Plotting Poetry
https://doi.org/10.7592/PP2025.02.helgadottir
Popular Voices: Computational Analysis of Poetry and Song. ELM Scholarly Press: Tartu 2025.

Textual Variation and Representative Selection of Texts: The Case of the Post-Medieval Icelandic *þulur*

Yelena Sesselja Helgadóttir

The Árni Magnússon Institute for Icelandic Studies, Iceland
sesselja@hi.is
0000-0001-8513-6238

Abstract

This article aims at an objective solution to the methodological question of effective text choice – for purposes of research and scholarly editing—for certain genres of folk poetry where texts are relatively short, but highly variable and at times have no clear boundaries. My focus is on Icelandic post-medieval *þulur*: versified, but not stanzaic, lists of names, short motifs and/or longer narrative episodes in very free poetic form. My hypothesis is that the selection which most adequately represents a useful overview of such texts, while also reflecting their specificity, should be based on the most typically encountered variant(s) of most common structural units of *þulur*, which are identified using quantitative methods. The article offers an algorithm for the text selection and discusses its advantages and shortcomings.

1 Introduction

In this article, I will propose and test a method for effective text choice for relatively short, but highly variable folk poetry—an urgent task for a scholar of Icelandic post-medieval *þulur* (hereafter PMÞ) or kindred poetic genres, which are under-researched due to a shortage of systematic and reliable editions.[1] The

[1] The edition that is usually cited is Jón Árnason and Ólafur Davíðsson 1888–1903; PMÞ texts, sometimes blended with other genres (as is the case in most editions), are in vols. 3: 384–397 & 4: 144–146, 175–221, 234–242, 246, 249–251 and passim (in Icelandic, as are most sources referred to in this note). This edition is rather unreliable (Aðalheiður Guðmundsdóttir 1997; Yelena Sesselja Helgadóttir 2020, pp. 70–72) and should be used with much caution. The few reliable editions containing PMÞ only have a very limited number of PMÞ texts each, as they are primarily editions of ballads and other related poetic genres (Jón Samsonarson 1964, PMÞ in vol. 2: 285–287, 305–314 and passim; Jón Helgason 1962–1981, vol. 8: 112–116) or sampler editions of different genres (Rósa Þorsteinsdóttir 2009, pp. 18, 20, 25, 39, 42–43, 46–47, 55, 64, 72, 74–75, 78, 82–83, 101, 104–106, 108, 112–113, 118). An article by Jón Samsonarson has a reliable, although limited, selection of PMÞ texts (2002, pp. 79–102). I also greatly benefited from careful reading of Jón Samsonarson's uncompleted edition of PMÞ (preserved at the Árni Magnússon Institute for Icelandic Studies). For more details on editions of PMÞ texts see my dissertation (Yelena Sesselja Helgadóttir 2020, pp. 70–74, cf. 67–69 (in Icelandic)).

central challenge is that of finding an algorithm that would yield a manageable, but representative selection that reflects the specificity of the genre in question. No recognised algorithm or publishing strategy exists at the moment for PMÞ and some kindred poetry due to their rich textual variation, which makes even identifying them as poems problematic.

The research is based on manuscript sources from the 17th to 20th centuries and on 20th century audio sources. Both types largely originate in the two most prominent folklore collection campaigns in Iceland. The former campaign, 1845 to ca. 1870,[2] resulted in over 50 manuscripts, each with the number of PMÞ texts ranging from one to several hundred (in large compendia such as DFS 67). The latter campaign, ca. 1960 to 1990,[3] yielded ca. 2,000 audio recorded PMÞ texts, several manuscript compendia (up to a hundred PMÞ texts each) and over 850 texts received in response to a questionnaire from 1984.[4] The total number of known PMÞ texts is about 4,500, a large majority of them unedited and undigitised. For this study, a corpus of about 2,500 PMÞ texts (from every kind of sources) was selected.

In Section 2, I will further introduce my project, its subject matter—PMÞ—and methodological challenges. Section 3 describes the proposed method and its distinct stages. Section 4 discusses some methodological problems and first results of the project, while Section 5 offers preliminary conclusions.

2 Icelandic Post-Medieval *þulur* and the Methodology for Their Analysis

Icelandic post-medieval þulur (PMÞ) are a form of folk poetry practiced from ca. the 15th to 20th centuries, 25 lines on average (ranging from 7 to over 100 lines),[5] characterised by:

1. A quite free poetic form, governed by much more flexible rules than Icelandic post-medieval poetry in general. It was analysed in some detail in my dissertation *Íslenskar þulur síðari alda* (Icelandic Post-Medieval *þulur*) (Yelena Sesselja Helgadóttir 2020, pp. 90–141 (in Icelandic; English

[2] Although the campaign was formally announced in 1845, in was never formally closed; rather, it became considerably less active around 1870. This explains my "ca." before the date.

[3] This campaign was hardly ever announced formally. Therefore, "ca." here covers both its end and its beginning.

[4] Audio archives are available at *Ísmús: íslenskur músík- og menningararfur* (https://www.ismus.is; the page is only in Icelandic) along with rich metadata; however, only few texts have been published. A (draft) transcript of the questionnaire and of a large number of responses is available at: *Sarpur: Menningarsögulegt gagnasafn* (https://www.sarpur.is/Spurningaskra.aspx?ID=531300; the page is only in Icelandic). Most manuscripts with PMÞ texts that are available online (information and/or photos) can be found at *Handrit.is* (https://handrit.is/?lang=en) and *NorS Sprogsamlinger* (https://sprogsamlinger.ku.dk/q.php?p=ds/hjem, under "Arnamagnæansk"; in Danish only).

[5] The minimal number of lines for PMÞ is defined as 7 in my dissertation (Yelena Sesselja Helgadóttir 2020, pp. 92–94), taking into consideration that this is the most prominent borderline between short forms (such as single stanzas) and longer poems (where PMÞ belong) in Icelandic post-medieval poetry. Occasional shorter PMÞ-like texts are most often PMÞ fragments and are treated as such.

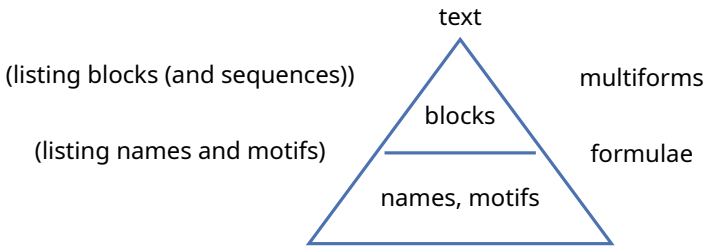

Figure 1: The two-level structure of PMÞ

Summary on pp. 283–285)); see also Yelena Sesselja Helgadóttir 2016 (in English).

2. A loose and highly variable two-level structure based on lists, as I have previously shown (Yelena Sesselja Helgadóttir 2020, pp. 144–249). The lists at the first level consist mainly of names (proper and common) and motifs which combine in various ways to produce relatively stable second-level units that I have designated as "blocks" (Figure 1), while blocks sometimes combine into larger and less stable units of the same level, called "block sequences". PMÞ texts can be made of units from both levels. Formulae and multiforms (in the sense of Frog 2017) are additional, cross-level compositional features of PMÞ supporting the structure (Figure 1; Yelena Sesselja Helgadóttir 2020, pp. 192–195, 218–224, cf. 2016).

3. A world-descriptive and world-creative function (Yelena Sesselja Helgadóttir 2020, pp. 261–268), rather than a mode of narrative or lyrical expression.

4. A strong relationship with, on the one hand, Old Norse/Icelandic þulur (lists of poetic names, out of which the structure and function of the PMÞ have developed), and on the other, continental Scandinavian post-medieval folk poetry (Yelena Sesselja Helgadóttir 2020, pp. 25–37, 262–264).

An example of a PMÞ text and its structure can be seen in Table 1.[6]

A problem in the study of PMÞ is their instability even at the very basic structural level. In many cases, there is no text (or group of texts) that is stable enough to be labelled a poem. For example, there is hardly any way to tell whether text A, containing four blocks labelled "ABCD", and text B, containing blocks "ABDEF", can be identified as being the same poem. The PMÞ are thus much more variable than, for example, ballads and most other Icelandic folk poems. This combination of variability and instability poses unique challenges with regard to producing a representative text selection. While it has long been standard to publish, for instance, ballads in accordance with type classification

[6] Spelling of quotations is normalised to Modern Icelandic. Punctuation is original. Necessary inserts are supplied in angle brackets. In Table 1, problematic English translations are in square brackets. The multifold meanings of the names here are discussed in (Yelena Sesselja Helgadóttir 2020, pp. 160–165).

Búkonan dillaði börnunum báðum	The housewife lulled both [sic] children	Introductory motif	BLOCK	BLOCK SEQUENCE *Heyrði ég í hamrinum* [I heard in the kliff]
Jórunni Þórunni Ingunni Kingunni Aðalvarði Ormagarði, Eiríki og Sveini,	Jórunn Þórunn Ingunn [Kingunn] Aðalvarður [Ormagarður], Eiríkur and Sveinn,	Name list (formulae-aided)		
mu<n>du fræn<d>urnar þínar[7]	remember your relatives	Connecting motif		
Kambur Skæringsson, Skæringur Brandsson Brandur Björgólfsson Björgólfur Hringsson Hringur Hreiða<rs>son [...] Bjór Brettingsson Brettingur Hakason Haki Óðinsson	[Kambur] son of Skæringur, Skæringur son of Brandur Brandur son of Björgólfur Björgólfur son of Hringur Hringur son of [Hreiðar] [...] [Bjór] son of Brettingur Brettingur son of Haki Haki son of Óðinn	Name list (formulae-aided)	BLOCK	
allra illra trölla faðir í hellri —	the father of all evil trolls in a cave —	Closing motif		

Table 1: PMÞ text (ÍB 816 8vo, 109r–109v) and its structure

systems (such as TSB), no type classification has been established for PMÞ or related Scandinavian rhymes, although this has been attempted (e.g., Nordlander 1886, pp. 7–8).[8]

The purpose of my project, which is eponymous to this article and is supported by the Icelandic Research Fund (grant 228363-051, 2022–2024), is to yield a representative selection of texts that reflects the specificity of the genre in question. The task is quite urgent, as availability of a fairly representative text selection is a pre-requisite for research—and therefore for a genre becoming involved in literary and/or folkloristic research-based discussion. Without such text selection being accessible, the genre in question runs the risk of remaining virtually invisible to many a researcher, even in allied areas of studies, as the case of PMÞ demonstrates.

In order for text selection to be representative and reflect the specificity of the genre in question, it is reasonable to include this specificity into the text selection process from the very start. PMÞ do not appear to have a single "most specific" feature (cf. Yelena Sesselja Helgadóttir 2020, p. 274); however, one of the most specific features of PMÞ—and probably the one that is most related to text selection—is their loose and highly variable two-level structure based on the listing principle, described earlier in this section.

The actual structure of PMÞ texts is certainly not always as clear-cut as the model shown in Figure 1, or even as the example in Table 1. PMÞ texts can contain (fragments of) numerous blocks or even of block sequences (as in the text in AM 277 8vo, 32v or AM 960 XX 4to, 3r), or even of texts belonging to other

[7] Words 2–3 are feminine in this line, but masculine in standard Icelandic.

[8] The question of applicable editorial strategies is not discussed here due to space limits. It may still be noted that eventual publication of the whole text corpus of PMÞ (most probably, online) is a long-run objective of my work, in spite of being very time-consuming. The time factor is one of the reasons for publishing a representative text selection first. Other reasons include the unreliability of the edition that is often used (cf. note 1 on Jón Árnason and Ólafur Davíðsson 1888–1903) and at least some readers' preference for a concise text selection to familiarize themselves with a genre.

genres (as in DFS 67, 256v–257r). This type of folk poetry is everchanging; a block (or a block fragment) may turn up inside another block, compound motifs and motif complexes make their way into PMÞ texts, and further levels appear within the two basic levels—nonetheless, the two basic levels defined above retain their foundational quality.

PMÞ are folk poetry; the methods of classical philology—or even new philology (cf. Wolf 1993)—are scarcely if at all applicable to such texts for the purpose of producing a representative text corpus. Making a stemma is most often impossible; besides, it is meaningless for the purpose of producing a representative text selection in terms of this project, since the oldest and, in this sense, most "original" text of a block is not necessarily representative of PMÞ structure. In a similar way, giving priority to a geographical approach would potentially lead to the listing structure of PMÞ being underrepresented in the resulting selection (besides, Icelandic PMÞ appear to be quite monotypic throughout the country). The same is likely to be true in text selections governed by chronological, topical or other traditional principles. This can be remedied by making the structure central in selecting texts.

In the case of folk poetry with unclear poem boundaries, but chiefly consisting of a limited number of structural units (such as PMÞ blocks) which have clearer boundaries than the poems, a manageable and representative selection can be based on the typology of these units. Thus, PMÞ structure can also serve as a tool for selecting texts. My hypothesis is that the selection which most adequately represents a useful overview of highly variable texts such as the PMÞ, while also reflecting their specificity, should be based on the most typically encountered variant(s) of the most common structural units of PMÞ, primarily PMÞ blocks. Chronological and geographical considerations should then be applied as second and third criteria of text selection. In the case of PMÞ, then, a very minimal representative selection (of probably 50–100 texts) can be based on one text of each of the roughly 50 blocks and their subtypes already identified; each text should contain the most common variant of the respective block. A properly representative selection should also contain the most common variants of other essential building units: motifs,[9] names, and preferably block sequences. Testing my hypothesis and evaluating the quantity of texts representing several building units—and thus possible economy (to result in the minimal number of texts) in the (final) corpus—is among the main aims of my project. Among the evaluation criteria is practicability of the selection process described below, as well as whether the ratio of texts representing several building units turns out to be high and the ratio of motifs, names and block sequences, not represented in the main corpus, proves to be low.

The methodology of this project is in line with Honko's (2000) conventional concept of "textuality" in that it prioritises meaning of the text and its connection with variation over formal aspects of variation or geographical and temporal factors. The purpose of this particular project, on the other hand, is not to create

[9] Motifs are defined in my dissertation as simple, repeated descriptions of various kinds of events, actions, conditions, and so on (Yelena Sesselja Helgadóttir 2020, pp. 179–180, in line with Jason 2007; cf. Gasparov 1997, p. 14; Neklyudov 2004; Thompson 1946, pp. 415–416). "Essential motifs" are those that occur in at least ¾ of all texts containing the respective block, as well as those that change further course of PMÞ-text.

a thick corpus of folk texts (in the spirit of Honko's ideas) but rather a rigorously limited text selection. Such selection will, nonetheless, prioritise meaning over time and place and have the potential to become the necessary foundation for a more extensive scholarly publication of PMÞ texts. This study, which focuses on the "typical" in variation and meaning, is a prerequisite for exploring less common variation and its potential connections to the location of texts and their cultural context.

3 Procedure

Section 3.1 describes the procedure of frequency analysis applied in order to ascertain the most common variants of the building units of PMÞ. Employed together with a comparison of texts, frequency analysis allows one to estimate which kinds of variation occur most typically and which variants are most representative. These methods are used for both levels of the structure model of the PMÞ as described in Section 3.2–3.3. Updating and refining the selection is addressed in Section 3.4–3.5.

3.1 Frequency Analysis

The most typically encountered variants of building units of PMÞ are identified through frequency analysis. This includes manually encoding the building unit in question throughout the PMÞ text corpus under consideration, allocating each sub-unit or its variant a letter. For instance, encoding a block involves labelling each motif and name or name combination; the first block in Table 1 would thus be described as "I 23: dfjioe" (where *I 23* is a unique text number). Sorting and counting—that can easily be done in, for example, Excel—reveals the most common structure for the building unit in question (for our block, it is "abcdfijoe", occurring in roughly 50% of the texts that include the block). It is also helpful for ascertaining rare and unique variants of the building unit in question.

It is important that sub-unit variants are generally not taken into account in the analysis of units, as this will be done at a different stage of the project (Section 3.3). Thus, when blocks are analysed, only major variants of motifs (and more seldom, of names) are considered: e.g., those dividing blocks into different types, or those that correlate with certain blocks (or other building units) preceding or following the block in question. An example is in the most frequent PMÞ block named *Sat ég undir fiskihlaða* (I sat under <a> fish stack). Close to the end, there is the motif *Sankti María gaf mér sauð* (<the> Virgin Mary gave me a ram), accompanied by either *það varð mér að miklum auð* (it turned out to be big wealth for me) (as in Holm. Papp. 64 fol., 99v) or *síðan lá hún steindauð* (then she lay stone-dead) (sic) (as in DFS 67, 250r). This is a significant distinction, correlating with different motifs in the beginning of the block—and, even more important, with different subsequent blocks. If *Sat ég undir fiskihlaða* has the motif *það varð mér að miklum auð* (allocated the letter x), it is usually followed by the block *Sofa, sofa hjónakornin bæði* (<They>

sleep, sleep, both husband and wife); but if *Sat ég undir fiskihlaða* has *síðan lá hún steindauð* (w), then *Kona mín í kofanum* (My wife in the hut) is usually the next block in the sequence (cf. Figure 4, Section 4.1). Both *það varð mér að miklum auð* (x) and *síðan lá hún steindauð* (w) qualify as motifs; however, they do not occur in PMÞ independently of *Sankti María gaf mér sauð* (v), which gives a reason to address x and w as variants in the compound motif "vx/w" or motif complex "vx/vw". This is a case when motif variants would be taken into account while analysing a block. On the other hand, variants like *hún var stokk- og steindauð* (she was [tree-]stock- and stone-dead) (as in AM 969 4to, 46v), although significant in their way as well, will be considered at later stages (Section 3.3–3.4).

3.2 Step 1: Identifying Most Common Variants of PMÞ Blocks

In Step 1, the whole corpus of PMÞ under consideration is reviewed in order to identify the texts containing the most typical textual variant(s) of each of the roughly 50 most frequently occurring blocks in PMÞ. This will result in a core selection, representing each of the block (sub)types ("Corpus A"). The procedure is as follows.

Starting with (I) the whole corpus of PMÞ texts, and (II) the list of 50+ most frequent blocks types and subtypes (cf. Yelena Sesselja Helgadóttir 2020, pp. 314–395; the blocks are additionally sorted by frequency):

1. If Block 1 on the 50+ list has no subtypes, proceed to (3).
2. If Block 1 has subtypes, identify the most frequently occurring subtype of Block 1, designating it as Block 1:1.
3. Identify all PMÞ texts containing Block 1 / Block 1:1.
4. Identify the most frequent variant of Block 1 / Block 1:1.
5. Identify the PMÞ texts containing this variant of Block 1 / Block 1:1.
6. Identify the oldest of these texts (the chronological aspect is given consideration here).
7. Tag that oldest text—containing the most frequent variant of Block 1 / Block 1:1—as part of Corpus A, representing there Block 1 / Block 1:1.
8. Tag all other texts, containing the most frequent variant of Block 1 / Block 1:1, as part of "Corpus A*" (needed for a later stage, see Section 3.4), representing there Block 1 / Block 1:1.
9. If Block 1 has other subtypes, return to (2) and proceed with Block 1:2 (1:3, etc.) until all subtypes of Block 1 have been analysed.
10. If Block 1 has no other subtypes, return to (1) and proceed with Block 2 (3, etc.) until all blocks on the 50+ list have been analysed.

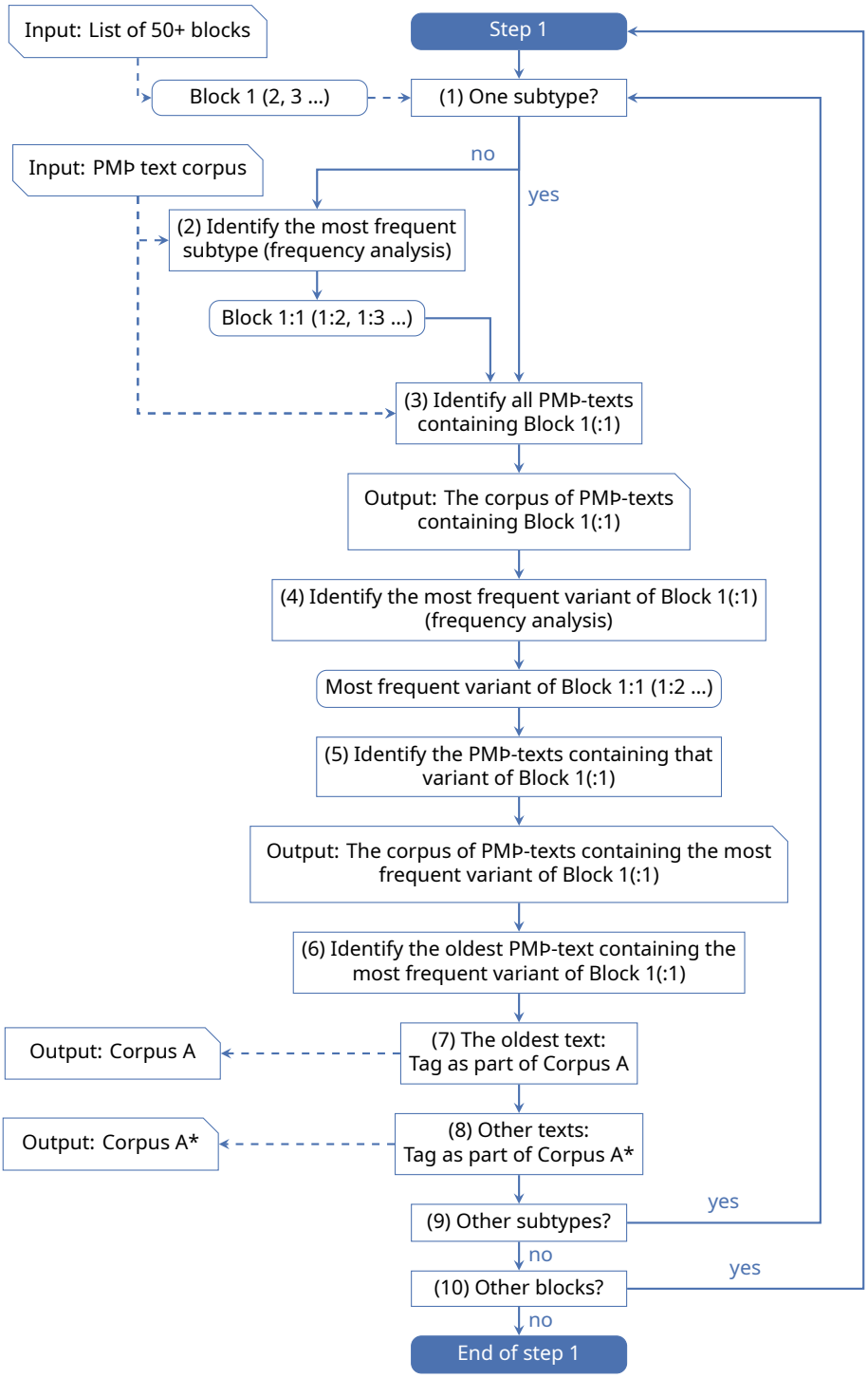

Figure 2: Flowchart for Step 1: Selecting texts representing block (sub)types

The procedure described here can also be presented as a flowchart (Figure 2), where simple connectors show the input data for next operation cell, while arrows can both point at the next operation and show input data.

Upon completing this procedure for all the blocks on the 50+ list, the texts in Corpus A are reviewed. Texts tagged as representing more than one block are identified and their ratio in Corpus A is determined, as one of the aims of the project is evaluating possible economy in the number of texts in the (final) corpus.

In addition to the core selection of oldest texts, representing each of the block (sub)types (Corpus A), this analysis identifies texts–other than the oldest ones—also representing their respective blocks (Corpus A*).

3.3 Steps 2–4: Identifying Most Common Variants of Essential Motifs, Names, and Block Sequences

In Steps 2–4, similar procedures are used for identifying the most frequent variants for other building units of PMÞ: essential motifs and combinations of names, on the lower level, and block sequences, on the higher level. For each of these units, the whole corpus of PMÞ under consideration is reviewed in order to identify the texts containing the most typical textual variant(s) of each of the units listed above. This will result in three core selections, representing each of the aforementioned (sub)types of building units: Corpus Bn (*n* stands for *names*), Corpus Bm (*m* = *motifs*) and Corpus Bs (*s* = *sequences*), as well as in three additional corpora containing other representative texts than the oldest ones: Corpus Bn*, Corpus Bm*, and Corpus Bs*. Then each corpus is reviewed; texts, tagged as representing more than one respective building unit, are identified and their ratio in the respective corpus determined. Texts that are also tagged as "Corpus A" are identified and their ratio in the corpora Bn, Bm, and Bs is determined.

3.4 Steps 5–7: Updating Corpus A

In Steps 5–7, the output of Steps 2–4 (Corpus Bn, Corpus Bm, and Corpus Bs) is compared to Corpus A, and Corpus A is updated with non-repetitive texts from the corpora Bm, Bn and Bs in order to maximise the number of essential motifs, name lists and block sequences represented in the corpus that is sought for. The procedure is as follows.

Starting with Corpus A and Corpus Bn:

1. If Text 1 (T1), representing Name 1(:1) or Names' Combination (NC) 1(:1) in Corpus Bn (i.e., the oldest text containing the most common variant of Name 1 / NC 1 or its most common subtype), also belongs to Corpus A, tag T1 as representing Name 1(:1) / NC 1(:1) and proceed to (7).

2. If Corpus A does not have T1, but has another text, containing the most frequently occurring variant of Name 1(:1) / NC 1(:1), tag the respective text in Corpus A as also representing Name 1(:1) / NC 1(:1) but not being its oldest representative. Proceed to (7).

3. If Corpus A has no text representing Name 1(:1) / NC 1(:1), then examine whether T1 also belongs to Corpus A* as a representative (albeit not oldest one) of any block (Block X). If not, tag Name 1(:1) / NC 1(:1) in Bn as unrepresented (adding it to Corpus A can still be considered at a later stage, cf. Section 3.5). Proceed to (7).

4. If T1 belongs to Corpus A*, examine whether T1 is inferior as a representative of Block X (structure-wise) than the text currently representing Block X in A. If T1 is inferior, tag Name 1(:1) / NC 1(:1) in Bn as unrepresented (cf. Section 3.5). Proceed to (7).

5. If T1 is not inferior structure-wise, the two texts are compared and factors such as their age and whether they represent other building units are evaluated. Based on this analysis, and provided that moving Block X's current representative to Corpus A* will not affect other units possibly represented by the text in question, T1 might be tagged as belonging to Corpus A—and the other text, as belonging to Corpus A* (with tag "A*<A" + represented building unit(s) ID). Proceed to (7).

6. If T1 is not tagged as belonging to Corpus A, tag Name 1(:1) / NC 1(:1) in Bn as unrepresented (cf. Section 3.5). Proceed to (7).

7. Return to (1) and proceed with Text 2 (3, etc.) until all names / NCs from Bn and their respective subtypes have been analysed.

For a visual presentation of the procedure above, see Figure 3.

Upon completing the above procedure, the ratio of names tagged as "UN" (i.e., unrepresented in Corpus A) is determined. The same procedure is then used to update Corpus A with relevant texts from the corpora Bm and Bs in Steps 6–7, ensuring that as few building units as possible are left unrepresented. Depending on time schedule of the project, repeating the procedure for units other than blocks (i.e., working with the corpora Bn*, Bm*, and Bs* instead of Corpus A* in Figure 3) can be considered.

3.5 Steps 8–9: Refining Corpus A

Up to Step 8, we have only been adding to Corpus A. In Step 8, the updated Corpus A is refined in an effort to decrease the number of texts without losing the representativity of the corpus. The procedure is as follows.

1. If Text 1 in Corpus A has more than one block, examine whether T1 is also in Corpus A* as a representative (not oldest) of its 2nd (3rd, etc.) block. If it is not, proceed to (5).

2. If it is, examine whether T1 is inferior as a representative of its Block 2 (structure-wise) than the text currently representing Block 2 in Corpus A. If it is, proceed to (5).

3. If T1 is not inferior, then T1 is compared with the text currently representing Block 2 in Corpus A (as in operation (5), Section 3.4).

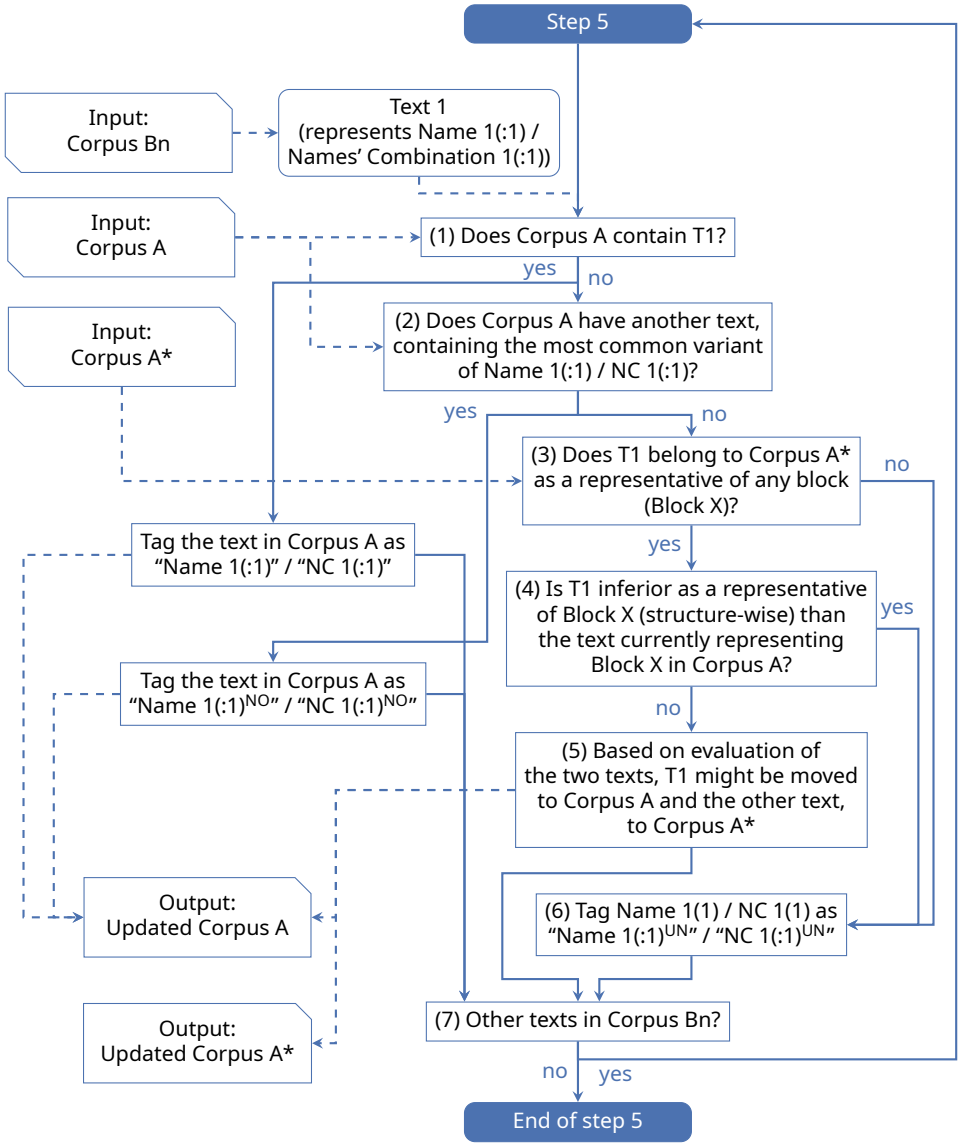

Figure 3: Flowchart for Step 5: Updating Corpus A from Corpus Bn

4. Based on this comparison, and provided that moving the current representative of Block 2 to Corpus A* will not affect other units that it possibly represents, T1 might be tagged as also representing Block 2 in Corpus A, while the other text is tagged as belonging to Corpus A* (i.e., "A*<A" + represented building unit's).

5. Return to (1) and proceed with next Block 3 (4, etc.) or Text 2 (3, etc.).

Upon completing the above procedure, the ratio of texts that were removed from Corpus A is determined with regard to the question of potential economy in the number of texts.

Depending on time schedule of the project, repeating the procedure for units other than blocks (i.e., working with the B-corpora instead of A-corpora) can be considered. In a project with a tight time schedule, examining only those texts in Corpus A that are already tagged as not oldest representatives of other units and those texts in B-corpora that are tagged as unrepresented can be considered. On the other hand, if time permits, it may be useful to review all texts in Corpus A* and B-corpora, with the view of substituting (occasional) less representative texts in Corpus A, provided that this would not affect other units represented by the texts in Corpus A.

In Step 9, the resulting Corpus A is studied for its representativity in terms of factors other than the structure of PMÞ, and the corpus is further refined. The questions are whether Corpus A is balanced in terms of the texts'

1. age: before 1900 vs. after 1900,

2. provenance: how well different regions of Iceland are represented,

3. source type: manuscript vs. voice recording,

4. folklore collection: how equally major collections are represented.

If the corpus is not balanced, ways to remedy this are considered, such as adding extra texts to Corpus A or substituting texts in Corpus A with other texts (for example, from Corpus A*) that are structure-wise not inferior as representatives for the block and other building units in question than the text to be replaced.

If the ratio of PMÞ block sequences, motifs, and names/NCs that still remain unrepresented in Corpus A is high, the ways to remedy this are considered. It is important to determine how many texts would then have to be added to Corpus A (noting that texts in B-corpora potentially represent two or more motifs that are yet unrepresented in A).

Finally, with time and size of Corpus A permitting, those texts that are neither in A- nor B-corpora ("Corpus C"), can be examined for rare variants in order to increase representativity of the final Corpus A.

4 Discussion

4.1 Methodological Considerations

In this section I will concentrate on the process of text selection, largely leaving aside the numerous problems of the preceding analysis and the encoding of each PMÞ text's structure, such as finding the subtle line between a compound motif and a complex of motifs, a complex of motifs and a small block, etc. in everchanging folklore texts (cf. Yelena Sesselja Helgadóttir 2020, pp. 179–185, 203–211).

Selecting the most representative text(s) of a block is seldom problematic, even when the block has numerous subtypes. For instance, *Sat ég undir fiskihlaða* (I sat under <a> fish stack), mentioned in Section 3.1, has four subtypes (cf. Figure 4), as well as "Ættartala" ("Genealogical list") (the second block in Table 1, Section 2) and *Ekki heiti ég Eiríkur* (My name is not Eiríkur), which often precedes "Ættartala" (however, not in Table 1). Rather, a problem can arise in blocks with minor and somewhat dispersed variation, not splitting easily into subtypes, such as *Kom ég þar að kvöldi* (I came there in <the> evening) or *Konan mín í kofanum* (My wife in the hut) (closer considered later in this section). It is debatable which would be more in keeping with the spirit of the method: using even minor textual differences to divide the block into subtypes, or considering virtually all its texts—apart from those containing very specific variants—as representing the block in question, although its texts would still be quite different. Following the working rule that sub-unit variants are generally not taken into account in the analysis of units (but considered at a different stage of the project), I am inclined to view all texts as representatives of the respective block if all the block's motifs/names are in place and in absence of clear distinctive features.

Most other cases that pose problems appear to be due to difficulties in discerning between (sub)types of building units or between two or more close types of units. On the block level—the higher structural level of PMÞ—which I am currently working on, a block participating in two (or more) different subtypes of the same block sequence can be problematic, as can an independent block which sometimes is a part of a block sequence and sometimes not. Examples of both cases are in the sequence *Sat ég undir fiskihlaða* (S in Figure 4). The block *Kona mín í kofanum* (C) is sometimes a part of the block sequence subtype "SC", but sometimes part of another—far less numerous—subtype of the same sequence, together with three short blocks (all or some of them) that I designated for simplicity as "I", "Í" and "J": "SCIÍJ". As both these combinations start with "SC", it is not obvious whether these are two distinctive subtypes of the sequence or whether they should be viewed as one subtype (and its subsubtype)—and therefore whether the text corpus underlying the analysis of the subtype "SC" should include texts containing "SCIÍJ" (cf. particularly operations (2–3), described in Section 3.2, and Step (4), Section 3.3). Needless to say, the most frequently occurring variant—and thus the whole result—could depend on this.

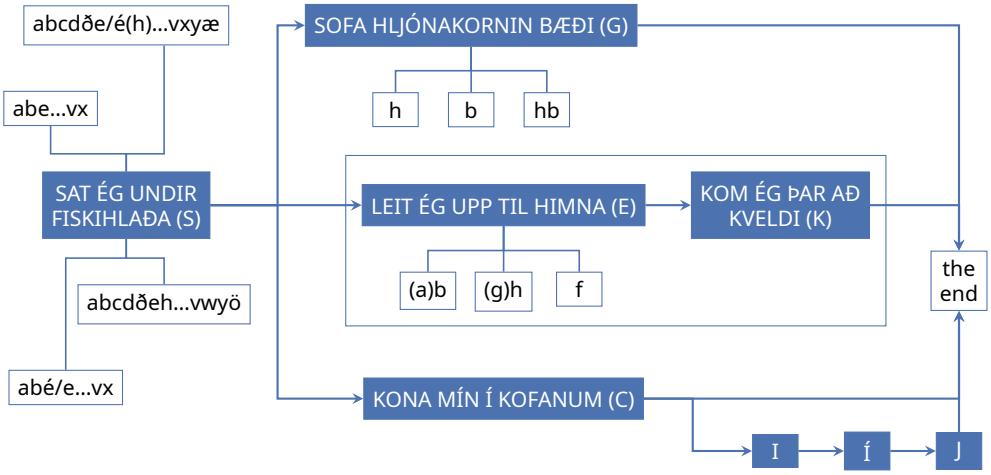

Figure 4: The structure of the *Sat ég undir fiskihlaða* sequence

The block *Kom ég þar að kvöldi* (K) is found in PMÞ texts both independently and as a part of several different sequences—among others, *Leit ég upp til himna* [I looked up to the skies] (E) (Figure 4). Besides, it is often the closing block in PMÞ texts starting with *Sat ég undir fiskihlaða* (usually with *Leit ég upp til himna* in between, but not always). Among several ways to handle this variability is viewing "S+(E)+K" as a subtype of the sequence *Sat ég undir fiskihlaða* ("SEK") or as a combination of block sequences, "S EK" (as in Yelena Sesselja Helgadóttir 2020, p. 392). The "S EK" way appears to be preferable, as it helps to avoid interpreting "S+(E)+K" as a very rare—probably unique—case of an independent block sequence which can also be a part of another block sequence.[10] But then, the analysis of the sequence *Sat ég undir fiskihlaða* lacks all texts containing K and E, as technically they belong to a different sequence,—albeit very closely related in oral tradition. Therefore, the "SEK" possibility deserves consideration, although it involves structural units of the same type tucked into each other—which is an issue that a position should be taken on: while exceptional on the level of blocks and sequences, this issue is almost usual on the lower level of PMÞ structure.

4.2 First Results on Representativity of Structural Units and Economy in the Number of Texts

When blocks that belong to the two largest sequences of the whole corpus of PMÞ have been analysed, there are reasons to be optimistic about possible economy in the number of texts in the final corpus. At least on the higher level of PMÞ structure, a PMÞ text, composed of several blocks, often appears to be representative of more than one block and/or (sub)type of the respective sequence. The oldest representative texts of the blocks—and apparently of their

[10] Cases like "S K"—which are not very numerous—can be interpreted similarly, but with E missing (although it is unusual for the head block of a sequence to be missing).

subtypes as well—also represent the main subtypes of the sequences, which makes unnecessary updating Corpus A with texts representing the sequences (cf. Step 7, Section 3.4).

Texts that are the oldest representatives of both their respective sequences' (sub)types and of each of the sequence's blocks are quite rare. More often, a text which is representative of all the units above is only the oldest representative for some of them. If the texts' age is disregarded, most of the blocks in the two sequences in question have a text that is also representative of the respective sequence (sub)type *and* of each block in the sequence. It could therefore be considered to disregard the texts' age, at the very least in those cases when the time difference is small—which is often the case, as the majority of the texts comes from the few folk poetry collection campaigns in Iceland.

When block subtypes are taken into consideration, they do not necessarily have a matching representative text in *each* subtype of other blocks (even if the texts' age is disregarded). However, in ca. two-thirds of cases there will be a text representing one (sub)type of each of the blocks and one (sub)type of the sequence, which is more than was expected.

Further research will tell whether these results will hold for all blocks and sequences on the higher level of PMÞ structure and how they will interact with its lower level.

5 Conclusion

In spite of some controversy (Section 4.1), testing the text selection method based on the typology of building units of PMÞ (described in Section 2–3) has so far shown that it is applicable and useful for selecting PMÞ texts representative of the structure of the genre. The method has few drawbacks as far as the higher level of PMÞ structure is being analysed.

A text that is representative of the sequence subtype and of each of its blocks is only in some cases the oldest of the texts that represent the units in question, and if the texts' age is regarded as crucial, it can be difficult to achieve substantial economy in the number of texts in Corpus A. On the other hand, first results show that if the requirements on the age of texts are reasonably relaxed—as can be easily justified—then there are texts representative both of the sequence subtype and of each of its blocks in a large majority of cases. In such cases, structure could be systematically given priority over chronology to achieve economy in the number of texts in Corpus A without losing the corpus' representativity structure-wise. Preliminary results thus show not only that the suggested method is practicable, but also that the economy in the number of texts is considerable and the ratio of block sequences, not represented in the main corpus, is fairly low, therefore supporting my hypothesis.

Acknowledgments

This article is a part of a project supported by the Icelandic Research Fund (grant 228363-051). I am grateful to Margaret Cormack and Ryder C. Patzuk-Russell, who read a version of this paper.

Manuscripts

The Árni Magnússon Institute for Icelandic Studies, Reykjavík

- AM 960 XX 4to
- AM 969 4to
- AM 277 8vo

National and University Library, Reykjavík

- ÍB 816 8vo

Royal Danish Library, Copenhagen

- DFS 67

Royal Swedish Library, Stockholm

- Holm. Papp. 64 fol.

References

Aðalheiður Guðmundsdóttir (1997). “(Ó)Traustar heimildir: Um söfnun og útgáfu þjóðkvæða”. In: *Skáldskaparmál* 4, pp. 210–226.

Frog (2017). “Formulaic Language and Linguistic Multiforms: Questions of Complexity and Variation”. In: *Formula: Units of Speech, ‘Words’ of Verbal Art*. Ed. by Frog. Helsinki: University of Helsinki, pp. 252–270.

Gasparov, Mixail Leonovich (1997). “«Snova tuchi nado mnoyu…» Metodika analiza”. In: *Izbrannye trudy*. Vol. 2. Moskva: Yazyki russkoj kul’tury, pp. 9–20.

Honko, Lauri (2000). “Thick Corpus and Organic Variation: An Introduction”. In: *Thick Corpus, Organic Variation and Textuality in Oral Tradition*. Ed. by Lauri Honko. Helsinki: Finnish Literature Society, pp. 3–28.

Jason, Heda (2007). “About ‘Motifs’, ‘Motives’, ‘Motuses’, ‘-Etic/s’, ‘-Emic/s’, and ‘Allo/s-’, and How They Fit Together”. In: *Fabula* 48, pp. 85–99. DOI: 10.1515/FABL.2007.008.

Jón Árnason and Ólafur Davíðsson, eds. (1888–1903). *Íslenzkar gátur, skemtanir, vikivakar og þulur*. 4 vols. Kaupmannahöfn: Hið íslenzka bókmenntafélag. URL: https://baekur.is/bok/09823197-c2c7-4436-9413-5335cd371cba.

Jón Helgason, ed. (1962–1981). *Íslenzk fornkvæði. Islandske folkeviser*. 8 vols. København: Munksgaard, Reitzel.

Jón Samsonarson, ed. (1964). *Kvæði og dansleikir*. 2 vols. Reykjavík: Almenna bókafélagið.
Jón Samsonarson (2002). “Barnagælur”. In: *Ljóðmál: Fornir þjóðlífsþættir*. Ed. by Einar G. Pétursson, Guðrún Ása Grímsdóttir, and Vésteinn Ólason. Reykjavík: Stofnun Árna Magnússonar á Íslandi, pp. 75–149.
Neklyudov, Sergej Yur’evich (2004). “Motiv i tekst”. In: *Yazyk kul’tury: Semantika i grammatika*. Ed. by Svetlana Mixajlovna Tolstaya. Moskva: Indrik, pp. 236–247.
Nordlander, Johann, ed. (1886). *Svenska barnvisor ock barnrim*. Stockholm: n.p.
Rósa Þorsteinsdóttir, ed. (2009). *Einu sinni átti ég gott*. 2nd revised. Reykjavík: Smekkleysa / Stofnun Árna Magnússonar í íslenskum fræðum.
Thompson, Stith (1946). *The Folktale*. New York: Dryden Press.
Wolf, Kirsten (1993). “Old Norse – New Philology”. In: *Scandinavian Studies* 65.3, pp. 338–348.
Yelena Sesselja Helgadóttir (2016). “Formulaic Language in Minimal Metrical Requirements: The Case of Post-Medieval Icelandic Þulur”. In: *The Ecology of Metre: A Special Issue of RMN Newsletter*. Ed. by Ilya Sverdlov and Frog. Vol. 11. Helsinki: University of Helsinki, pp. 49–61. URL: https://www.helsinki.fi/assets/drupal/2022-12/rmn_11_2015-2016.pdf.
Yelena Sesselja Helgadóttir (2020). “Íslenskar þulur síðari alda”. ISSN: 978-9935-9491-9-6. PhD thesis. University of Iceland. URL: https://hdl.handle.net/20.500.11815/1939.