga saga	1149	Morkinskinna	1301

The results are largely consistent with different parameters:

Table 19: The two parts of *Heimskringla* compared to five historical texts.

Text	Lowest delta	Difference	Most frequent words
Heimskringla A	Egils saga	17.4 %	500
Heimskringla B	Egils saga	6.2 %	500

---

3 Males (2015) argues that Óláfr Þórðarson composed *Egils saga*. The stylometric observations here would not necessarily rule this out since they do show some stylistic affinity between *Egils saga* and *Knýtlinga saga*.

Table 19: (Continued)

Text	Lowest delta	Difference	Most frequent words
Heimskringla A	Egils saga	16.4 %	1000
Heimskringla B	Egils saga	2.7 %	1000
Heimskringla A	Egils saga	12.6 %	2000
Heimskringla B	Egils saga	1.8 %	2000

These results demonstrate that the stylistic affinity between *Heimskringla* and *Egils saga* is not merely the consequence of a closeness in age or a closeness between *Egils saga* and the kings' sagas. Consistent with previous research, *Óláfs saga helga* is found to be even more similar to *Egils saga* than the other parts of *Heimskringla* are.

Comparison with other methods

This is not the first attempt to determine the degree of stylistic affinity between *Egils saga* and *Heimskringla* and it is natural to ask how it compares with previous research. The principal difference between this method and Hallberg's (1962) influential pair word investigation is that here we are dealing with common words while the pair words belong to the rare part of the vocabulary. There is, thus, essentially no overlap between these two studies – they complement each other.

Hallberg's pair words method was criticized on various methodological grounds by Marina Meier (1963). In particular, she pointed out Hallberg's lack of statistical sophistication in compensating for the problem that the texts he studied were of various lengths (a problem returned to in Leoni 1970 and Louis-Jensen 2009). Hallberg replied (1964), clarifying his methods and arguing that his simple methods were adequate to the task. In my view, his reply is largely satisfactory – more sophisticated methods are unlikely to yield significantly different results.

In addition to his pair word research, Hallberg also conducted investigations into a few particular words and collocations. Already in his first study, he noted that the verb *kveðask* has an unusually low frequency both in *Heimskringla* and in *Egils saga* (1962, 52–56). This was followed up by various similar observations. One of the traits studied was the frequency of sentence initial *ok er* versus *en er* (both meaning "when") (Hallberg 1963, 10–11; Hallberg 1968, 200–202), where a high frequency of *en er* is the putative Snorri trait. In this case, the stability of the manuscript transmission has been investigated (Haukur Þorgeirsson 2014).

The problem with studies of individual words is that the investigator is open to the charge of opportunism. Why did we choose to focus on *kveðask* and *en er* rather than a hundred other common words or phrases? In comparing any two given texts, it is surely possible to find *some* shared trait if we allow ourselves to pick any convenient target, an issue sometimes referred to as the *multiple comparisons problem*. While this does not negate the value of studying individual traits, it highlights the advantage of methods like Hallberg's pair word method and Burrow's delta in which the investigator proceeds according to a pre-defined plan.

Following in Hallberg's footsteps, Jonna Louis-Jensen has investigated certain stylistic traits in *Heimskringla* and *Egils saga* which may have diachronic implications. She correctly points out that *Óláfs saga helga* and *Egils saga* have a preference for *til þess er* over *þar til er* (both meaning 'until') and a preference for *hitta* over *finna* (in the sense 'meet') (Louis-Jensen, 2009, 108–110). In the remainder of Heimskringla, there is a balance between *til þess er* and *þar til er* and a preference for *finna* over *hitta*. According to Hallberg's diachronic studies, this suggests that "Heimskringla B" is a younger work than *Egils saga* and *Óláfs saga helga*. Louis-Jensen further backs this up with observations on some archaic features in *Egils saga* which are not present in *Heimskringla* (Louis-Jensen, 2013, 142–145).

Louis-Jensen's evidence does suggest that *Óláfs saga helga* may be an older text than the rest of *Heimskringla*. This should not be a surprising result since it has been commonly accepted since Sigurður Nordal 1914 that this was the order in which these texts were written.[4] The style of a given author changes throughout his or her life Stamou (2007) and such changes can reflect broader trends in the language community (Can/Patton, 2004, 2010). Thus, the chronological development which Louis-Jensen points to does not constitute strong evidence against common authorship. As pointed out by Wright (2015, 9–11), all the archaic features she identifies fit comfortably within Snorri's working lifetime.

Conclusions

This article confirms that Burrows' delta is an effective tool for authorship attribution of Icelandic texts, performing well on the test case of 19th century prose fiction. When this test is then applied to *Heimskringla* and the sagas of Icelanders, *Egils saga* turns out to have the lowest delta by a large margin.

---

4 For radically different views on Snorri Sturluson's oeuvre see Sigurjón Páll Ísaksson 2012, 2014 and Berger 2001.

In a more challenging test, *Heimskringla* is compared with *Egils saga*, *Jómsvíkinga saga*, *Morkinskinna*, *Íslendinga saga* and *Knýtlinga saga*. Even here, Egils saga has the lowest delta for both parts of *Heimskringla*. These results constitute substantial support for the theory of common authorship of *Egils saga* and *Heimskringla*.

Literature

Primary Literature

Ármann Jakobsson and Þórður Ingi Guðjónsson (eds.) 2002: *Morkinskinna*. (= Íslenzk fornrit XXIII–XXIV). Reykjavík.

Heimskringla. [http://heimskringla.no]

Netútgáfan. [https://www.snerpa.is/net/index.html]

Rafbókavefurinn – Íslenskar rafbækur í opnum aðgangi. [http://rafbokavefur.is/]

Secondary Literature

Ármann Jakobsson 2002: "Our Norwegian Friend: The Role of Kings in the Family Sagas". In: *Arkiv för nordisk filologi* 117. 145–160.

Berger, Alan J. 2001: "Heimskringla is an abbreviation of Hulda-Hrokkinskinna". In: *Arkiv för nordisk filologi* 116. 65–69.

Bjarni Guðnason 1982: "Formáli." *Danakonunga sǫgur*, v–cxciv. (= Íslenzk fornrit XXXV). Reykjavík.

Burrows, John 2002: "'Delta': A measure of stylistic difference and a guide to likely authorship." In: *Literary and Linguistic Computing* 17. 267–287.

Can, Fazli and Jon M. Patton 2004: "Change of Writing Style with Time". In: *Computers and the Humanities* 38. 61–82.

Can, Fazli and Jon M. Patton 2010: "Change of Word Characteristics in 20th-Century Turkish Literature: A Statistical Analysis". In: *Journal of Quantitative Linguistics* 17 (3). 167–190.

Cormack, Margaret 2001: "*Egils saga*, *Heimskringla*, and the Daughter of Eiríkr blóðøx." In: *alvíssmál* 10. 61–68.

Eder, Maciej and Jan Rybicki 2013: "Do birds of a feather really flock together, or how to choose training samples for authorship attribution". In: *Literary and Linguistic Computing* 28 (2). 229–236.

Elín Bára Magnúsdóttir 2015: *Eyrbyggja saga. Efni og höfundareinkenni*. Reykjavík.

Finlay, Alison 2014: "Jómsvíkinga Saga and Genre". In: *Scripta Islandica* 65. 63–79.

Fotis, Jannidis, Steffen, Pielström, Christof, Schöch and Thorstein, Vitt 2015: "Improving Burrows' Delta – An empirical evaluation of text distance measures". Digital Humanities Conference 2015, Sidney.

Guðrún Nordal 2002: "Egill, Snorri og höfundurinn". *Lesbók Morgunblaðsins*, December 21st 2002. 8–9.

Hallberg, Peter 1962: *Snorri Sturluson och Egils saga Skallagrímssonar: Ett försök till språklig författarbestämning*. Reykjavík.

Hallberg, Peter 1963: *Ólafr Þórðarson hvítaskáld, Knýtlinga saga och Laxdæla saga: Ett försök till språklig författarbestämning*. Reykjavík.

Hallberg, Peter 1964: "Snorri Sturluson och Egils saga Skallagrímssonar: Kommentarer till en recension". In: *Maal og Minne*. 12–20.

Hallberg, Peter 1965: "Om språkliga författarkriterier i isländska sagatexter". In: *Arkiv för nordisk filologi* 80. 157–186.

Hallberg, Peter 1968: *Stilsignalement och författarskap i norrön sagalitteratur: Synpunkter och exempel*. Göteborg.

Haukur Þorgeirsson 2014: "Snorri versus the copyists: An investigation of a stylistic trait in the manuscript traditions of *Egils saga*, *Heimskringla* and the Prose Edda". In: *Saga-Book* 38. 61–74.

Hoover, David L. 2005–2015: *The Delta Spreadsheets*. [Published online.]

Leoni, Federico Albano 1970: "Sagas islandaises et statistique linguistique: Quelques observations". In: *Arkiv för nordisk filologi* 85. 138–162.

Louis-Jensen, Jonna 1997: "Heimskringla – Et værk af Snorri Sturluson?" In: *Nordica Bergensia* 14. 230–243.

Louis-Jensen, Jonna 2009: "Heimskringla og Egils saga – samme forfatter?" In: *Studier i Nordisk 2006–2007*. 103–111.

Louis-Jensen, Jonna 2013: "Dating the Archetype: Eyrbyggja saga and Egils saga Skallagrímssonar." In: Mundal, Else (Ed.): *Dating the Sagas: Reviews and Revisions*. Copenhagen. 133–147.

Males, Mikael 2015: "Er Ólafur Þórðarson höfundur Eglu?" In: *Són* 13. 173–179.

Meier, Marina 1963: "Om et nyt forsøg på at løse Eigla-gåden." In: *Maal og Minne*. 94–101.

Sigurður Nordal 1914: *Om Olaf den helliges saga*. Copenhagen.

Sigurjón Páll Ísaksson 2012: "Höfundur *Morkinskinnu* og *Fagurskinnu*". In: *Gripla* 23. 235–285.

Sigurjón Páll Ísaksson 2014: *Ólafs saga helga og Heimskringla*. Reykjavík.

Stamou, Constantina 2007: "Stylochronometry: Stylistic Development, Sequence of Composition, and Relative Dating." In: *Literary and Linguistic Computing* 23 (2). 181–199.

Vésteinn Ólason 1993: "Íslendingasögur og þættir." In: *Íslensk bókmenntasaga II*. Reykjavík. 23–163.

Vésteinn Ólason 2002: "Inngangur." In: *Ritsafn: Snorri Sturluson I*. Reykjavík. xi–lxxvi.

West, Ralph 1980: "Snorri Sturluson and Egils saga: Statistics of Style." In: *Scandinavian Studies* 52. 163–193.

Wright, Jon 2015: Whose *Edda*? Investigating the textual unity and authorial attribution of the *Prose Edda*. MA thesis at University College London.