Epic Formulas and Intertextuality in 16th Century Hungarian Historical or Epic Songs

Levente Seláf, Villő Vigyikán, Petr Plecháč, Margit Kiss

1. Introduction

1.1. Characteristics of the 16th century Hungarian epic poetry

The first great period of Hungarian literature is the 16th century. From earlier times only a very limited number of texts, and even less poems, have been conserved: a real literary tradition seems to emerge only shortly before the devastating battle of Mohács, causing the collapse of the powerful medieval Hungarian Kingdom.

In the literary landscape of 16th century Hungary the vernacular production in Hungarian is still somewhat weak, but a wider range of poetical texts have been composed during this period. One of the most important, and (in scientific discourse at least) the most popular poetical genre of the century is the epic or historical song, called "históriás ének". Of a total of 1523 poems (conserved, or lost but at least known by mentions) from before 1600 no less than 186 belong to this genre. The text of 173 such historical songs (and some minor fragments) are conserved (25137 strophes, 99060 lines, 529455 tokens). Although the genre is above all attached to the telling of contemporary warlike events, several other topics have been treated in that form: the plot might have been historical or fictitious, antique, medieval or contemporary in inspiration, related to Hungarian or biblical history, or treating even other European political events not related to Hungary. In a current research project we are exploring the corpus' different metrical and poetical patterns, characteristics and the compositional technique of the poets with computational tools.[1]

1.2. Oral v. written patterns

The main peculiarity of the corpus is its double, oral and written character. We know from several (inner and outer) sources that Hungarians were pleased to sing these compositions on different occasions, festivities (Oláh, Galeotto Marzio, Sidney) (Oláh 2000, p. 58; Seláf 2020). The poems have several references to an oral presentation. All of them were written to tunes, and most of the melodies are known from contemporary songbooks and were used for singing lyrical songs as well. These patterns show that epic songs could be orally performed. But there is no precise description of any of these performances. And we know from the study of Ancient and Renaissance epics that

[1] The current members of the group are, beside the authors of this article, Szilvia Maróthy, Eszter Simon, Artjoms Šeļa, Mária Finta and Boglárka Pardi, among others. The authors warmly thank all former and current participants of the research project for their contribution to this article.

rhetorical figures related to orality might be simply tributes to the Latin learned literary tradition, without any reference to reality. And most of the conserved poems themselves also contain traces of a written, fixed form of writing. The most important of them is the acrostic: many of the historical songs include a Hungarian or Latin text if we read together only the first letters of each strophe. It is also common that the conditions of composing the poem are recorded in the closing stanza of the text, which is also clearly a pattern of writing.

The presence of oral and written patterns in the historical songs shows the genre as a very specific mixture of a lost, traditional, sung and performative poetic tradition, and a learned, rather sophisticated, written epic poetry, influenced by Ancient models (like the *Aeneid*). The heterogeneity of the genre is perceivable also by the different levels of the two kinds of patterns in the poems; in some cases the rhetorical structure of a historical song reflects a prominently oral character. This is most obvious when a poem is full of epic formulas: roughly verbal clichés employed several times, generally in the same metrical position, and sometimes occurring also in other poems of the corpus.

1.3. The formula in other literary traditions

The notions of epic formula and formula-system were created by Milman Parry, who proved their importance in Homeric epic (Parry 1928). The search for formulas was enlarged to other prominent literary traditions (French chanson de geste, Middle English epic, Italian cantari, the *Poema del mio Cid*, etc.), and their importance was also proven in oral composition of folk poetry in the Balkan peninsula (in the Serbo-Croatian heroic ballads) (De Chasca 1970; Kay 1983; Seláf 2020; Tatlock 1923).

We consider that it is necessary to provide a definition specific of the epic formula to each literary tradition, including the Hungarian historical songs. This corpus has some special features that must have influenced the use of formulas. The poorness and monotony of rhyming acercates this corpus to the chansons de geste written in laisses. But while the ancient Greek epic is using mostly hexameters, and the French chansons de geste only decasyllabic lines, and rarely the alexandrins, in Hungarian historical songs there is a higher variety of meters (with 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19, 23, and even 25 syllabic lines), although the most common is the 11-syllabic verse. This metrical variability preconises a somewhat higher variability of the formulas as well, just like in Middle English.[2]

[2] Windelberg and Miller refer to Fry who "in addition to permitting variation in both lexical classes and syntactic-semantic structure, ... also admits variation in the metrical pattern" (Windelberg and Miller 1980, p. 32).

1.4. Definitions of the epic formula in Hungarian literary studies

Since the second half of the 20th century several analyses succeeded to identify formulaic expressions in the corpus of 16th century Hungarian poetry, and some authors tried to give a definition of the epic formula fitting to this particular poetic tradition. In lack of computation tools it was not evident to have a look on the entire corpus.

In the Hungarian scholarly tradition Béla Varjas was the first to search for epic formulas in the historical songs. He claimed that formulas were applied mostly in epic songs, but not exclusively: they could appear in all versified genres, also in lyrical poetry (Varjas 1982, p. 202). He extracted several formulas from the curious epic song entitled *Cantio de militibus pulchra* (*Nice song of soldiers,* RPHA 0369). Varjas also tried to identify the formulas of the *Cantio* in other historical songs of the period. He supposed that the anonymous author of the *Cantio* has constructed his poem of elements borrowed from a contemporary common treasury of formulas and poetical tools. Varjas was searching for the "word-groups created according to the grammatical rules of versification, and used regularly to express specific thoughts" (1982, p. 202). He considered as formulas not only the lexical, but also the grammatical, structural and compositional repetitions. But on the other hand he did not take into consideration the metrical position of the formulas.

Varjas also made a catalog of the epic formulas found in the *Cantio* and in one or more of the dozens of historical songs he read through. In a longer excerpt of Tinódi (16 verses), he identified formulas in every single line, and some of them occur also, in a somewhat altered way, in the *Cantio de militibus pulchra*, that uses very similar sentences in the description of a battle (1982, p. 206–207). The number of inner formulas or repetitions is also very high inside the poems. In the next picture we highlighted some of the formulaic lines, and indicated with the same color the lines having a similar or absolutely identical lexical content [Figure 1].

Tinódi (1552) - RPHA 1334	**Gyulai ének (1561) - RPHA 0369**
Magyarok (N) es (C) álló (P) seregök (N) meghagyák (V), Az (A) szárnya (N) mindkét (Pr) fél (N) öszveroppanának (V), Nagy (Ad) erős (Adj) viadalt (N) akkoron (Ad) tartának (V), Ott jeles terekek nagy sokan elhullának. Nagy sok dob, trombiták oly igen harsagnak, Nagy rettenetösen üvöltnek, kiáltnak, Nagy szép festett lovak az mezőn jargalnak, Kikről fő terekek elestek, megholtanak. Az viadal köztök ám ott meglassódék, Álló (P) seregekből (N) két fél takarodék, Jézust (N), Allát (N) meg (VP) másodszor (NN) üvöltének (V). Mindkét felől hamar taraszkokból lüvének.	Két (NN) fél (N) öszve (VP) erősen (Ad) roppanának (V) Az (A) törökök (N) mind (Pr) Allát (N) kiáltának (V), Az (A) magyarok (N) mind (Pr) Jezust (N) kiáltának (V). Két (NN) fél (N) öszve (VP) erősen (Ad) roppanának (V), Reg/gel/tül fogva mind estvéig vívának, Szegény (Adj) magyarok/ban/ (N) sokan (NN) megholtanak (V). Törökök előtt magyarok futamának, Magyarokat messze nem űzték vala, Nap immáron hogy alámegyen vala. Falka barmot távoly földen látának, Magyar seregnek azt alítják vala, Az törökök rajta rémültek vala.

Gyorsan nagy vakmerőn öszveroppanának (V), Nagy (Ad) erős (Adj) viadalt (N) akkor (Ad) es (C) tartának (V), Szekereket magyarok meg nem bonthaták, Mert taraszk, puskákból sok golyóbist szórának. (201—216. sor)	Kevés magyar ismét nékiek téré, Nagy (Ad) erős (Adj) viadalt (N) vélek (Pr) tartának (V), Szegény (Adj) magyarok (N) sokan (NN) megholának (V) (115—129. sor)

[Figure 1]

In the formulaic lines we gave the grammatical categories of the words, in order to show the level of grammatical parallelism (N = Noun, C = Conjunction, P = Participle, V= Verb, A= Article, Pr = Pronoun, Ad = Adverb, Adj = Adjective, VP = Verbal Prefix, NN = Number Noun).[3]

In the 1980-90's Amedeo di Francesco dedicated several studies to the epic formulas of the 16-17th century Hungarian poetry (Di Francesco 2005). He applied the notion of formulaic style to explain the presence of formulas in a poetical tradition that involves written and oral compositions as well. This formulaic style was adopted by every author, independently of the primarily written or oral way of creation they have chosen (Di Francesco 2005, p. 148). He claims that the authors composing in writing tried to imitate by the use of formulas the style and the patterns of the oral epic poetry.

Di Francesco already took into account the metrical position of the formula in the line. He proposed a midway between Varjas' very large and open formula-definition, and the formula as it was interpreted in most of the relevant occidental literary analyses, that prescribed the combination of lexical elements and a grammatical structure in a specific metrical position. But his most relevant contribution to the debate was to distinguish inner (occurring only in one text, but several times) and external or shared formulas (present in more than one text).[4] He also proposed a typology of formulas according to their complexity: for him a simple formula is an exact repetition, a composite formula allows some variations, while the complex formula involves the fixed conjunctions and verbal locutions. His collection of examples was also limited to a smaller corpus, not the entire group of epic songs (Di Francesco 2005, p. 156–164).

[3] The lines in yellow, gray and turquoise contain external or shared formulas (common to both authors), the blue and red lines the inner formulas. We did not highlight with color the lines that have parallels only in other parts of the poems.
[4] In the rest of his study Amedeo di Francesco calls formula only the external formulas, the inner ones are called simply repetitions. There is in fact no difference between the two classifications for him. Nevertheless we find it necessary to analyze also the inner repetitions of a text, because it determines the structure of the text, and also because any lines repeated inside the text could have appeared also in other texts (of the same or of other poets). We maintain the distinction and terminology of inner and external formulas, and take into account both of them.

Di Francesco also differentiated the formulas according to their length: from very short ones, often reduced to a noun and an adjunctive (ADJ+NOUN) structure, via as longs as a hemistich, to those occupying a good part of a strophe (but despite the similarity of the meaning and the grammatical structure, the resemblance of the two strophes allows a very high variability).

2. The current research project

2.1. First steps

On the basis of the material collected by Varjas and Di Francesco, we tried to identify lexical and structural repetitions and parallelisms in the corpus, and to examine their variety by different computational tools and methods.

The creation of a digital (.txt and .json formats) corpus of epic songs allowed us first to search simply for more occurrences of beforehand identified formulas. The results confirmed to us the importance of variation in the use of formulas. For instance while Di Francesco was collecting the occurrences of the formulaic expression: "vala nagy bánatja" [he/she had a great sorrow], he omitted to find the same expression in another verbal tense: "lőn nagy bánatja" [in consequence he had a great sorrow], identified in the song RPHA 1328. In the same way he finds in the poem RPHA 1189 the expression: "sokat gondolkodék" (see the line "Harpagus ű róla sokat gondolkodék" [Harpagus was thinking about him/her a lot]), but misses to find "Róla Cresus király sokat gondolt vala" [King Cresus had been thinking of him a lot] in the poem RPHA 0525, where the difference is again the conjugation of the verb "gondol". He finds the expression "nem sok idő múlván" [in a while] in Gyergyai Albert's poem (RPHA 0053), but misses to identify in the same poem its variant "nem sok idő múlva". According to him the first form occurs altogether in 12 texts, in 15 lines, but the computer-based search could identify 7 more texts containing it. The variant "nem sok idő múlva" was found altogether in 11 texts. Besides Gyergyai's poem, the Fortunatus (RPHA 0560) is the only text containing both. The formula occurs more frequently in the beginning of the line, but not always!

We can assume that the level of variation is much higher than it was supposed to be by Di Francesco (DiF). Varjas accepted a much higher level of flexibility. He considered as formula so simple expressions as the naming of the fortress of Gyula as Vég-Gyula (word composed with the epitheton "vég" ('of the border'): "az Vég-Gyula. . .", "Vég-Gyulában", "Vég-Gyulábul" lines 3, 39, 41, 64, 79, 92 and 95 of the *Cantio* (RPHA 0369).

Another example of a too loose formula-definition is when he considers as variations of the same formula some very different expressions containing the lemma "számlál" [to count]:

- „Megszámlálásra hagyom ..." (Tinódi, RPHA 1244)
- „Számlálok majd én is ..." (Szakmári Fabricius István, RPHA 1246)

- Megszámlálók egynéhány vitézeket. (Cantio)

Di Francesco criticized Varjas for a very loose selection of formulas, but in fact some of the formulas identified by Di Francesco are also rather simple, and do not seem to be specific to poetry. It is true that the examples gathered by Di Francesco often fill precisely one hemistich, so they have a precise metrical function in that way. But still so, they seem to be more typical expressions, or even just frequent collocations, not idiomatic phrases, because their degree of lexical variability is very high. Some of his examples also appear in the Old and Middle Hungarian corpus of informal language database (OMH) (https://tmk.nytud.hu/about.php) [Figure 2].

- Formula identified by De Francesco (DiF): **Ezt hallván - - (-) I - - - - - (-) (-) (Tinódi, Szegedi, etc.)**
  Example in the OHM (OMH) : "**Ezt haluan** vajat es mezet vevek kezembe."
- (DiF): **Csudálatos vala I - - - - - (-) (Batizi, Ilosvai, Varsányi, Hunyadi)**
  (OHM): **"chudalatos**, hogy az kegyelmed ioszagan nem bekesegesek az en emberim."
- (DiF): **Nem sok idő múlván I - - - - - (-) (Tinódi, Sztárai, Ilosvai, Dézsi, stc.)**
  (OHM): "**nem sok idö muluán**, ezen Molnárra Feleségéuel edgiüt sok ideigh valo hidegh lölésnek giötrelme szálot."
- (DiF): **- - - - - (-) I az hatalmas Isten (Farkas, Varsányi, Batizi, etc.)**
  (OHM): "**A hatalmas Isten** tartsa meg kegyelmedet nagy jó egészségbe'!"
- (DiF): **Kinek talám mássát ti/tü nem/sem hallottátok (Istvánfi, Ráskai, Tinódi)**
  (OHM): "soha **mását nem** láttam, **melynek mását** szörnyűség **hallani**," [variation]

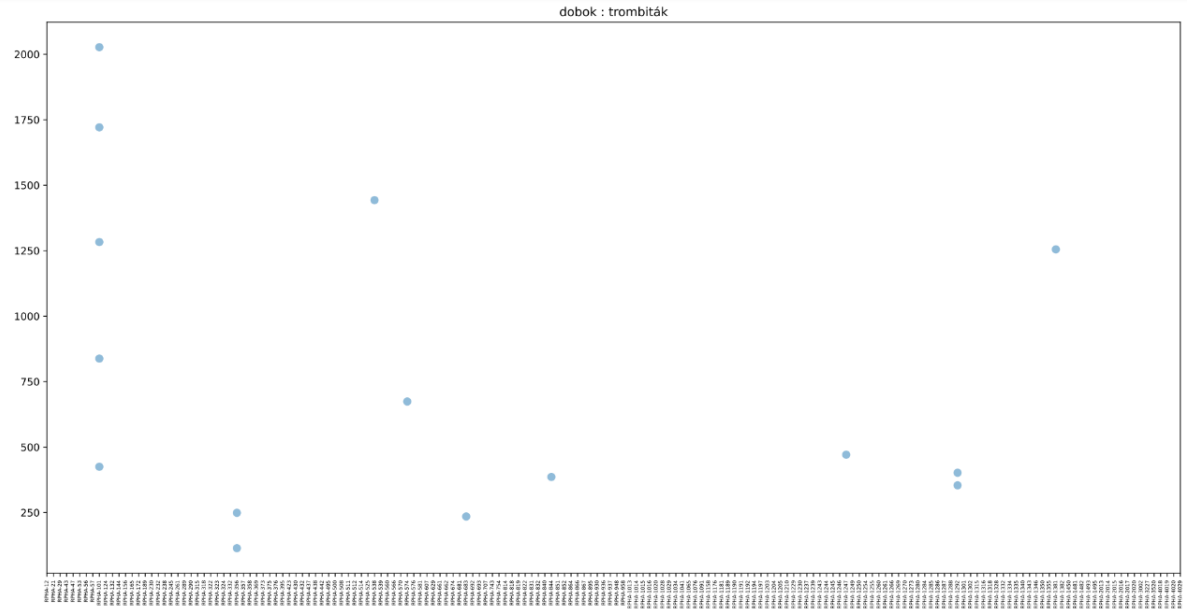
[Figure 2]

In order to understand the real nature of grammatical and lexical repetitions in the historical songs we tried to detect them by computational tools in our corpus. In this process we analyzed the grammatical structures in rhyming position (not only the rhyme-word), and we intended to collect the repetitions (parallel grammatical structures, partly or totally identical lines).

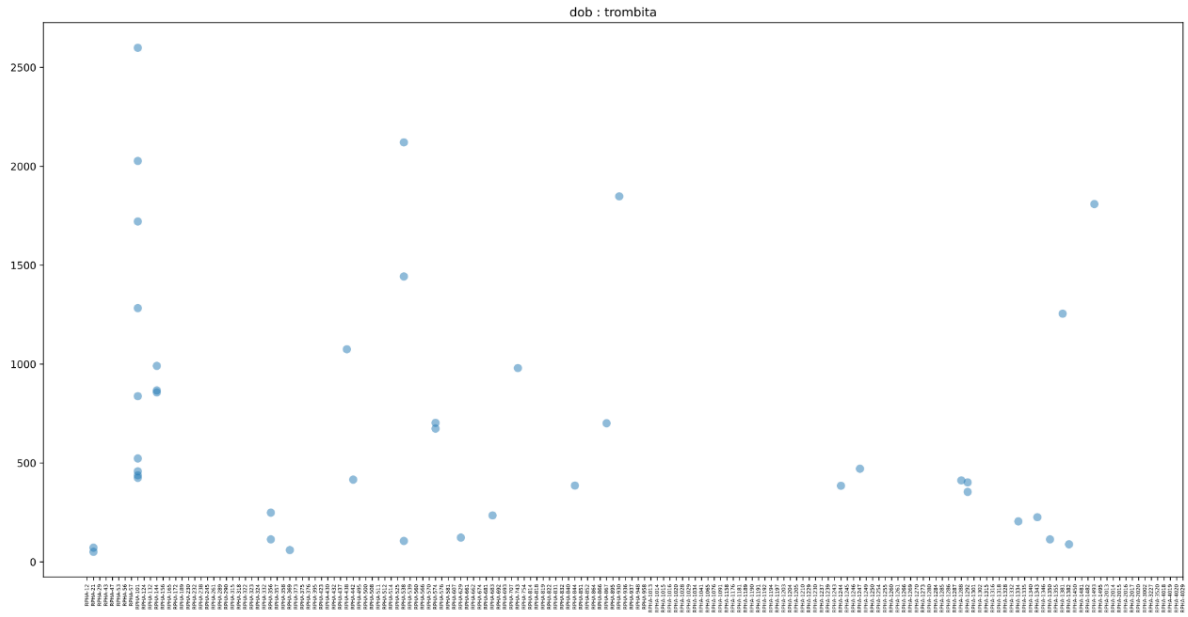
2.1.1 Keywords

The complementation of the catalog of hitherto identified formulas was tried by simple search for keywords, for the most frequent word collocations, and also for combinations of two lemmata. For example we have found as a frequent collocation the words "dobok" and "trombiták". The lemmata "Dob" [drum] and "trombita" [trumpet] appeared also among the 100 most common lemma-

collocations [Figure 3, 4].[5]



[Figure 3]



[Figure 4]

Both words are very common in the genre, mostly in the descriptions of battles. The sound of these two instruments is a topos in these scenes, and we suspected that it might be part of formulaic expressions. But neither the search for lemmata nor for word collocations could not help us to find a solid, frequently used formula. The word collocation of "dobok" and "trombiták" occurs 15 times in 9 texts, while the lemmata collocation "dob" and "trombita" are combined in the same line 41

[5] The purpose of the diagrams is merely to illustrate the appearance of word collocations and lemma-collocations in the corpus. Horizontal axis: poems; Vertical axis: lengths (in lines); Dots: word collocations and lemma-collocations

times in 25 different texts. So the search for the lemmata could reveal much more similar expressions than the search for word-forms, but the variations of the verses containing them were too high to identify an epic formula in all verses containing just one of the instruments, and not even those that contained the collocation of both.

2.1.2 Regular expressions

We also tried to search for formulas identified by Varjas and Di Francesco, or by the members of our group, by regular expressions. It gave much better results than the simple searches, and significantly augmented the number of the findings. These two examples of REGEX queries illustrate the difficulties of the identification of the variants of the same expression in this specific corpus:

> 'mikor(on)? .*?jut(á)?nak[,.:]?' - (when they arrived to) (- 39 hits against 24 found by Di Francesco)
> '(t[ié]rd|f[eö]j).*?hajt|hajt.*?(t[ié]rd|f[eö]j)' (to bend his knee or to bow his head) (62 examples, formula not examined by Di Francesco)

This method worked only in the case of formulaic expressions identified beforehand, and with a great circumspection in order to formulate the regular expression in such a way that they help to find all potential variations of a formula. But again, the agglutinative character of the language and the orthographic variety of the corpus would make it too difficult, and very long, to do and to verify all similar searches, even with the regular expressions; the computational rule-based searching methods are confronted with too many morphological variations in the corpus [Figure 5]:

Searching for 'mikor(on)? .*?jut(á)?nak[,.:]?'.
RPHA-101: 1
Brassó havassára **mikor jutának**,
RPHA-1014: 1
Pannoniába **mikor** be **jutának**,
RPHA-1065: 1
Végezetre **mikor** űk az tenger mellé Varanához **jutának**,
RPHA-1210: 1
Cvik várashoz **mikoron** ők **jutának**,
RPHA-1250: 1
Panaszolkodással **mikor** bé**jutának**,
RPHA-1285: 1
Strázsásokra **mikoron** ők **jutának**,
RPHA-1286: 1
Az vitézek **mikor** közel **jutának**,
RPHA-1288: 1

Egyiptom felé **mikoron jutának**,
RPHA-1328: 2
Ide Erdélybe **mikor** be**jutának**,
Ide ez országba **mikor jutának**,
RPHA-1334: 1
Az mezén el-alá **mikoron jutának**,
RPHA-1340: 1
Az Trójában **mikoron** be**jutának**,
RPHA-1355: 1
Szalkai mezőre **mikoron jutának**
RPHA-1381: 1
Ötödnapra **mikoron** ők **jutának**,
RPHA-1382: 1
Kozári mezőre **mikoron jutának**,
RPHA-373: 2
Az tengernek szigetéhez **mikoron jutának**,
Rómaságnak várasához **mikoron jutának**,

[Figure 5]

2.1.3 POS-tags

We also focused on the variability of grammatical patterns. Namely, we measured the entropy of each stanza based on the sequences of the four line-final POS-tags in order to have a picture of the importance of parallel structures in the creation of the strophe. The morphological analysis was done with the help of a version of the emMorph analyzer specifically designed for the treatment of Old and Middle Hungarian texts (Váradi, Simon, Sass, Mittelholcz, Novák, Indig, Farkas, and Vincze 2018; Indig, Sass, Simon, Mittelholcz, Vadász, and Makrai 2019). This way not only the lexical repetitions, but parallelism of the grammatical structure of subsequent lines of a stanza could have appeared:

Grammatical parallelism of the lines in the same strophe [Figure 6]:[6]

(1) Az Olaszországnak egy szegeletiben,
[Det|Pro] + [N][Dat] + [N|Pro] + [N][PxS3.Pl=i][Ine]
(2) Az Vezule hegynek ű kerületiben,
[N|Pro] + [N][P] + [N][Dat] + [N|Pro][S3] + [?]
(3) Nagy sok szép városok vadnak ű fektiben,
[Adv] + [Q] +[Adj] + [N][Pl] + [V][P3] + [N|Pro][S3] + [N][PxS3=i][Ine]

[6] The false results are highlighted with color.

(4) És jeles szín népek laknak a mentiben.”
[C] + [Adj] + [N] + [N][Pl] + [N][Dat] + [Det|Pro] + [N][PxS3=i][Ine]
(Istvánfi Pál, RPHA 0318)

[Figure 6]

The lines (1) and (2) have the same structure, just like the lines (3) and (4).

Despite the many corrections and developments the emMorphOMH still doesn't function properly. One of the typical faults was the parts of speech clarification, because the grammatical homonyms caused several false analyses. Like pl. *laknak* [N][Dat] or *laknak* [V][P3], where the second analysis is the correct one, but the analyser failed to recognize it. The historical morphological forms equally caused false interpretations of some of the totally identical word structures, even in rhyme position:

szegeletiben [N][PxS3.Pl=i][Ine]*,
kerületiben [?]*,
fektiben [N][PxS3=i][Ine],
mentiben [N][PxS3=i][Ine].

In this case two of four analyses are only correct. At the actual state of the analyser the relatively high rate of failure (above 6%) makes it impossible to offer a reliable computer-assisted morphological analysis of the corpus. Further improvements are needed, and will be done, in order to obtain precise results concerning the grammatical parallelisms. We keep on working on this (Horváth, Maróthy, and Simon 2023).

So unfortunately all these attempts failed to show an unequivocal result. The formulas could not be systematically revealed by these methods. The differences between the poems were not strong enough, and the identification of the most important recurrent lexical elements was not possible. The main reason for the failure of these attempts was the specificity of the corpus: the orthographic variability of the word forms, the high number of morphological variations of cases and suffixes of the Hungarian language of the 16th century, as well as the dialectal differences of the texts.

2.2. Character bigram: formulas and line sharing

Because of the unequal and unreliable results of the aforementioned methods, and the difficulties of identifying formulas (and in general to define the epic formula in its intrinsic variability as it appears in this corpus), we needed a new tool. As in our corpus a line corresponds most of the time to a clause, it seemed to be meaningful to search for the repetitions on the level of the lines. We decided to try a character bigram search in order to identify the closest lexical parallels of the lines in the entire corpus. As we are dealing with poems, a search of parallelisms and repetitions on the level of the verse was absolutely appropriate. This method allowed us to discover a very rich network of

intertextuality, and also some unexpected long formulas. We followed the approach of Maciej Janicki (Janicki, Kallio, and Sarv 2022) and represented each line as a vector of character bigram frequencies. Cosine distance was then calculated for each pair of lines in the corpus. (With a corpus consisting of nearly 100k lines, this meant several billions of combinations, making the task way beyond the capacity of a regular PC. We are grateful to the Institute of Czech National Corpus for kindly allowing us to use their servers for this purpose.) We have included the punctuation, hence the number of perfectly identical lines (cos.dist. = 0) is very small (52 pairs); the presence or absence of a line-final period means cos.dist. > 0. After a thorough examination we have decided to investigate the pairs having cos.dist < 0.3. This way also line pairs of very different length were included, where the formulas were rather a case of substring matching. These cases suggest that the formulaic style in this epic poetry is not strongly meter-dependent!

Of course our main goal was to identify epic formulas in a more complete and more secure, automatized way than by simply searching for previously detected formulas, word collocations and expressions. Finding even very closely related, (almost) identical lines in the corpus is just an extra gain (see in Chapter 3.2 some of these shared lines). The analysis of the identical or similar lines allowed us also to propose a typology of the formulas.

3. Results

3.1. Inner repetitions

We kept Di Francesco's distinction of inner and external formulas. First we wanted to measure the proportion of almost identical lines in a single poem, and to detect by this way the inner formulas that could help the creation and the memorisation of the poem and to give a rhythm to it by the repetition of the (almost) identical lines at different points of the text.

We have several tables corresponding to the level of similarity of the lines. The absolute and the relative values [Figure 7, 8] of the repeated lines have been collected and represented. The more lines of the poem are identical (to a certain degree), the higher is the relative value. According to the decreasing level of correspondence in the lines there were of course more and more similar lines found inside the poems. But six of the first ten positions of the lists were always held by the same poems.

If we look at the absolute values, the first ten positions are held by the following poems:

Absolute value	cos.dist. ≤ 0.1	cos.dist. ≤ 0.15	cos.dist. ≤ 0.2	cos.dist. ≤ 0.25	cos.dist. ≤ 0.3
1.	Zombori-1255 46	Zombori-1255 75	Zombori-1255 108	Zombori-1255 198	Zombori-1255 393
2.	Hunyadi-538 31	Hunyadi-538 51	Hunyadi-538 88	Hunyadi-538 158	Hunyadi-538 343
3.	Drávamelléki-373 26	Drávamelléki-373 33	Valkai-1328 68	Valkai-1328 121	Valkai-1328 227
4.	Valkai-1328 15	Valkai-1328 32	Drávamelléki-373 49	Drávamelléki-373 68	Drávamelléki-373 113
5.	XXX-369 12	Ilosvai-692 14	Ilosvai-692 34	Ilosvai-692 61	Ilosvai-692 112
6.	Ilosvai-692 11	XXX-369 14	Görcsöni-101 28	Görcsöni-101 49	Cserényi-1493 111
7.	Sztárai-1016 11	Sztárai-1016 14	Sztárai-1016 26	Cserényi-1493 46	Görcsöni-101 98
8.	Batizi-124 10	Sztárai-1015 13	Sztárai-1015 20	Tinódi-1381 38	Tinódi-1381 79
9.	Tinódi-1381 9	Batizi-124 12	Cserényi-1493 19	Sztárai-1016 31	Sztárai-1016 68
10.	Sztárai-1015 8	Cserényi-1493 10	Tinódi-1381 18	Sztárai-1015 30	Szebeni-560 53

[Figure 7]

These tables show the first five poems in a relatively stable position. Allowing a greater difference of the lines does not change radically the position of the poems in the ranking: apparently the repetition of lines with minor or major differences characterizes the texts in the same way. Of course the absolute numbers of repetitions depend a lot on the length of the poems. In each table Antal Zombori's poem on the fights of the tribes of Israel, based on the Bible, is the leader. Zombori's poem contains 2051 lines in 342 stanzas that allows a very high number of inner repetitions. While the *Cantio de militibus pulchra* is the 5th in the list with 12 very similar lines at level 0.1, it disappears from the top ten at level 0.2 , and reaches the 11th position only, with 17 identical lines.

This is due to its relative short length. The statistics change seriously if we divide the total number of lines with the number of repeated lines. The strongest value of the relative number of the closest repetition of lines (0.1) is by far that of the *Cantio de militibus pulchra* (0.0689) followed at a high distance by another anonymous poem, *Rusztán császár históriája* (0.0287). This pivotal position is held by the *Cantio* if the distance is augmented to 0.15 (value: 0.0804), but the poem is only the second at the distance 0.2 (value: 0.0977), and only the fourth at 0.3 with a value of 0.1322.

Relative values	cos.dist. ≤ 0.1	cos.dist. ≤ 0.15	cos.dist. ≤ 0.2	cos.dist. ≤ 0.25	cos.dist. ≤ 0.3
1.	XXX-369 0.06897	XXX-369 0.08046	XXX-1190 0.10476	XXX-1190 0.14286	Zombori-1255 0.19161
2.	Drávamelléki-373 0.02876	Zombori-1255 0.03657	XXX-369 0.09770	XXX-369 0.10920	XXX-1190 0.17143
3.	Batizi-1192 0.02778	Drávamelléki-373 0.03650	Drávamelléki-373 0.05420	Zombori-1255 0.09654	Hunyadi-538 0.14913
4.	XXX-4018 0.02609	XXX-1190 0.02857	Zombori-1255 0.05266	Drávamelléki-373 0.07522	XXX-369 0.13218
5.	Zombori-1255 0.02243	XXX-570 0.02825	XXX-570 0.04520	Hunyadi-538 0.06870	Drávamelléki-373 0.12500
6.	Szkhárosi-576 0.02083	Batizi-1192 0.02778	Hunyadi-538 0.03826	XXX-570 0.06780	XXX-570 0.08475
7.	Csáti-376 0.01911	XXX-4018 0.02609	Sarlóközi-840 0.03652	Batizi-1192 0.05093	Valkai-1328 0.06998
8.	XXX-1190 0.01905	Csáti-376 0.02548	Batizi-1192 0.03241	Batizi-124 0.04787	Sztárai-1015 0.06860
9.	Batizi-124 0.01773	Sarlóközi-8400 0.02528	Csáti-376 0.03185	Sztárai-1015 0.04573	Fekete-1284 0.06566
10.	Sarlóközi-840 0.01685	Hunyadi-538 0.02217	Sztárai-1015 0.03049	Sarlóközi-840 0.04494	Batizi-1192 0.06481

[Figure 8]

The two first rankings show clearly the specific character of the *Cantio de militibus pulchra* in the corpus. The highest number of lines of the highest similarity (distance 0.1 and 0.15) is reached by far by this poem. As a higher distance of closeness is allowed, the *Cantio* falls somewhat back in the ranking. At 0.2 it is still very close to *A babiloniabeli Bél és sárkány bálvány istenekről való história* (RPHA 1190), which is a fragmentary paraphrase of two episodes of the biblical *Book of Daniel*. As the second half of this poem is missing, we cannot really judge if it contained in this lost part as many of repetitions as in the first part. Probably yes. This poem has a very loose versification, the meter is not regular, rhyming is accidental, so the main pattern that consolidates its structure is the repetition of line-long lexical elements: mostly parts of a dialogue, either introductions, like: "Szóla az király Dánielnek" [The king said to Daniel] and its pair "Szóla Dániel az királynak" [Daniel said to the king] or replicas, like "Meghallgassad felséges király" [Listen to this, mighty king], etc.

All these patterns (the rudimentary versification of that piece, its archaic character, the way how it reconstructs dialogues between the king, Daniel and the priests) explain perfectly well the outstandingly high rate of repetitions in the biblical paraphrase. It is much more interesting that at the distance 0.3 two other, very long poems precede the *Cantio* in the ranking: Zombori's biblical historical song (RPHA 1255), and Ferenc Hunyadi's poem on the *Siege of Troy* (RPHA 0538). In Zombori's poem the rate is so high that it means that almost every fifth line has at least one pair in the poem (0.1916)! Of course, we cannot exclude that the distance 0.3 is too high to notice only real

pairs; it is possible that it allows to assimilate lines which would not be so close to each other if they were not calculated by the algorithm, but still, the next value is much smaller (*A babilóniabeli Bél...*, RPHA 1190: 0.1714), and differences are very clear between the poems, and even the authors. The value of Zombori's poem is 1,4496 times bigger than that of the *Cantio*. It might be explained also by the fact that we considered that Zombori's poem has some non-rhyming lines, and cut into two two long lines (the metrical structure is a14(7,7), a14(7,7), a21(7,7,7), a21(7,7,7) in the RPHA, but a14 (7,7), a14 (7,7), x14 (7,7), a7, x14 (7,7), a7 in our analysis). If long lines of 21 syllables would have been compared, probably the number of almost identical lines would have been limited. If the seven-syllable lines would be considered only as hemistichs, then it could be very interesting to examine the formulaic repetitions on the level of hemistichs, and not on the level of the lines, as we do here and now. Anyhow, the fact that Zombori and Hunyadi pass over the anonymous author of the *Cantio* in reuse of lines shows that our assumption that the *Cantio* has by far the strongest oral features in the corpus, must be somewhat alleged. The high rate of inner repetitions suggest a special poetic and aesthetic conception in case of these two long epic poems: a stronger relationship to orality or at least to a more archaic poetical practice of composing.

The *Cantio* is very strongly structured, the repetitions concern the dialogues and also the action (the descriptions of battles, plenty of narrative topoi), and apparently the almost identical imitations all along the poem offer a strong and compact unity to it. The length of 57 stanzas does not allow to fill with repetitions all the poem: to treat all topics the author needs a larger variation of lines as he is advancing in the preparation of the battles and the military acts themselves; or at least, the level of variability becomes too high to be perceived by the algorithm.

The most important gain of the comparison is to show that the *Cantio* does not stay alone with its high density of inner repetitions or formulas. When the value of repetitions is higher than 0.1 at the distance rate 0.3, we can suspect a strong impact of a traditional (orality-based) versification technique. Repetition of longer lexical units, sometimes of entire lines, was a mnemotechnical device that could facilitate the composition and the memorization of the texts. With the *Cantio* there are five poems above that limit: Zombori's (RPHA 1255), the paraphrase of the *Book of Daniel* (RPHA 1190), Hunyadi's text (RPHA 0538), and the paraphrase of the *History of emperor Rusztán* from the *Gesta romanorum*, by the Anonymous of Drávamellék (RPHA 0373).

It is important to state that almost every subgenera are represented among these poems: biblical paraphrase, contemporary report song of a war-event, epic song based on a plot taken from the Bible, and also rewritings of medieval Latin narrative texts! So this poetic technique or pattern is *not only* related to the genre of the report song, as an occasionally performed poem, supposed to be more spontaneous, reflecting on very recent events.

Further investigation is needed to establish if there are other poetical and metrical patterns joining together these five poems more than the rest of the corpus. But we can already observe that most of

the poems with a high ratio of inner formulas have no acrostic (highlighted in orange), or contain only the name of the author in the first letters of the first stanzas (highlighted in blue) [Figure 9].

Relative values	cos.dist. ≤ 0.1	cos.dist. ≤ 0.15	cos.dist. ≤ 0.2	cos.dist. ≤ 0.25	cos.dist. ≤ 0.3
1.	XXX-369 0.06897	XXX-369 0.08046	XXX-1190 0.10476	XXX-1190 0.14286	Zombori-1255 0.19161
2.	Drávamelléki-373 0.02876	Zombori-1255 0.03657	XXX-369 0.09770	XXX-369 0.10920	XXX-1190 0.17143
3.	Batizi-1192 0.02778	Drávamelléki-373 0.03650	Drávamelléki-373 0.05420	Zombori-1255 0.09654	Hunyadi-538 0.14913
4.	XXX-4018 0.02609	XXX-1190 0.02857	Zombori-1255 0.05266	Drávamelléki-373 0.07522	XXX-369 0.13218
5.	Zombori-1255 0.02243	XXX-570 0.02825	XXX-570 0.04520	Hunyadi-538 0.06870	Drávamelléki-373 0.12500
6.	Szkhárosi-576 0.02083	Batizi-1192 0.02778	Hunyadi-538 0.03826	XXX-570 0.06780	XXX-570 0.08475
7.	Csáti-376 0.01911	XXX-4018 0.02609	Sarlóközi-840 0.03652	Batizi-1192 0.05093	Valkai-1328 0.06998
8.	XXX-1190 0.01905	Csáti-376 0.02548	Batizi-1192 0.03241	Batizi-124 0.04787	Sztárai-1015 0.06860
9.	Batizi-124 0.01773	Sarlóközi-840 0.02528	Csáti-376 0.03185	Sztárai-1015 0.04573	Fekete-1284 0.06566
10.	Sarlóközi-840 0.01685	Hunyadi-538 0.02217	Sztárai-1015 0.03049	Sarlóközi-840 0.04494	Batizi-1192 0.06481

[Figure 9]

The exception here is again Zombori's long poem with a complicated dedication in Latin in the acrostic. The song *Ím, megromlottál, ó, jó keresztyénség* (RPHA 0576) by András Szkhárosi Horvát has also a very advanced place in the ranking at 0.1 (6th, with 0.020833), but it is probably due to the sermon-like character of the poem, which is quite far from that of the other historical songs. This rather short composition applies repetitions of lines in a didactic purpose, and disappears from the top ten already at the level of distance 0.15.

This table shows the bottom of the lists: the number of poems that have no line-long repetitions at all [Figure 10]:

Level	0.1	0.15	0.2	0.25	0.3
poems with 0 repetitions	89	65	47	32	20

[Figure 10]

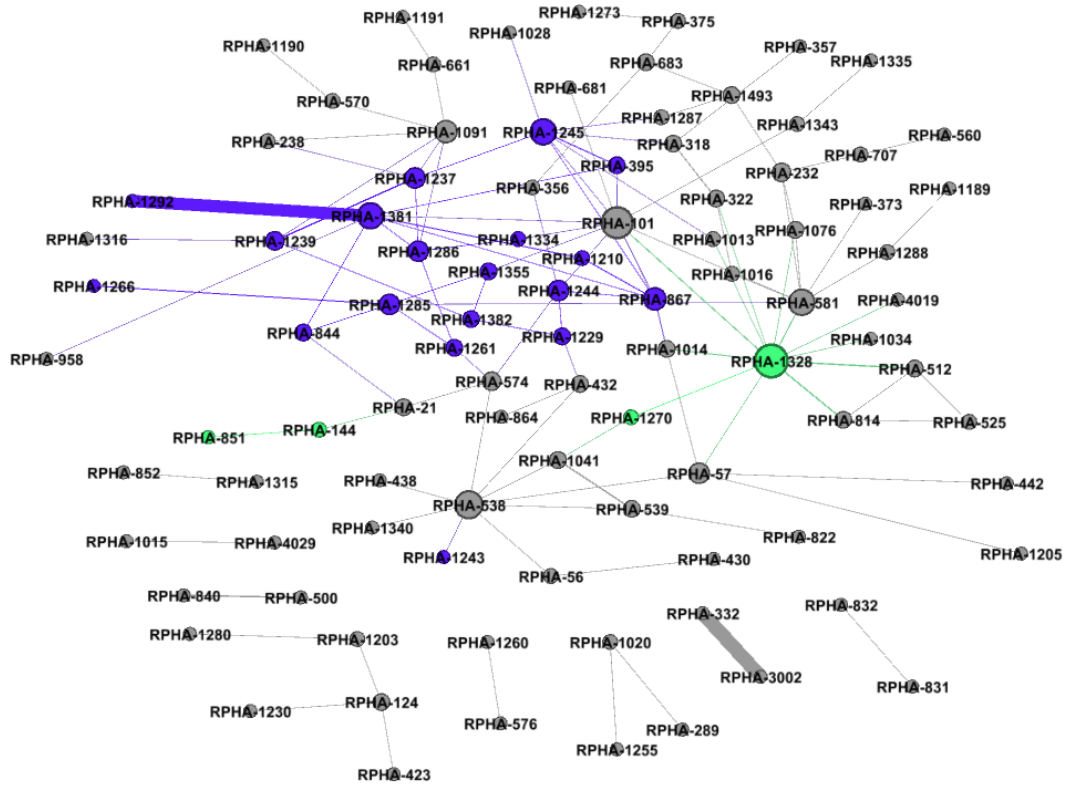
This statistics reveals also that more distant similar lines appear more easily than the closest variants, and that only 20 poems have no line-repetition at the highest allowed distance. These 20 exceptions include 8 poems by Miklós Bogáti Fazakas, 3 by Sebestyén Tinódi, and some other poets are present with only one composition in that list. As all the 15 poems of Bogáti Fazakas present in the list of historical songs are much closer to the bottom of the list than to its top, in his case we can really notice a refusal of a compositional technique based on line-long repetitions. His poem with the highest ranking is 82nd in the list 0.3, and this is actually not even a traditional historical song, but a paraphrase of the Song of Songs. It is much less epic in character than the others: its repetitive lines are due much more to the structural repetitions of the Biblical original than to the personal style of Bogáti or to a wish to imitate the formulaic style of his contemporaries and predecessors.

3.2. Borrowings, Imitations, Intertextuality: External or Shared Formulas

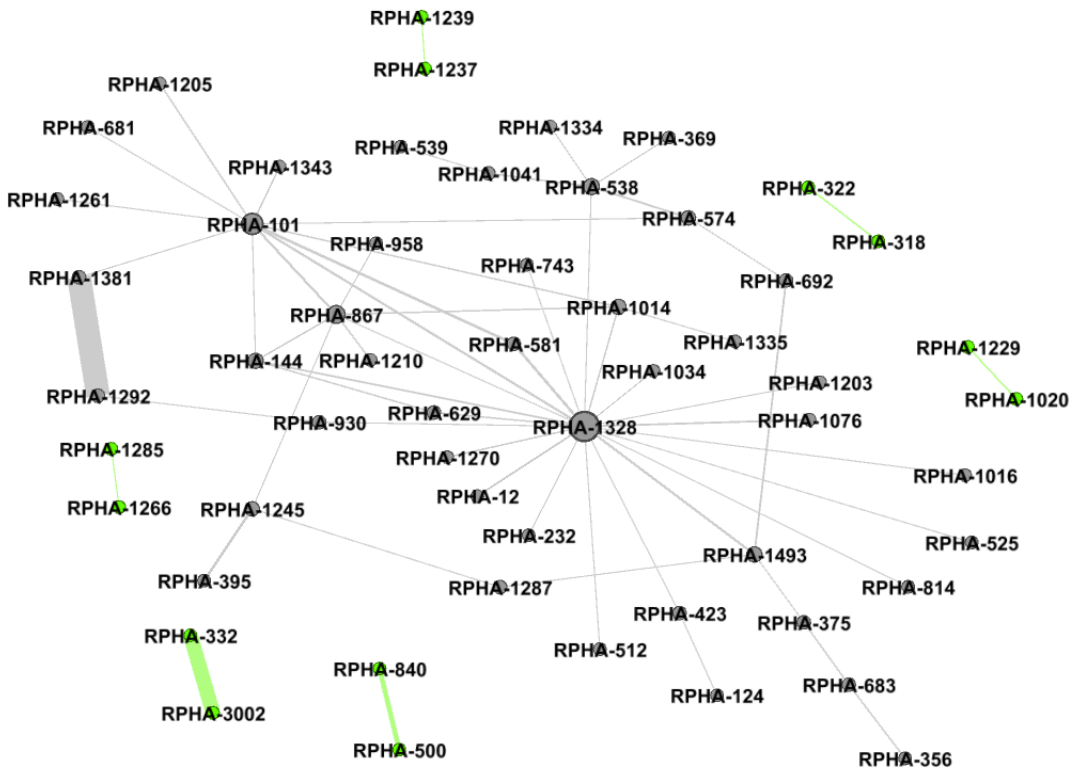
The analysis of the inner formulas has shown something about the individual compositional techniques of the poets. The identification of external formulas in form of common, almost identical lines to several poems, gives us another insight to their poetic devices: sharings of a high number of textual elements could reveal strong intertextual connections, or a common stock of line-long formulas. 5798 very similar (below 0.3 distance) line-pairs have been detected in the corpus by our method. Out of the 98503 lines of the corpus 6727 different verses were present at least once in the parallels: almost 7 percent (6,83 %) are involved in this table of the intertextual sharings.

The first picture shows the closest relationships between the texts [Figure 11]. It indicates that a very little number of poems share almost identical lines (96 nodes and 121 weighted edges). Tinódi and Valkai are present with the most poems and connections in the graphs. Increasing the value to 0.2 offers a much more complicated network [Figure 12]. At that point several pairs appear that share lines between each other, but not with the central network. Augmenting the potential distance of the lines some of these couples' relationships become stronger, while others disappear off the graph, as their common lines do not attain the necessary level (8 lines at the distance 0.3) [Figure 13]. In some cases we might explain these pairs with a common author, or eventually by a common topic.

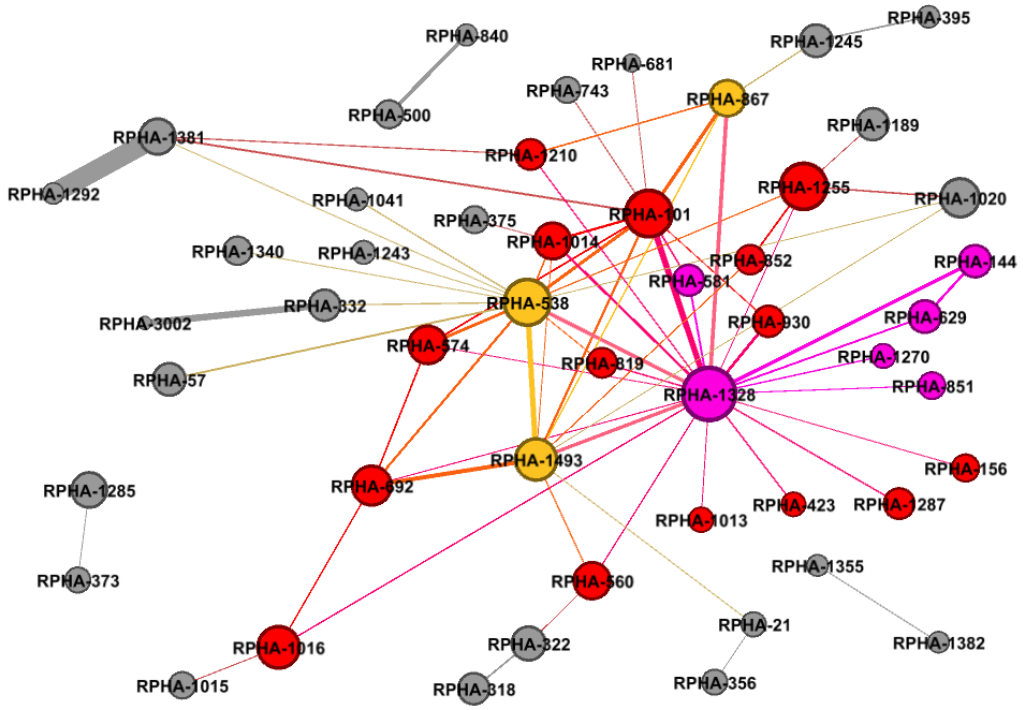
The more complex graphs above the value of 0.2 show in a very strong position the poem *Genealogia historica regum* (RPHA 1328) by András Valkai, a chronicle of Hungarian kings. It shares many lines not only with other poems by Valkai, but also with Cserényi (RPHA 1493), Tinódi (RPHA 0867), and Hunyadi (RPHA 0538). Of course, it is not independent of the length of these poems: respectively 811, 784, 297, and 575 stanzas of the four texts, some of the largest compositions in the corpus. But the thematic closeness of these poems is also evident: all the four are chronicles of older periods of history: three of Hungary, and one of Persia.



[Figure 11] - min. 1 line-sharing at the distance 0.1. Blue: Tinódi; Green: Valkai



[Figure 12] - min. 3 line-sharings at the distance 0.2. Green: single couples



[Figure 13] - min. 8 line-sharings at the distance 0.3. Pink: Valkai; Red: adjacents of *Genealogia historica regum* (RPHA 1328); Yellow: Cserényi (RPHA 1493), Tinódi (RPHA 0867), and Hunyadi (RPHA 0538)

In the frames of this article we cannot present all the parallelisms of the texts and all the possible conclusions of the line-sharings. But we would like to present some specific cases when the line sharing might be explained in different ways.

3.2.1. Authority

We observed in several cases a very strong parenthood between texts of the same authors. Tinódi's 22 poems are present in the table in 888 pairs, and 400 of his lines are coupled with lines of his other poems. András Valkai has only 5 historical songs, but 474 parallel lines, which is a very high number, assuring him a very central role in the network.

In two cases parenthood is especially strong, but they are easy to explain. In the first case one poem of our corpus is part of another one: a shorter part, the lamentation of Aeneas (RPHA 3002), of Huszti's paraphrase of the *Aeneid* (RPHA 0332) circulated as an independent text: in this case the number of almost identical lines is of course equal to the lines of the excerpt (40). All variants are only orthographic, and due to the vicissitudes of the textual transmission. The other case concerns Tinódi. From the level 0.15 the strongest connection is seen in the case of two poems by Tinódi, both dedicated to the siege of Eger. The earliest (RPHA 1381) poem was the object of the

summarized version, composed somewhat later (RPHA 1292). Tinódi borrowed a lot from his own poem, so it is not surprising at all that the two compositions share the most lines (113) at level 0.3.

3.2.2. Genre or editor?

Two love stories or romances of the corpus, Gáspár Ráskai's *History of the knight Franciskó* (RPHA 0322), and Pál Istvánfi's translation of the 100th novella of Boccaccio's *Decameron* (RPHA 0318) equally share a high number of lines (0.1: 2 lines, 0.15: 2, 0.2: 4, 0.25: 7, 0.3: 13) [Figure 17]. Istvánfi wrote his piece in 1539, Ráskai in 1552. Some lines are so similar that it would suggest a common author of the two texts: but we know that it is not the case. There are two ways to explain this: it is possible that Ráskai knew the work of his predecessor and voluntarily borrowed some lines from it. The other possible explanation is a radical intervention into the text of one of the poems (or in both) by the first publisher of both, András Komlósi, who printed these songs in the same year, 1572, in Debrecen. The tune given for Istvánfi's poem in the edition is precisely that of Ráskai: so they shared the melody. The following expressions are very similar in the two poems [Figure 14].[7]

[7] The words in red are identical; the words in orange are dialectal variants of the same expression; the words in yellow contain the same lemmas; the words in green are the same word-groups, except the word order.

Ráskai	Kinek talám mását tü nem hallottátok;	Istvánfi	Kinek talám mását ti nem hallottátok,
Ráskai	Nem hiszek ég alatt oly nyomorult embert,	Istvánfi	Nem hiszek én mostan ég alatt oly embert,
Ráskai	Minden múlatságát hátrahagyta vala,	Istvánfi	Egyéb múlatságát hátra hagyta vala.
Ráskai	Ez vitéz az asszont csak szemlöli vala,	Istvánfi	Az asszont az Volter csak szemléli vala,
Ráskai	Az vénasszony ebben nem resten jár vala,	Istvánfi	Az szolga nem resten ebben eljárt vala,
Ráskai	Róla sokat akkor ő nem gondolt vala,	Istvánfi	Nem sokat ő azon gondolkodott vala,
Ráskai	Keserűségében csak meg nem holt vala.	Istvánfi	Ű nagy örömében csak meg nem holt vala,
Ráskai	Abban egy vén jámbor akkor lakik vala,	Istvánfi	Az vén jámbor akkor kiballagott vala.
Ráskai	Ebben haladékot semmit nem halaszta,	Istvánfi	Ebben haladékot semmit nem kívánunk,
Ráskai	Atyjának, anyjának öröme lött vala,	Istvánfi	Atyjának, barátnak nagy öröme vala.
Ráskai	Nagy szép beszédekkel úgy kérdezi vala.	Istvánfi	Nagy szép beszédekkel urának szólt vala,
Ráskai	Ő nagy örömében csak meg nem holt vala,	Istvánfi	Ű nagy örömében csak meg nem holt vala,

[Figure 14]

3.2.3. Traces of the oral epic tradition?

As we told at the beginning of the article, several 15-16th century authors claim that Hungarians used to sing songs about the deeds of their ancestors. In the corpus of historical songs there are some

poems that are dedicated to old Hungarian (and Hun) history, and it was interesting to check if they have some shared formulas that could reflect a common background of oral composition. Actually we found three poems on this topic that share some surprisingly similar lines, and two parallels, that are formulated in such a similar way that they might come from an earlier common source. Görcsöni's poem has some parallels with the two others, but they might be explained by other reasons, the similarity is somewhat more superficial. But Valkai and Gosárvári both tell the death of the Hun king Attila revealing his high age:

Százhuszonnégy esztendős korba vala [He was 124 years old.] (Valkai RPHA 1328)
Százhuszonnégy esztendős korban vala (Gosárvári RPHA 1014)

They also tell the story of the German hero Dietrich, who survived of being shot by a Hun-Hungarian arrow in his forehead:

Homlokában Rómába vitte vala, [He brought it to Rome in his forehead] (Valkai RPHA 1328)
Homlokában Romában vitte vala, (Gosárvári RPHA 1014).[8]

It is worth taking into consideration that these poems tell the ancient story of the Hungarians, and it is possible that these shared lines are borrowed from some lost poems dealing with that topic. Nevertheless, even in this case we must be prudent: both Valkai's (1576) and Gosárvári's (1579) texts were first edited in Kolozsvár, by the same publisher. For this reason we cannot exclude that the printer fashioned both texts in the same way at that point.

3.2.4. Topic-dependence

In some cases also between those that are discussing the same topic**.** And in some cases the high number of shared lines between two poems might lead to a new authorial attribution.

The Anonymous of Sarlóköz (RPHA 0840) wrote a poem on the Last Judgment in circa 89 stanzas in 1552. It shares 25 similar lines with a poem attributed to Péter Bornemissza on the same topic, composed in or before 1582, of 202 stanzas. It is a rather high number. Péter Dobai also wrote a

[8] The lines in context:
Csak kevesen az Deitrich elszalada, / Homlokba nyíllal azt is lőtték vala, / Homlokában Rómába vitte vala, / Azért haláltalannak híják vala. (RPHA 1328)
(Transl.: *Shortly after Deithrich ran away, / He was shot by an arrow in his forehead, / He brought it to Rome in his forehead, / That's why he is called immortal.*)
Homlokában nyilat belőtték vala, / Deitrich hadnagy fejét találták vala, / Homlokában Romában vitte vala, / acrin viadalban elveszett vala. (RPHA 1014)
(Transl.: *He was shot by an arrow in his forehead, / Captain Deitrich was shot in his head, / He brought it to Rome in his forehead, / Macrin was killed in a battle.*)

poem on the Apocalypse (RPHA 0012), but it is less connected to the two others. The 47 stanzas long surviving fragment shares only 5 lines with RPHA 0840, and only one with RPHA 0500. This is the earliest poem on the topic, contained in the same edition as Péter Bornemissza's, so Bornemissza must have known it. Still, it seems clear that despite the identical topic, not all the three poems share the same lines, and not in the same proportion (although the hypothetical discovery of the lost part of Dobai's poem could maybe change the data). We might conclude from this that sharing a theme did not oblige the authors to share lines too.

3.3. Thematic groups of the formulas

The analysis of parallel or shared lines allowed us to have an overview of the topics that are frequently associated with formulaic expressions. Some of these recurrent elements were not surprising at all, while some others could be observed only thanks to our approach. The identity of a structural function or an identical narrative element explains in many cases the appearance of almost identical formulas in different poems.

3.3.1. Structural function - Composition and dating

The opening and closing sections of the poems contain many similar expressions. It is really frequent that the authors describe in the last, colophon-strophe of the poem the conditions of the composition. The name of the author, a verb meaning roughly "to write" or "to compose", and an object are the obligatory elements of such a formula.

Examples:

- Ezt énekbe szerzé az Batízi András, RPHA 1192
- Tinódi Sebestyén írá könyvében, RPHA 1229
- Nem jelenté meg nevét ő ezekben, RPHA 01270

Many poems contain dates: either historical ones, or the year of the composition of the poem (these in the colophon). No less than 193 different lines begin with "Ezer" (One thousand) and the huge majority of the lines contain a date.

Examples:

- Ezernégyszáz írtak az negyvenötben, RPHA 0101
- Ezerötszázban és ím az negyvennégyben, RPHA 01230
- Ezerötszáz után és az hatvannyolcban, RPHA 01254

3.3.2. Identical motifs

Some acts, gestures are expressed with very similar expressions in our corpus. It is highly probable that they were used also in everyday communication in a very similar way. The expression “hálát ad” (to give thanks) is surprisingly frequent in the list, with many different actors and recipients of the thanks: God, a king, a lady, a lord, etc.

Examples:

- Oláh az Istennek nagy hálát ada. RPHA 0101
- Nagy hálákat jóvoltáért Istennek adjanak, RPHA 1482
- Ez nyereségön nagy hálát adának. RPHA 1244

Several scenes of meeting were described in the corpus, that gave occasion to present the high esteem to another person, kneeling one’s knees and/or head(“térdet/fejet hajt”). This act is also really frequent in the corpus.

Examples:

- Tirdet, fejet hajtván, söveget hányának, RPHA 1254
- Térdet, fejet Hectornak hajtnak vala. RPHA 538
- Ez egy úrnak térdet, fejet hajtana; RPHA 1189

As there are numberful military acts, wars, battles, sieges narrated in the corpus, it is not surprising that a high number of them were expressed in a stereotypical way. In this case we cannot speak of a single formula, but much more of a group of formulas, related to different moments of the confrontation: the preparation (sounds of drums, trumpets, or even cannons) engagement, the clashes, the loss and the victory. We can give here just a very little number of examples, without entering into details.

Examples:

- Sok dob, trombita erősen harsoga, RPHA 0629
- Trombitát fútata, dobot üttete, RPHA 0144
- Sípok, dobok, trombiták szünögtenek, RPHA 1381

We do not claim that all the topics of the identified formulas have been described in this list. Neither that all the appearances of the aforementioned formulas have been found, because our method allowed us to identify only the closest parallels. But still, it gives a very solid starting point for the definition of the formulas.

3.4. Definition of the line-long formulas in the historical songs: The limits of variation

The above-mentioned two eminent scholars, Béla Varjas and Amedeo di Francesco described the Hungarian historical song as a genre full of formulas and repetitions. But they could not offer a precise definition of the epic formula and they had no tools to measure and to show the real impact of this poetical tool on the genre. Some of the formulas they selected are maybe not really "formulaic": they are common, basic verbal locutions, frequent also in everyday communication, and in their form too flexible to be considered as enrooted formulaic expressions, way less constrained than the above-mentioned expressions of gratitude or warlike acts (Chapter 3.3.2).

As we could identify quite clearly formulas at the length of a verse, we are proposing a somewhat different conception and definition of the formulas than the previous scholarship. It seems to us that it is impossible to give a concise definition of the formula based on shorter expressions, or any other recurrent elements of 16th century Hungarian poetry.

According to our definition the formula:

- stays in the frames of a single line (only with very-very few exceptions)
- contains at least three core elements, and some optional or accidental, not obligatory ones
- lexical variability may also affect *some of* the core elements, but the variations are always *synonyms* or *belong to the same semantic field* (so the number of variations is rather limited in case of these few core elements).

These points are different from Di Francesco's categories, while some of his samples of formulas do not fulfill these criteria, being too short or too trivial. At the same time we keep Di Francesco's definition of the distinction between the functions of the inner and shared formulas. The difference of these functions will be more deeply analyzed in another paper.

The formulaic style reflects a special compositional technique. It shall be identified in only a small part of the poems in our corpus, when the main poetic device of the text is the reuse and variation of some expressions or lines of the same composition. The *Cantio de militibus pulchra* and the *History of emperor Rusztán* are the most evident surviving examples of that technique.

In the formulaic style

- the syntactic structures of the lines are *parallel*, either *identical* or *very similar*
- stating the parallelism of the syntactic structures of the lines, the morphological variation (number, flection, verbal tense, verbal prefixes) of the same lexical elements is allowed. The function words and other optional elements might change
- rhyme position affects our perception of the formula: if the rhyme is the same in case of two similar lines, their formulaic character seems to be stronger. In the same way, the identical beginning of the verses strengthens the formulas.

3.4.1. Example: A shared formula in all its forms

The next table illustrates this conception of the formula with all occurrences of the expression: “valamilyen viadalt tartani” (‘to fight or to joust somehow’, literally: ‘to hold some kind of joust’) [Figure 15].

1. Nagy erős viadalt ők ott tevének, RPHA 1266
2. Nagy erős viadalt vélek tartott vala. RPHA 1243
3. Nagy erős viadalt mindkét fél tartának, RPHA 1334
4. Nagy erős viadalt mindkét fél tarta, RPHA 0101
5. Csuda nagy viadalt velök tartának, RPHA 1335
6. Az törésön ők nagy viadalt tartának, RPHA 1245
7. Csuda erős viadalt vélök tarta, RPHA 1335
8. Nagy erős viadalt vélek tartának, RPHA 0369
9. Nagy erős viadalt akkor es tartának, RPHA 1334
10. Nagy erős viadalt akkoron tartának, RPHA 1334
11. Új viadalt Cignussal kezdett vala RPHA 0538
12. Nagy erős viadalt esmét tartának, RPHA 0538
13. Rettenetes viadalt indítának, RPHA 0538
14. Erős viadalt pogánokkal tarta, RPHA 0144
15. Mert erős viadalt ők ott tartának, RPHA 0867
16. Vízárokból viadalt es tartottak: RPHA 0245
17. Nagy viadalt velek törletének. RPHA 0245
18. Nagy erős viadalt hajdúkkal kezdének, RPHA 1334
19. Szekér környül nagy viadalt szörzének, RPHA 0867
20. Az szekér mellett viadalt tart vala, RPHA 0867
21. Törökekkel ott nagy viadalt kezde, RPHA 0867
22. Hogy végső viadalt ők mívelnének. RPHA 1014
23. Nagy viadalt egymással tartnak, vesznek, RPHA 1014
24. Bátor szívvel vélök nagy viadalt tőn; RPHA 1335
25. Új viadalt Hectorral kezdett vala. RPHA 0538
26. Mindkétfelől erős viadalt tőnek, RPHA 1261
27. Az pórokkal nagy viadalt ő tarta, RPHA 0395
28. Az víz mellett erős viadalt tévének, RPHA 0683
29. Ott nagy erős viadalt Israelnek fiai - RPHA 1255 (ellenséggel tartanak.)
30. Ott nagy erős viadalt Israelnek népei - RPHA 1255 (pogányokkal tartanak.)
31. Mind napnyúgatig nagy viadalt tarta, - RPHA 0432
32. Belől terekekkel nagy viadalt tartnak, - RPHA 1245
33. Ím nagy viadalt hamar ők kezdének. - RPHA 0574

34. Kapuközbe nagy viadalt tart vala, - RPHA 0844
35. Kikkel öszve erős viadalt tartának, - RPHA 0356
36. Rettenetes viadalt véle tarta, - RPHA 1328
37. És nagy viadalt vélök tartának, általmennek vala. - RPHA 0324
38. Ütközének, nagy viadalt tartának, - RPHA 0867
39. Derék harcot, sebes viadalt kezdének, - RPHA 0819

Nagy/erős/retternetes/sebes/új/csuda/végső [jelző(k) / attributive(s)]+viadalt [tárgy/object]+kezd/tart/tesz/művel/szerez/törlet/indít [állítmány/predicate]+(alany/subject)+(időhatározó / adverbial of time)+(helyhatározó / adverbial of place)+(társhatározó / comitative adverbial)

[Figure 15]

As we can see, the order of the three obligatory elements of the formula follow each other in the same order. An adjective marking the noun 'viadal' (as an object), and a verb are present in each line, except in 16 and 20, where an adverb replaced the adjective. Some complementary elements might appear at different places of the line, but this core of the formula is always present. The tense and the number of the verb might change, and several synonyms of the verb 'tartani' [to hold] might occur in the formula ("szerezni", "tenni", "mívelni"), and if not synonyms, than some verbs belonging to the same semantic field: "kezdeni", "indítani" [to begin]. The two most frequent adjectives are "nagy" [big] and "erős" [strong], occurring frequently together, but "rettenetes" [terrible], "sebes" [fast], "új" [new], "végső" [last] do occur as well. The adjective "csuda" [marvelous] accompanies/complements either "nagy" or "erős". These variations are allowed in the frames of the formula. 17 examples were identified by the bigram character search, and 22 others found manually. 2 more verses containing the word-form "viadalt", and revealed by the automatic bigram search as parallels to the 17 formulaic lines were judged not to belong to the formula.[9]

3.4.2. Intertwining of the formulas

This illustrates perfectly that in the historical songs the formulas create a network, not independently of their strong variability. These formulas are not static, and they are not simply enumerated one after the other. By their meaning and their often changing structure they are intertwined and strongly joined to each other, strengthening the poetical effect of the work.

The character bigram search identified a very high condensation of formulas in the first part of the *History of emperor Rusztán* by the Anonymous of Drávamellék. This poem shows many signs of oral

[9] Another way to express the same meaning is when "viadal" is the subject of the phrase. There are 11 cases in the corpus, five of them in poems by András Valkai. The adjectives "nagy" and "erős" occur also in some of these 11 lines, and in six the same verb "tartani" occurs, but in a different meaning: "valamennyi ideig tart" [to last for a period].

composition (lack of acrostics, clumsiness of rhyming, etc.). As a last example of the compositional technique in a formulaic style we show the imbrication of three formulas in an excerpt of this poem.

The three formulas [Figure 16]:

- Úristen + meghallgatni + X (birtokos) + könyörgés [God + to listen + someone's + prayer]
- fölvenni + jelző (szent, nagy) + ke(ö)resztsége(ö)t/szegénységet/gazdagságot [take + ADJ + christianity/poorness/richness] the adjectives "nagy" and "erős" occur also in some of these 11 structures, and in six lines the same verb "tartani" occurs, but in a different meaning: "valamennyi ideig tart" [to last for a period]
- szeretni + ifjúságban/vénségben + szegénységet/gazdagságot [to love + in youth/in old age + poorness/richness]

8. „Kérlek, Uram, hallgassad meg én könyörgésemet,
Ne nézd, Uram, pogányságink, de nézd jóvoltodat,
Én is fölveszem Rómaságban az szent körösztséget,
Körösztyénség mellett szegénséget és nyomorúságot."

9. Az Úristen meghallgatá asszony könyörgését,
Eustachius mert mikoron ágyban fekünnék,
Fényösséggel, villámlással ő környülvéteték,
Az Istennek ő követe előtte állapék:
...
11. Azt izené, hogy fölvegyed az szent körösztséget,
Rómaságban ő ez hármon hagyott szabadságot:
Ha szereted ifjúságban az nagy gazdagságot,
Vagy szereted ifjúságban az nagy szegénységet.

12. Harmadikon ezen hagyott neked szabadságot:
Vénségedben ha szereted az nagy gazdagságot,
Vagy szereted szegénységet, mind szabadon hagyott,
Mert az Isten körösztségedben ezzel látogatott."
...
15. Meghallgatá az Úristen ifjú könyörgését,
Meghallgatá Eustachiusnak az ő könyörgését,
Megmutatá: azki nem tűr, nem veszi gyümölcsét.
Halljátok meg, mint mutatá ezekhez szerelmét.

16. Ím fölvevé ifjúságban az szent körösztséget,
És az mellett nem vevé föl az nagy gazdagságot,

Meggondolá az háromnak sanyarú vétkeit,
Azt gondolá, föl kölly vönni az nagy szegénséget.

17. *„Ha én - úgymond - ifjúságban szegénséget veszek,*
Ifjúságinkban mi erőnkkel kenyerünket esszük,
Aratáshoz, nagy munkához derekunkat hajtjuk,
Mi testünket nagy munkával be is födözhetjük.

18. Vénségemben inkább veszem az nagy gazdagságot,
Mert akkoron mi erőnktűl inkább megfosztatunk,
Akkoron kölly inkább nekünk az mi szegedelmünk."
Ez szó hallván az Istennek követe elmúlék.

[Figure 16]

The intermingling of the three formulas shows that it is a very conscious, well-trained use of a technique of composition. The main poetical principle of this poem is without any doubt the use of formulas as inner repetitions, and not the quality of rhyming or the grammatico-syntactical parallelism, despite the fact that there is rhyming and of course also some lines built with the same syntax. The rhymes are rather poor, and phonetically they don't always match to each other (see: gazdagságot - szegénységet). The composition reflects the orality in all respects, and approaches the poem to the *Cantio* and the paraphrase of the biblical *Book of Daniel* (RPHA 1190).

4. Conclusion

Our investigation tried to propose a better understanding of one of the most salient features of 16th century Hungarian versification, the use of parallel syntactic structures by the verses of the same strophe, and the rhetoric technique of using epic formulas. Among the numerous methods we tried, the most fitting for the analysis of the corpus proved to be the character bigram search. With its help we identified a huge number of very similar or identical lines inside the poems, and many shared lines between the poems of the corpus. This dataset offered us a huge stock of "potential formulas". A closer view on the data allowed us to identify several groups of formulas, having some core elements. On the basis of the results we proposed a new definition of the epic formula as it appears in our corpus. The computational method also helped us to identify some poems that are certainly influenced by, or largely dependent on an oral compositional technique. The line-long formula-definition seems to be a promising starting point for further discussion about intertextuality, simulated orality, and the epic formulas in the Hungarian poetry of early modernity.

Acknowledgments

The research in this study is supported as a National Research, Development and Innovation Office–NKFIH, OTKA 135631 project. Data and code are available at http://github.com/versotym/oldhun.

References

- De Chasca, Edmund. (1970). "Toward a Redefinition of Epic Formula in the Light of the Cantar de Mio Cid." In: *Hispanic Review,* 1970.3, pp. 251–263. DOI: https://doi.org/10.2307/471929
- Di Francesco, Amedeo (2005). "A históriás ének mint formulaköltészet." In Ead.: *Kölcsönhatás, újraírás, formula a magyar irodalomban – Tanulmányok*. Budapest : Universitas.
- Dömötör, Adrienne, Gugán Katalin, Novák Attila, Varga Mónika (2017). "Kiútkeresés a morfológiai labirintusból – korpuszépítés ó- és középmagyar kori magánéleti szövegekből". In: *NyK*, 2017, pp. 85–110. OMH. URL: https://tmk.nytud.hu/about.php) (visited on 04/08/2023)
- Horváth, Andor, Maróthy Szilvia, and Simon Eszter (2023). "A históriás énekek számítógépes elemzésének módszertani kérdései". In: *A históriás ének: poétikai és filológiai kérdések*. Ed. by Seláf Levente. Budapest : Gépeskönyv. URL: https://f-book.com/book/2023/A-historias-enek-kerdesek/index.php?chapter=6. (visited on 04/08/2023)
- Horváth, Iván, Font Zsuzsa, H. Hubert Gabriella, Herner János, Szőnyi Etelka, and Vadai István 2022, *Répertoire de la poésie hongroise ancienne*, v. 7.3. URL: https://f-book.com/rpha/v7/ (visited on 17/08/2023)
- Indig, Balázs, Bálint Sass, Eszter Simon, Iván Mittelholcz, Noémi Vadász, and Márton Makrai (2019). "One format to rule them all – The emtsv pipeline for Hungarian". In: *Proceedings of the 13th Linguistic Annotation Workshop*. Florence, Italy: Association for Computational Linguistics, pp. 155–165. DOI: 10.18653/v1/W19-4018. URL: https://www.aclweb.org/anthology/W19-4018. (visited on 04/08/2023)
- Janicki, Maciej, Kati Kallio, Mari Sarv (2022). "Exploring Finnic written oral folk poetry through string similarity". In: *Digital Scholarship in the Humanities*, 2022.1, pp. 180–194. DOI: https://doi.org/10.1093/llc/fqac034. URL: https://academic.oup.com/dsh/advance-article/doi/10.1093/llc/fqac034/6643198 (visited on 04/08/2023).
- Kay, Sarah (1983). "The epic formula: a revised definition". In: *Zeitschrift für französische Sprache und Literatur*, 1983.2, pp. 170–189. URL: https://www.jstor.org/stable/40617027 (visited on 09/03/2023)

- Novák, Attila, Gugán Katalin, Varga Mónika, Dömötör Adrienne (2018). "Creation of an annotated corpus of Old and Middle Hungarian court records and private correspondence". In: *Language Resources and Evaluation*, 2018, pp. 1–28. OMH. URL: https://tmk.nytud.hu/about.php) (visited on 04/08/2023)
- Oláh, Miklós (2000). Hungária. Athila (Millenniumi magyar történelem - Források, 2000). Budapest : Osiris. ISBN: 963-379-916-3.
- Parry, Milman (1928). *Les formules et la métrique d'Homère.* Paris: Les Belles Lettres. URL: https://chs.harvard.edu/read/parry-milman-les-formules-et-la-metrique-dhomere/ (visited on 01/02/2023).
- Seláf, Levente (2020)." Between Lyric and Epic: The Great Turkish War in German, Italian and Hungarian Ereignisliedern". In: *Controversial Poetry 1400–1625*. Ed. by Judith Keßler, Ursula Kundert, and Johan Oostermanby. Radboud Studies in Humanities 11. Leiden ; Boston : Brill, pp. 61–86.
- Tatlock, John S. P. (1923). "Epic Formulas, Especially in Lazamon". In: *PMLA*, 1923.3, pp. 494–529. URL: https://www.jstor.org/stable/457331 (visited on 09/03/2023)
- Váradi, Tamás, Eszter Simon, Bálint Sass, Iván Mittelholcz, Attila Novák, Balázs Indig, Richárd Farkas, and Veronika Vincze (May 2018). "E-magyar – A Digital Language Processing System". In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Ed. by Nicoletta Calzolari *et alii*. Miyazaki, Japan: European Language Resources Association (ELRA). ISBN: 979-10-95546-00-9
- Varjas, Béla (1982). "A históriás ének és társadalmi funkciója". In: Varjas, Béla *A magyar reneszánsz irodalom társadalmi gyökerei*. Budapest : Akadémiai. URL: http://real-eod.mtak.hu/12045/1/AkademiaiKiado_002342.pdf (visited on 01/02/2023).
- Windelberg, Marjorie, and D. Gary Miller (1980). "How (Not) to Define the Epic Formula." In: *Olifant*, 1980.1, pp. 29–50. URL: View of How (Not) to Define the Epic Formula (iu.edu) (visited on 04/08/2023)