Acta Ethnographica Hungarica 69 (2024) 1, 187–210
DOI: 10.1556/022.2025.00011

1828
AKADÉMIAI KIADÓ

The Possibilities of Enhanced Reading. Analyses with Voyant Tools on the Hungarian Verbal Charm Corpus

EMESE ILYEFALVI*

Institute of Ethnography and Folklore, Eötvös Loránd University, Hungary

Received: November 15, 2023 • Accepted: February 14, 2025

© 2025 The Author(s)

ABSTRACT

This study explores the potential of digital text analysis in examining Hungarian verbal charms. Using the Voyant Tools platform, it applies keyword frequency analysis, collocational network mapping, and distribution visualizations to a digitized corpus of over 6,000 annotated texts and metadata. The paper compares historical and ethnographic subcorpora, emphasizing how collection methods, researcher preferences, and questionnaire design have shaped the resulting textual patterns. Rather than drawing universal conclusions, the study highlights how the structure, editorial processes, and data preparation of the corpus itself influence what patterns can be detected and what interpretive questions can be asked.

KEYWORDS

Hungarian verbal charms, digital text analysis, folklore corpus, computational folkloristics, corpus visualization, Voyant Tools

RESEARCH OBJECTIVES AND ANALYTICAL FOCUS

In my paper, I summarize the lessons learned from some of the analyses carried out on the text corpus prepared for the digital database of charms in Hungarian. The primary goal of my investigations using data-driven computer text analysis and data visualization was to search for different patterns and carry out their interpretive processing: how and to what extent do the algorithmic analyses of the large-scale corpus support the findings of earlier Hungarian charms

* Corresponding author. E-mail: ilyefalvi.emese@btk.elte.hu

AKJournals

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

research? In addition, I implicitly reflect with my analyses on the question of how different publishing practices, the method of preparing texts and the theoretical concepts behind them, i.e., textualization techniques, influence the interpretation. In general, with various digital rearrangements of the texts (Rockwell – Sinclair 2016:40; Wittek 2016:130), what new questions can be asked of the current corpus with a tool suitable for enhanced reading?

THE DIGITAL DATABASE OF HUNGARIAN VERBAL CHARMS

The database construction and the digital preparation of the texts started in the Fall of 2014,[1] in this study, I analyse different corpora generated from the material of 6,037 records.[2] The assistants of the research group added the metadata for each record to the database in an Excel spreadsheet:

- typology (based on Éva Pócs's typology published in 2014: Pócs 2014)
- function (what purpose was the text used for: e.g., curing headaches, love magic, etc.),
- ritual-gesture (standard Hungarian description of the ritual-gesture related to the text),
- type of source (text collected from orality or from different types of historical sources, cf. Ilyefalvi 2014),
- place of collection,
- collector/scriber
- informant (name, gender, age, and other details, if known),
- language of the data,
- location of the data,
- and bibliographical reference of the original publication.[3]

We chose a tabular format due to the relatively large number of metadata fields and the very short, simple nature of the texts. During digital processing, each charm record was placed in a separate Excel cell, distinct from the metadata. However, we applied additional coding within each charm record, as it included more than just the reading text.

Our textualization procedures were guided by the aim of capturing the data as fully as possible – including co-text and context – within a single data item. Within each record, we annotated the title (< >), the context ( ), the specific text ({ }), and the collector's question (# #) so that these elements can be analyzed separately in later stages.[4]

[1]Building the database was launched within the framework of the ERC project: *Vernacular Religion at the Boundary of Eastern and Western Christianity: Continuity, Changes and Interactions* under the European Union's Seventh Framework Program (FP7/2007–2013)/ERC grant agreement No. 324214 http://eastwest.btk.mta.hu/. (accessed: 14 February 2024).

[2]This includes slightly more charm texts, but the number of texts under discussion is approximately accurate. Quantifying the material is complicated by the fact that, in some cases, it was difficult to clearly mark the beginning and end of individual charms – sometimes two or three texts may be embedded within a single data record. For a more detailed discussion of this issue, see: Ilyefalvi 2017:55–56 and Ilyefalvi 2019:161–165.

[3]For more detail on metadata, see: Ilyefalvi 2019:167–174, as well as Ilyefalvi 2021.

[4]For more detail on textualization problems, see: Ilyefalvi 2019:158–166.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

The construction of the database is still in progress, and as such, the entire collection is not yet available for publication.[5] However, I have made the various subcorpora – generated from 6,037 records and forming the basis of the analysis presented in this paper – publicly accessible, accompanied by detailed documentation. A complete bibliography of the volumes, publications, and manuscripts that serve as the sources of the database is also provided.[6]

VOYANT TOOLS

Voyant Tools is an online text analysis and data visualization platform, made publicly available in its current form in 2016 by Stéfan Sinclair (McGill University) and Geoffrey Rockwell (University of Alberta), following several earlier versions and prototypes. It was specifically designed to enable scholars in the humanities, social sciences, and library sciences to conduct digital text analysis without requiring advanced programming skills, and to explore the interpretive possibilities of enhanced reading in a digital environment.

The program supports a variety of file formats: users can upload individual documents or entire corpora, allowing for the comparison of different texts and datasets. Voyant Tools offers a suite of built-in text analysis and visualization tools that can be freely combined, customized, and used dynamically within the interface, depending on the user's needs.[7] It also supports multiple export options for visualizations and accommodates hybrid forms of publication – including printed materials with associated online tools, visualizations, and interactive content – which are increasingly common in digital humanities research.

EXPLORATORY ANALYSES OF THE HUNGARIAN VERBAL CHARM CORPUS

Word frequency, word cloud, word distribution

In many cases, especially at the outset of research and without predefined questions, data-driven analysis of large-scale corpora reveals its value through the way computer programs highlight textual patterns and characteristics (Rockwell – Sinclair 2016:282). Stephen Ramsay described this exploratory, playful engagement with texts as the hermeneutics of "screwing around," emphasizing its importance to scholarly inquiry (Ramsay 2014:111–120). It resembles a reader browsing a library's open shelves, aimlessly drifting in the hope of a serendipitous discovery. So,

[5]These developments do not affect the results presented in this paper, primarily because the majority of the records added to the database since August 2018 do not contain full charm texts but only refer to them. Although the analysis follows Chapter 6.2 of Ilyefalvi 2019, the dataset used here has been further refined since that publication. Therefore, the corpora were regenerated and the visualizations were created specifically for the present study.

[6]https://real.mtak.hu/118305/ (accessed: 14 February 2024).

[7]For a total of 24 built-in digital tools and their description, with illustrative examples of use, see the following link: https://voyant-tools.org/docs/#!/guide/tools (accessed: 14 February 2024).

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

what can we learn about the charms corpus without asking targeted questions? What properties emerge from a basic word frequency analysis? Figure 1 below presents a word cloud based on the word frequencies of all charm-related data (ritual, context, and text) linked to specific dates.[8]

The visualization revealed several important properties of the corpus. For instance, it suggested that a significant portion of the texts may be related to healing (e.g., *gyógyítása* ['curing']), and that numbers – such as three, three times, and nine – play a prominent role. The high frequency of nouns associated with Christian religiosity (e.g., *szent* ['holy' / 'saint'], *Isten* ['God'], Jézus ['Jesus'], *Mária* ['Mary'], *kereszt* ['cross']) is also notable, as are recurring terms such as *árpa* ('stye' / 'barley') and *szem* ('eye') In the initial phase of text analysis, I did not apply a stopword list or exclude functional words, as these can sometimes yield meaningful insights.[9] In our case, the prominence of subordinating conjunctions like 'when' and 'if' proved especially noteworthy. According to Éva Pócs, in Hungarian charms, syntactically marked patterns of parallel wishes and commands often rely on structures such as 'as…, as', 'then…, when', and 'if…, then', all of which are built around subordinating conjunctions (Pócs 2014:37).

In the previous analysis, I did not visualize the most frequent words from the charm texts alone, but from the entire textual dataset – that is, the full set of charm records, including titles, possible ritual instructions, as well as additions by informants, collectors, or editors. For example, the word *gyógyítás(a)* ('curing') appeared with high frequency because it is commonly

Fig. 1. Word cloud visualization of the charms corpus

[8]In creating the charms database, we aimed for a diplomatic transcription; however, it is important to note that certain words may appear more prominently simply because their spelling is more standardized within the corpus, due to dialectal and orthographic variation.

[9]It is primarily used for stylistic analysis. See Labádi 2018:17–34, 17, 20–21; Archer 2009:1–15; Rockwell – Sinclair 2016:155.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

used as a title in verbal charm collections, and the data records often include titles assigned by researchers and/or informants. Examples include: 'Cure for styes', 'Cure for headaches', 'Cure for diphtheria', and so on.

Thanks to proper data preparation, however, we are able to generate a filtered version of the corpus that includes only the speech acts – that is, only the actual charm texts. After removing function words, we can observe the most frequent lexical items in the speech act corpus (see Table 1). Among the most common words ('holy', 'Jesus', 'eye', 'Mary', 'virgin', 'God', 'Holy Spirit', 'Christ', 'beautiful', 'blessed'), the word 'eye' stands out as the only one not directly connected to Christian religiosity. In contrast, adjectives like blessed and beautiful typically refer to the Virgin Mary, a central figure in both European and Hungarian charms.

Since the corpus data was arranged chronologically, a distribution diagram can be used to visualize how the most frequent words are spread across time, based on their relative frequencies. Figure 2 below shows the distribution of the top three most frequent word stems (*jéz** [Jesus], *szem/szöm** ['eye'], and *szent/szönt** ['holy' / 'saint']) across the corpus.[10] Notably, *szem/szöm** ('eye') is almost entirely absent from the first tenth of the corpus. This highlights the importance of examining frequent word distributions across different time periods, as these are far from uniform.

Of course, a natural question arises: what can we actually learn about the corpus from word frequency alone? Word lists and frequency data may point us toward topics by highlighting high-frequency lexical items, but they are insufficient for identifying actual topics, text types, motif groups, or charm types in themselves (Rockwell – Sinclair 2016:175).

Table 1. Most frequent words in the charms text corpus with their frequencies

szent ('holy' / 'saint')	2,491
jézus ('Jesus')	1,023
szem ('eye')	904
mária ('Mary')	897
szűz ('virgin)	703
isten ('God')	687
szentlélek ('Holy Spirit')	608
krisztus ('Christ')	603
szép ('beautiful')	578

[10]Voyant Tools allows for various types of word form searches. For instance, using the command *jéz** will retrieve all occurrences beginning with jéz- as a single result. Similarly, it is possible to search for word endings, identify words within a five-word proximity, or combine multiple word forms in a single query.
For a detailed overview of all possible search combinations and syntax, see: https://voyant-tools.org/docs/#!/guide/search (accessed: 14 February 2024).

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

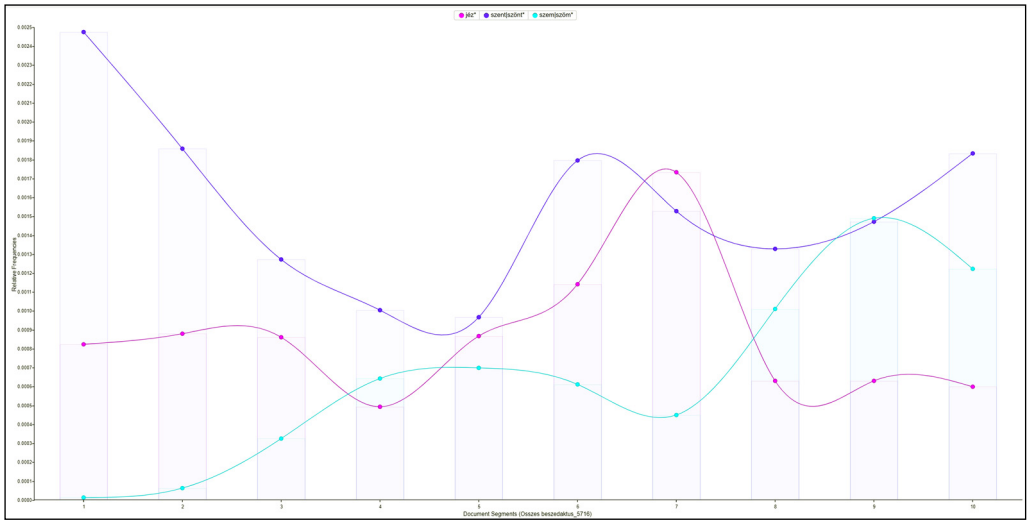

Fig. 2. Distribution of the top three most frequent word stems (*jéz**, *szem/szöm**, and *szent/szönt**) across the charms corpus

However, collocational network graphs – showing the words that most frequently co-occur with the most frequent terms – can offer more insight. In Figs 3 and 4, we see network graphs of the three most frequent words discussed so far ('Jesus', 'holy / 'saint', and 'eye'). Figure 5 presents a more detailed collocational graph for the word eye.

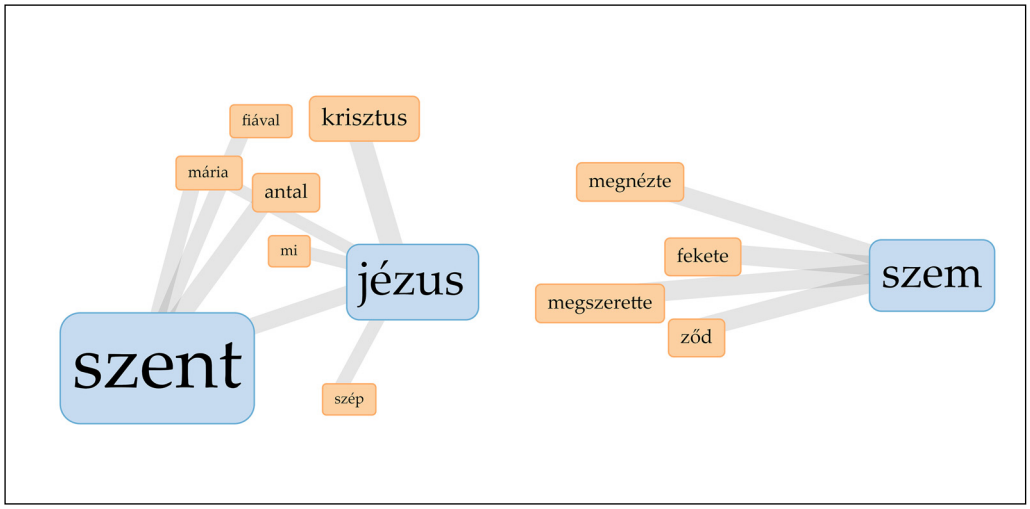

Fig. 3. Collocational network of the word forms *Jézus*, *szent*, *szem*, based on a 3-word co-occurrence window

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

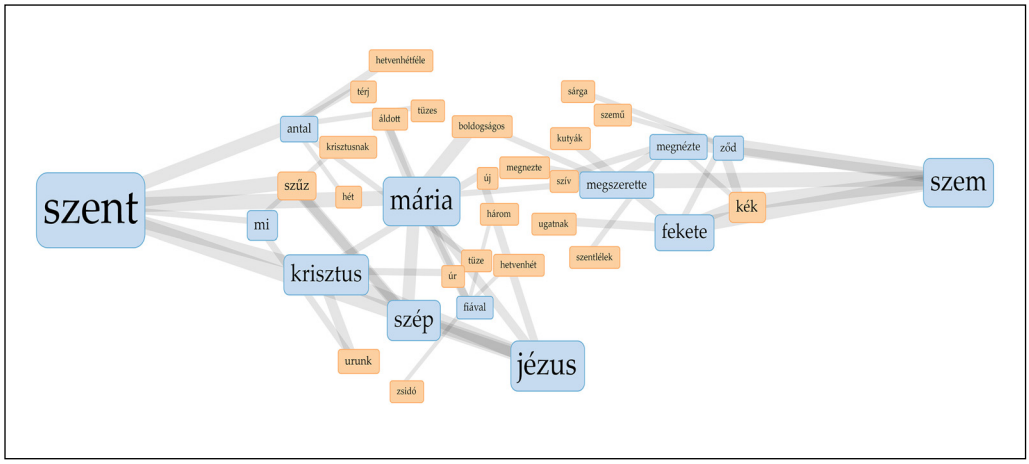

Fig. 4. Collocational network of the word forms *Jézus*, *szent*, *szem*, based on a 5-word co-occurrence window

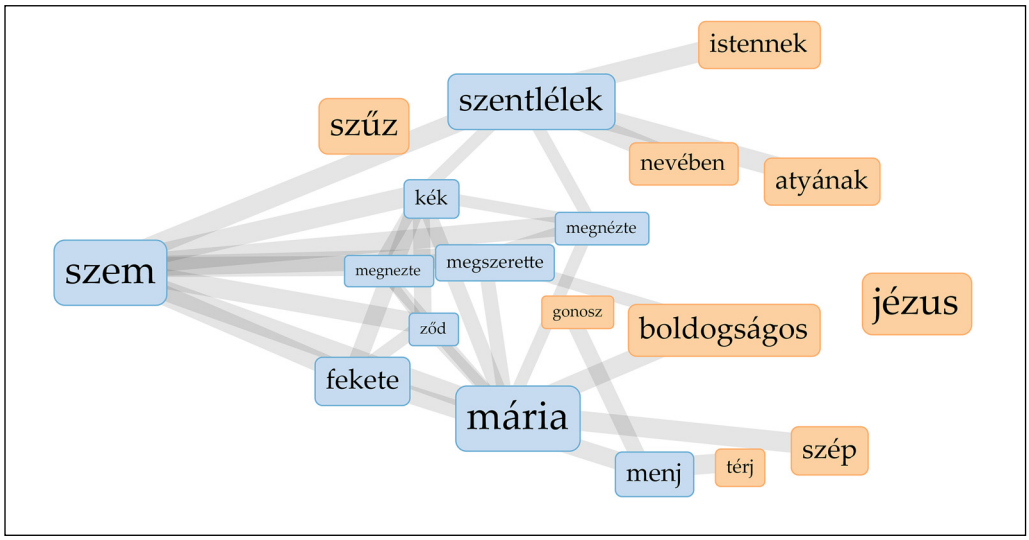

Fig. 5. Collocation network graph of *szem*, based on a 17-word co-occurrence window

A well-known charm type clearly emerges from the collocates of the word eye – such as *megszerette* ('loved'), *megnézte* ('looked at'), *zöld* ('green'), *fekete* ('black'), and *kék* ('blue'). This group corresponds to Group XI in Éva Pócs's typology, where the core structure of the charm involves the naming of harmful agents, followed by a wish, command, supplication, blessing, or curse. These functions are also reflected in the keywords of the network graph.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

This charm type was typically used to diagnose and prevent harm – usually associated with bewitchment or affliction caused by the evil eye. Such conditions represent one of the most common categories of ailments in Hungarian folk belief, often treated through verbal charms (Pócs 2014:577).

The network graph in Fig. 3 illustrates simpler texts that primarily list the causes of the illness. In contrast, Fig. 5 shows how this type frequently merges with other charm genres, such as narrative charms and/or religious supplications – as indicated by words like blessed, Mary, etc.

> "The healer »blew on his head « and recited the charm three times: * {Looked at him with a blue eye, / a green eye, / a white eye, / a dreadful eye, / an evil eye. / May the Blessed Virgin Mary descend into his head and eyes, / into every bone and limb, / and let it be as if he were born and brought forth by the Virgin Mary! / Ptü, ptü, ptü [spitting]!"[11]

> "I make the sign of the cross on her forehead. Then I say: *{A blue eye looked at her, / a black eye cast the spell. / Comforting, beautiful Virgin Mary, / comfort this woman!} It must be recited three times, then you must touch her forehead with your tongue and pat it down"[12]

> "They say that the one who casts the evil eye is a woman or man whose child was taken from the breast and then nursed again. I heard it said this way, half in jest, as in an old tale. She came home, and her head hurt so badly she could barely get inside. She put down what she had brought, lay down, and while she slept, I read this over her: * / {Blue eye, green eye, mottled eye, yellow eye looked at her, a virgin girl loved her. I sprinkle her with the same words that were used to sprinkle Jesus, when he, too, was looked at. Our sweet mother Mary sent Lady Saint Helena to bring three drops of water from the Jordan River. Lady Saint Helena went and brought back three drops. Jesus was washed with it, and he was better right away. May it help Anna in the same way it helped Jesus.} / * Then I blew over her and said she should pat it down and spit. away from the place where I read it, so the illness would leave her, and the place as well. That night, she came back to herself."[13]

> "For protective sprinkling against enchantment *{The Blessed Virgin Mary set out / on a very long journey. / She met three Jewish girls, / they undressed her and kissed her. / Go now, saints and holy women, / and throw three sacred coals / into the Jordan River. / May my sprinkling help this person / just as the Blessed Virgin Mary / helped her blessed holy Son. / An eye looked, a heart was enchanted, / and may the Holy Spirit / heal them. Our Father. / Praised be Jesus Christ.}"[14]

One of the key challenges in computational text analysis is that software cannot distinguish between polysemous words or homonyms, as it does not interpret meaning but processes only sequences of characters. However, KWIC (Keyword-in-Context) tools offer a useful way to address this limitation.[15]

[11]Gajcsána (Găiceana, România) place of collection: Egyházaskozár, Hungary, 1968, collected by Tekla Dömötör and Éva Pócs. Pócs 2014:663–664.

[12]Gyimesközéplok (Lunca de Jos, România) 1971–1972, collected by Sándor Bosnyák. Bosnyák 1982:120.

[13]Gyimesközéplok (Lunca de Jos, România), 1972, collected by Anikó Salamon. Salamon 1975:138; Pócs 2014:859–860.

[14]Székelyudvarhely, (Odorheiu Secuiesc, România) 1974–1999, collected by Árpád Daczó. Daczó 2003:277; Pócs 2014 849.

[15]For the relationship between KWIC-type tools and printed concordances, see Rockwell – Sinclair 2016:45–67.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

Left	Term	Right
megyek bűnös Ádám földjére / termő	szem	tördelni, / áldott búza veszteni." / Ez
adj te nekem többet! Vak	szem	élő' légy, / Isten földet teremtett
léleknek nevében. — Föld, édes anyám! [...]	szem	megnézte, szív megverte, sz. háromság
a használjon az egész világnak.	szem	megnézte, szív megverte, szent háromság
Fiú roncsa, / Szentlélők eloszlassa! »Kék	szem	, fekete szem. Kézzel, vízzel lemosom
Szentlélők eloszlassa! »Kék szem, fekete	szem	. Kézzel, vízzel lemosom. Hogyha férfi
hét, ez nem hat, „Kék	szem	, fekete szem, sárga szem, ezer
nem hat, „Kék szem, fekete	szem	, sárga szem, ezer látta, száz
Kék szem, fekete szem, sárga	szem	, ezer látta, száz levágta, Jézus
nevében a kősziklához. Ámmen!" [...] „Kék	szem	igézte, fekete szem igézte, verje
Ámmen!" [...] „Kék szem igézte, fekete	szem	igézte, verje meg az Isten
nevében. Ámen. / Jézus, Mária szerelmiért. /	szem	megverte, szív szerette: / Vetem kalap
szív szerette: / Vetem kalap alá. /	szem	megverte, szív bágyasztja: / Úr jézus
szív bágyasztja: / Úr jézus orvosoljа! /	szem	megverte, szív szerette: / Vetem fejkötő
szív szerette: / Vetem fejkötő alá. /	szem	megverte, szív bágyasztja: / Úr jézus
szív bágyasztja: / Úr jézus orvosoljа! /	szem	megverte, szív szerette: / Vetem pántlika
szív szerette: / Vetem pántlika alá. /	szem	megverte, szív bágyasztja: / Úr jézus
legyen apád, anyád, minden rokonod! „	szem	Piétër iés Szem Pá, [c
minden rokonod! „Szem Piétër iés	szem	Pá, [c] mon még [d
Ujvesárnop napküötökor emëntem ién Paradicsomkerbe,	szem	mise hagatásáro. Fötekintek aranuómaféro, beleesett
e, mikor alfelemet idë ütöm!" „	szem	Piétër iés Szem Pá, [c
idë ütöm!" „Szem Piétër iés	szem	Pá, [c] mon még [d
Úrjézus bocsássa elébe a gonosztevőket!	szem	látott, / Szív vert meg! / Szent
istem mëgatta, ámmën. Isten elindúlt	szem	[a] Pétër után a palláson
mëgattad, ámmën. Segics më, istenëm,	szem	[a] Mária, szent anyám, hogy
pártájába, ez kis Bandikánknak a	szem	ne ártson." Elindula szűz Mária
alu gyüttié, párta alá mönny! [...] »	szem	meglátto, szüv megverte, Isten lásso
megkeresztelteteti! [...] Ellen mondok a ragályos	szem	megnézésének! [a] Ugy ne álljon
igézetek: kékszemű, feketeszemű, sárgaszemű emberek.	szem	megnézte, szűv megszerette, szem megigézte
emberek. Szem megnézte, szűv megszerette,	szem	megigézte. Én is tiltom tűzzel
az Atya, Fiú és Szentlélekisten.	szem	meglátta, szív megszerette, vigasztald meg
meg, / János verte meg, Kék	szem	látta, barna szem látta, mindenféle
meg, Kék szem látta, barna	szem	látta, mindenféle szem látta. Boldogságos
látta, barna szem látta, mindenféle	szem	látta. Boldogságos Szüz Mária jöjj
férfi, akkor «kalap alá». «Valahány	szem	megnézte, annyi angyal szájjon melléje
meg, / hasadjon ki a csicse	szem	megnezte, / szű megszerette, / vigasztaló szent
szent lélek, / vigasztald meg Pétert.	szem	megnezte, / szű megszerette, / vigasztaló szent

Fig. 6. Voyant Tools' Contexts tool displaying occurrences of different forms of the word *szem*

In Voyant Tools, for example, the Contexts window allows users to quickly determine whether the word *szem* ('eye') in the corpus consistently refers to the human eye, or whether it also appears with other meanings (see Fig. 6).

In Fig. 7, we can see that the automated search may also return words that are irrelevant to the analysis – for example, *szömőcs/szemölcs* ('wart') – simply because they share the same initial sequence of characters as the target word. In this case, however, the number of such occurrences is negligible.

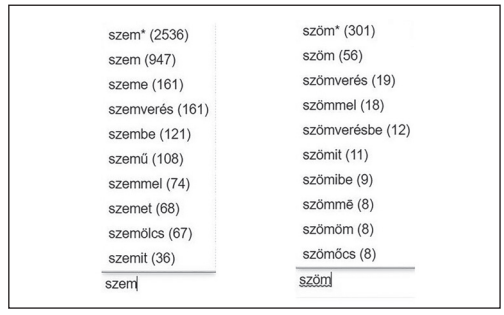

Fig. 7. Voyant Tools' Contexts tool displaying a pop-up window with concordance lines and the various word forms of the search term

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

Nonetheless, we can examine the distribution of word forms beginning with *szem/szöm** (e.g., *szem, szemek, szöm*, etc.) separately. To explore this further, I divided the corpus into two subcorpora: 1) Historical texts – texts recorded or published before 1851; 2) Ethnographic texts – texts collected from oral tradition after 1851 through ethnographic fieldwork. Using Voyant Tools' Bubblelines visualization tool (see Figs 8 and 9), I analyzed the distribution of *szem** and *szöm** word forms across these two subcorpora.

The main result of the word distribution analysis confirmed a key finding from previous charms research: there are no recorded instances of charms for healing bewitchment or the evil eye in historical sources – that is, in texts originating from ecclesiastical or secular literate culture. According to Éva Pócs, this supports the conclusion that such charms existed only in oral, folk practice (Pócs 2014:579).

Texts healing styes and the evil eye

A comparison of the two subcorpora shows that one of the most characteristic words in the ethnographic texts is *árpa* ('stye' or literally 'barley'), which occurs 469 times. This figure aligns with previous findings: according to Éva Pócs, alongside the so-called "counting/reading" texts associated with the evil eye, *árpa*-healing charms form the most significant group within the known corpus of Hungarian verbal charms. Moreover, like charms against the evil eye, *árpa*-healing charms are entirely absent from historical (written) sources (Pócs 2014:20, 257). It is worth noting that the Hungarian word *árpa* is a homonym: it refers both to the grain 'barley' and to a common eye infection known in English as a 'stye'.

If we examine the distribution of *árpa** within the corpus, we can see that from 1960 to 1979 the presence of the word is exceptionally high, even in terms of relative frequency (Table 2 and Fig. 10).

What accounts for this rise? An earlier analysis of the corpus revealed that the 1960s and 1970s can be considered the "heyday" of collecting activity in terms of data volume. However, this is partly due to the fact that this period coincided with the collection work for various

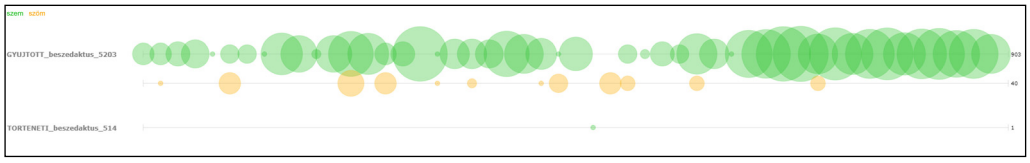

Fig. 8. Intra-corpus distribution of *szem** and *szöm** in sub-corpora before and after 1851

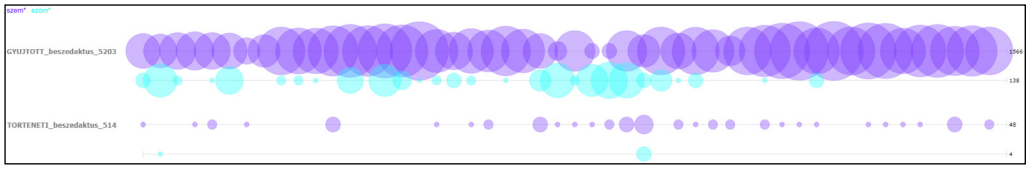

Fig. 9. Intra-corpus distribution of *szem** and *szöm** in sub-corpora before and after 1851

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

Table 2. Ten most frequent words in the pre-1851 and post-1851 subcorpora, with number of occurrences

Pre-1851 subcorpus	Number of occurrences of the word form	Post-1851 subcorpus	Number of occurrences of the word form
szent ('holy' / 'saint')	1823	*szent* ('holy' / 'saint')	668
szem ('eye')	903	*minden* ('all')	265
Mária ('Mary')	854	*Istennek* ('for God')	220
Jézus ('Jesus')	840	*Jézus* ('Jesus')	183
szűz ('virgin')	631	*Isten* ('God')	151
szép ('beautiful')	554	*Krisztus* ('Christ')	146
boldogságos ('blessed')	553	*amen* ('amen')	125
szentlélek ('Holy Spirit')	536	*x* ('x')	124
Isten ('God')	536	*szűz* ('virgin')	90

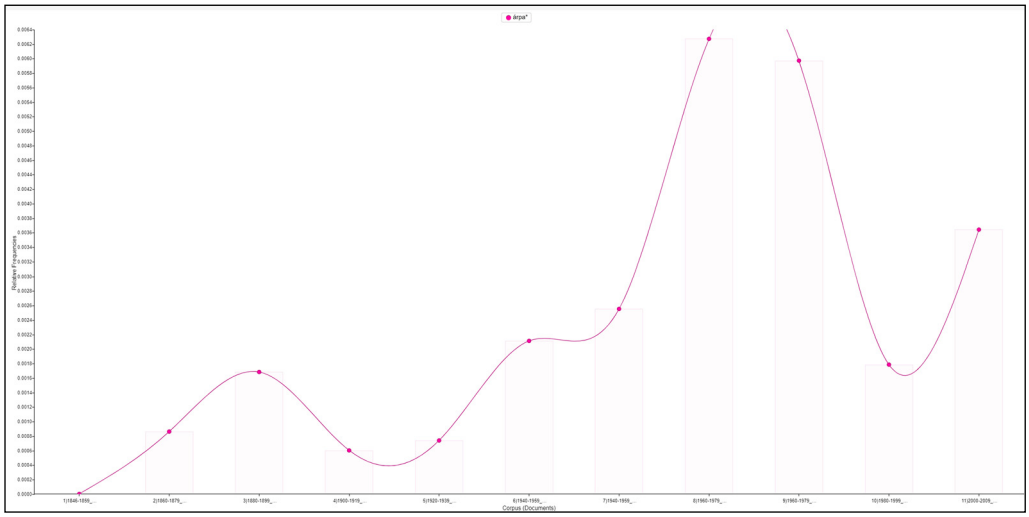

Fig. 10. Distribution of *árpa** word forms in 20-year periods across the collection corpus

ethnographic atlases – most notably, the *Magyar Néprajzi Atlasz* (Hungarian Ethnographic Atlas) – which may significantly distort the representation of certain topics (Ilyefalvi 2021:116–123).

For example, Questionnaire IV of the Hungarian Ethnographic Atlas included a dedicated question (Question 197) on methods for curing styes affecting the eye (Barabás 1958:110).

This connection to the Atlas project becomes even more evident when we examine the presence of the noun *árpa* in the corpora of individual collectors. A comparison of the ten

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

researchers who contributed the most entries to the Hungarian digital charm database shows that *árpa* reaches its peak in the corpus of Éva Pócs, with 181 occurrences.

When we analyze the relative frequency of *árpa** forms excluding the data from the Hungarian Ethnographic Atlas (MNA), the Topography of Hungarian Folk Belief (MNT),[16] and the Ethnographic Atlas of Szolnok County (SZMNA), we get the following distribution (see Fig. 11):

The figure shows that, if the Atlas data are not taken into account, in terms of relative frequency, the occurrence of the term *árpa** is at least as significant in the time period 2000–2009 as in the time period 1960–1979. So far, however, I have only worked with the speech act corpus, but there are known *árpa*-healing texts that do not contain the term *árpa** itself, for example in the case of texts denying the disease or other simpler types of charms. The two examples below illustrate this well. The actual speech act in curly brackets does not contain the term *árpa**:

> < Curing styes > / "Oh! That was done out of superstition, like this – quickly, backhanded. They said: * {Not that! Not that! Not that!} / * Then they slapped the ground with their hand. And then – so they said – you had to walk around the well and throw a grain of barley into it. Then the stye would go away. But mine didn't. Only the doctor healed it."[17]

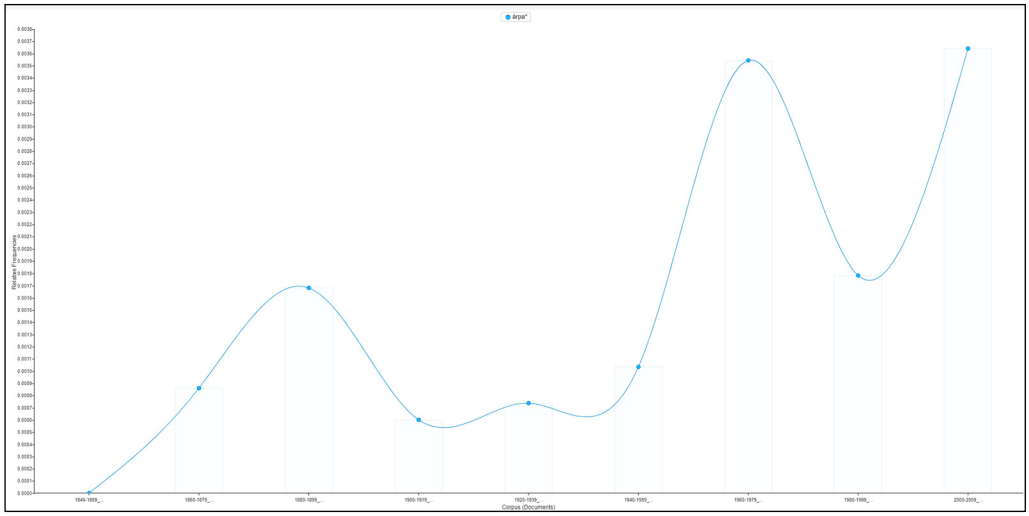

Fig. 11. Relative frequency of *árpa** word forms in the corpus, excluding data from the Hungarian Ethnographic Atlas (MNA), the Topography of Hungarian Folk Belief (MNT), and the Ethnographic Atlas of Szolnok County (SZMNA)

[16]For more on the research project Topography of Hungarian Folk Belief and its manuscript materials, Pócs 2019:114–123.

[17]Kiszombor,1969, collected by Imre Ferenczi. Ferenczi 1977:338; Pócs 2014:187.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

< Curing styes > / "If someone had a stye in their eye, it was »harvested« with a knife. The following was said three times, after that they spat three or four times: * {Let it swell today, / let it shrink tomorrow, / and by the next day let it be gone completely!}"[18]

Therefore, this investigation requires access to the full charm corpus – that is, not only the texts themselves, but also the accompanying ritual descriptions and, more importantly, the titles, which often indicate function. The MicroSearch visualization below (Fig. 12) displays the relative proportions of the subcorpora within the full corpus, as well as the frequency of *árpa*-forms in each subcorpus.

At the same time, in the MNT questionnaire, they asked not only about the *árpa* in the 1960s and 1970s, but also about the previously discussed texts against the evil eye.[19] In what follows, I will attempt to determine whether or not the Atlas collections had a similar effect on the quantity of texts on the cure for styes and against the evil eye.

Both charts (Figs 13 and 14) show that the frequency of *árpa** increases much more sharply in the 1960s than that of *szem/szöm** word forms. However, in the case of *szem/szöm** searches, the results may be distorted by unrelated word forms such as *szemölcs* or *szömölcs* ('wart'), which are irrelevant to the research question. To reduce this distortion, I created the following visualization by combining selected word forms:

"szem|szöm|szemverés|szömverés|szeme|szöme|szemmel|szömmel|szemű|szömű|szemit| szömit."

Fortunately, in all instances, *árpa* appears in the corpus only in word forms that are relevant and semantically appropriate for the analysis (see Fig. 15).[20]

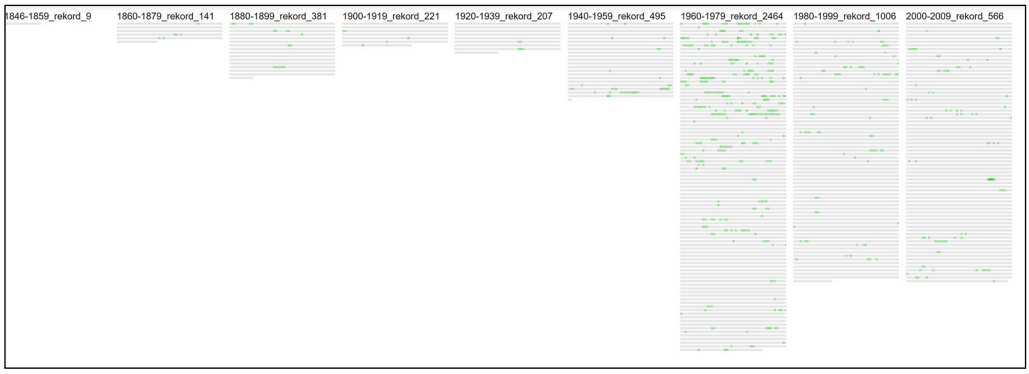

Fig. 12. Occurrences of *árpa** **in the entire corpus by 20-year period**

[18]Gyimesközéplok ((Lunca de Jos, România), 2005, collected by Barnabás Csörge. Pócs 2014:350.

[19]Cf. MNT questionnaire, V.

[20]A total of two pieces of data appear in the database from the Moldavian collection, where the word *jar/korsó* is used to treat *árpa* ('styes' / 'barley'), but due to the smallness of this number, I omitted the combined forms of *árpa** and *korsó** ('jar') in the analysis. Regarding *korsó,* see: Pócs 2014:293; Takács 2015:544–545.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

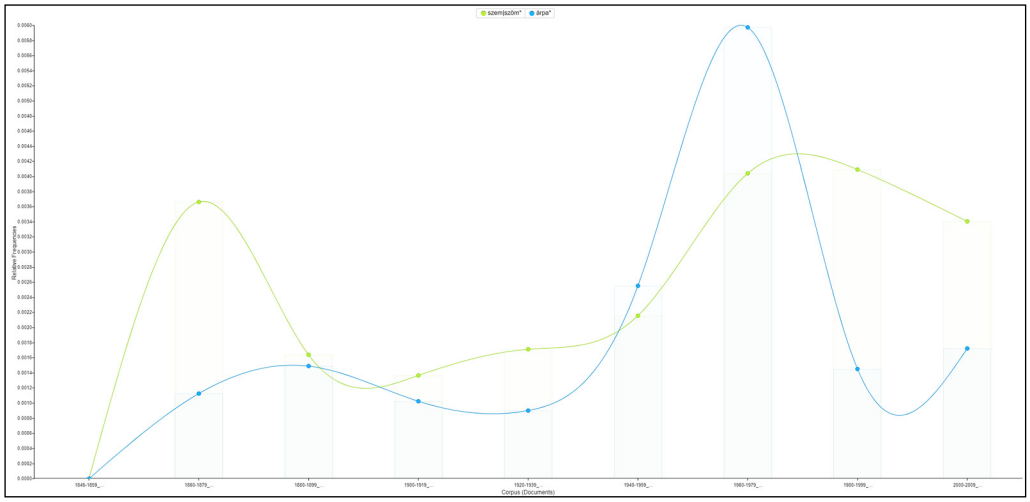

Fig. 13. Distribution of *árpa** and *szem/szöm** word forms in the charms corpus by 20-year period

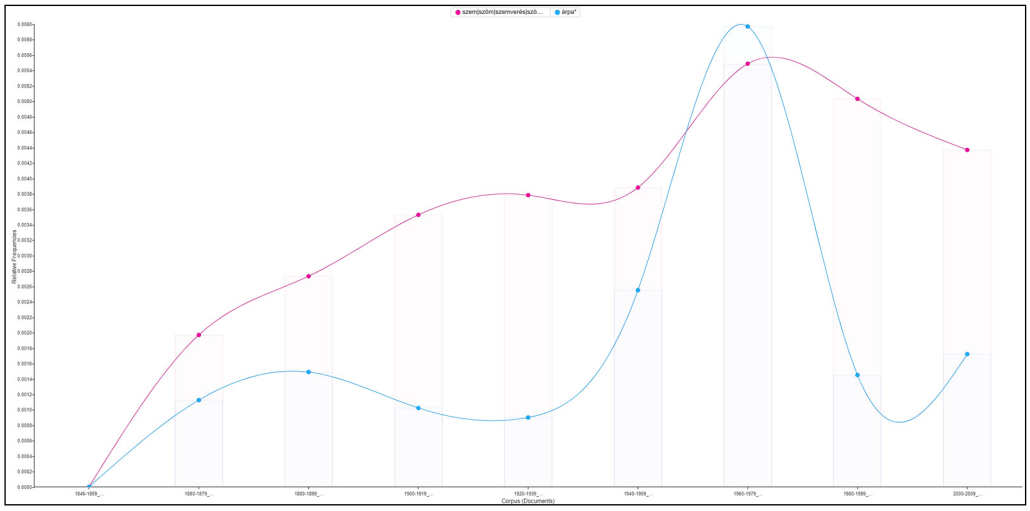

Fig. 14. Distribution of *árpa** and *szem/szöm** extended forms (e.g., *szemverés, szöme, szemmel*) in the charms corpus by 20-year period

The resulting visualization (Fig. 16) further confirms that the number of texts related to the evil eye was less affected by the Atlas questionnaire collections of the 1960s.

Based on the distribution diagrams (Figs 13, 14 and 16), we can draw two conclusions. On the one hand, the number of texts related to the evil eye in the corpus is constantly increasing, and although it reaches its peak in the period between 1960 and 1979, the number of data does

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

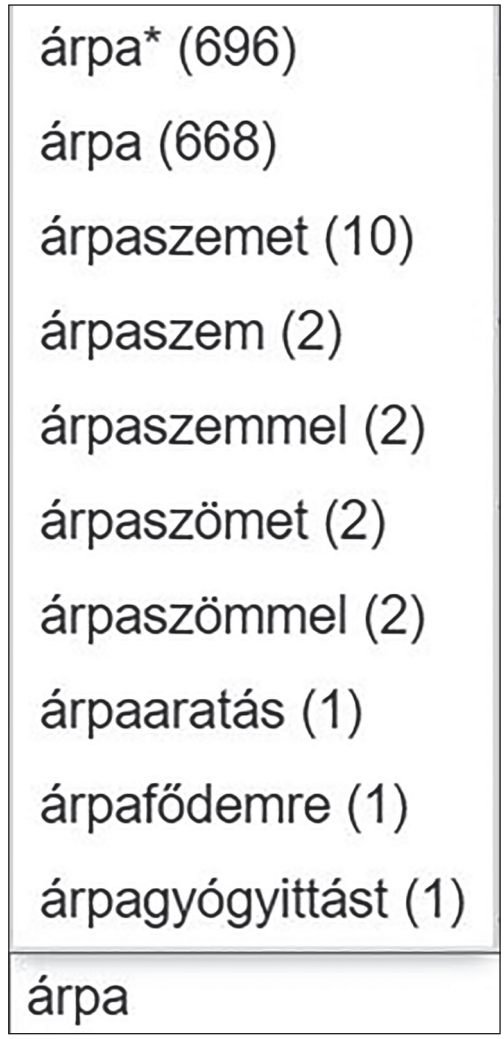

Fig. 15. Concordance view in Voyant Tools' *Contexts* tool showing occurrences and forms of *árpa**

not decrease significantly even after the Atlas research has been completed. On the other hand, in the case of the texts healing *árpa* we can observe that it was the Atlas research that put the type in the forefront of interest, neither before nor after that project was this type intensively collected. As I already noted several times above, the *árpa*-healing texts are sometimes the simplest one-sentence or dialogic texts containing in many cases even obscene expressions, which were widespread throughout the country and known to everyone. Therefore, the data of the corpus do not shed light on the extent to which the texts related to *árpa* or the evil eye were known in different time periods, but rather draw attention to the lack of research interest in

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

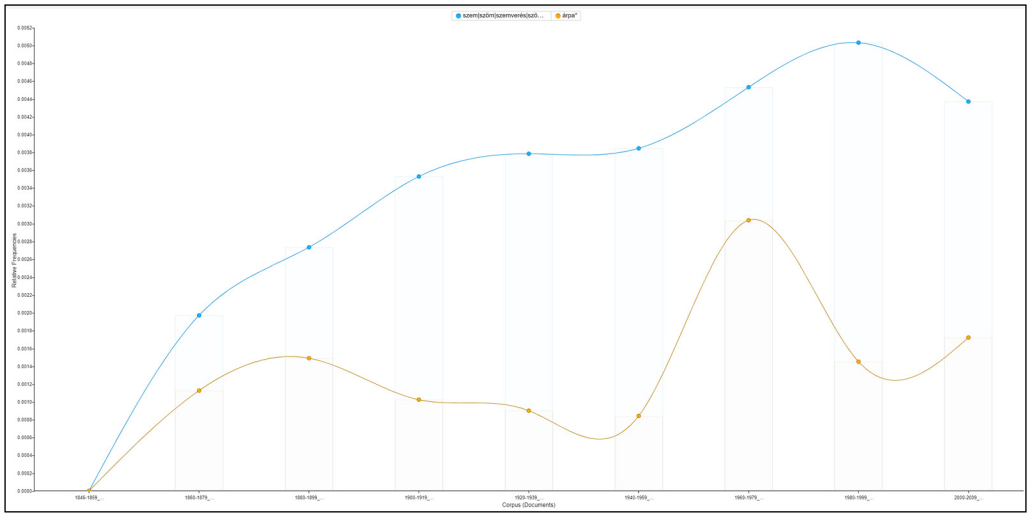

Fig. 16. Distribution of *árpa** and extended *szem/szöm** word forms in the charms corpus (excluding MNA, MNT, and SZMNA), by 20-year period

certain types of texts, as well as its existence. So, it seems that the all-too-well-known *árpa*-healing texts were considered less worthy of being collected or even published by researchers.

Collectors' corpora

In terms of folklore corpora, not only can the methodology of collection, the conceptualization of the research and the period in the history of the discipline be extremely decisive, but also the personality and interests of the collector that also influence the folklore text that is finally recorded.

Filtering by collector allows us to create subcorpora for each individual, enabling us to examine how – and to what extent – their personal preferences, research interests, genre conceptualizations, and collection methods are reflected in their corpora. It also allows us to analyze how each collector's corpus aligns with the regional charm repertoire characteristic of their field site. Continuing the analysis from the previous subsection, the distribution of *árpa*-charms and charms against the *evil eye* is shown below (Fig. 17), based on the subcorpora of the ten most prolific collectors in the Hungarian digital charm database:

Figure 17 clearly shows that Éva Pócs and Vilmos Diószegi stand out in terms of the number of *árpa*-healing texts – a result that is not surprising, given that both participated in the compilation of the Hungarian Ethnographic Atlas.

When we compare this with the distribution excluding the MNA, MNT, and SZMNA data (Fig. 18), it becomes evident that Éva Pócs documented this type of charm independently of the Atlas collections as well.

The peaks in keywords related to the evil eye are strongly associated with the collectors' geographical focus: Takács, Halász, Bosnyák, and Diószegi all conducted fieldwork primarily in

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

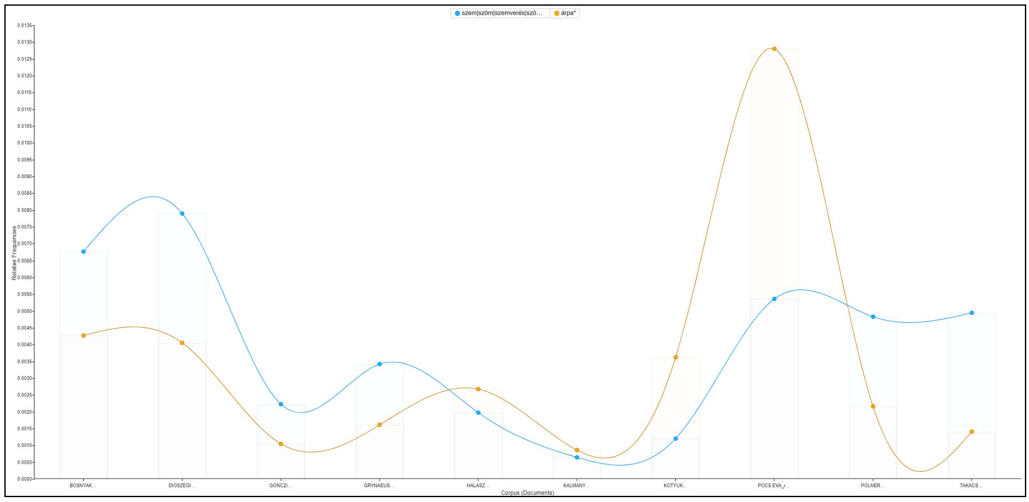

Fig. 17. Distribution of *árpa** and *szem/szöm** word forms in the corpus of the ten most prolific collectors

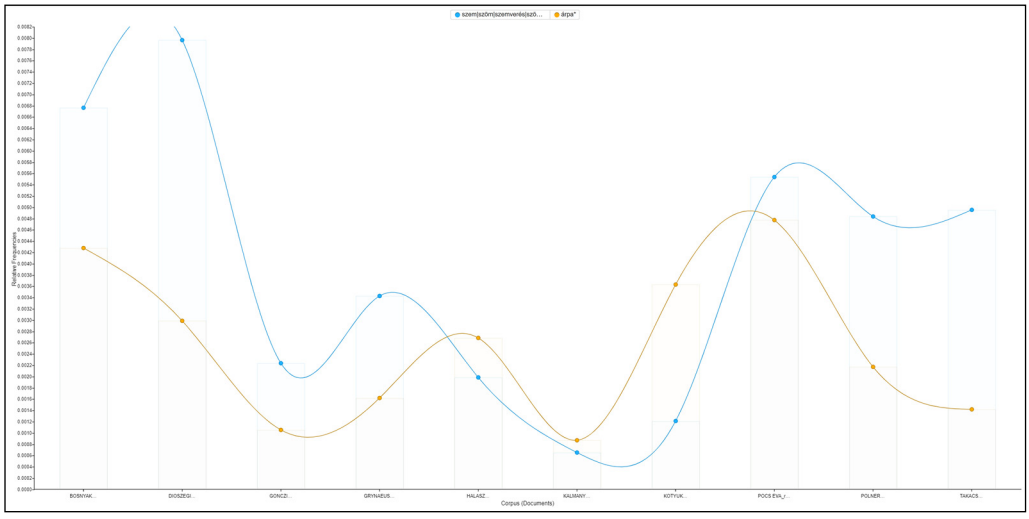

Fig. 18. Distribution of *árpa** and *szem/szöm** word forms in the ten collectors' corpora, excluding MNA, MNT, and SZMNA

Moldavia, Bukovina, and the Gyimes region. (Diószegi worked among Moldavian and Bukovinian Csángó communities resettled in Baranya County, Hungary).

According to Éva Pócs, rituals aimed at preventing or curing the effects of the evil eye – particularly the associated water-sprinkling practices – "are especially intensively preserved

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

among the eastern groups of Hungarians" (Pócs 2014:577–579). As the diagram shows, this pattern is clearly reflected in the collectors' corpora as well.

This pattern can also be observed in collocational network graphs generated from the most frequent words in each collector's speech act subcorpus (Fig. 19). Consistent with the results above, the network graphs of Takács, Halász, Bosnyák, and Diószegi prominently feature keywords typically associated with evil-eye and bewitchment charms – such as *megnézte* ('looked'), *kék* ('blue'), *zöld* ('green'), and *szem* ('eye') – as well as terms that link these texts to longer, more narrative or religious forms (e.g., *boldogságos* ['blessed'], *Mária* ['Mary'], *megtalálkozott* ['encountered']).

In the case of Vilmos Diószegi, we can also observe the prominence of counting charms, which Diószegi also wrote about in separate paper (cf. e.g., Diószegi 1957:162–71) In the corpus of Éva Pócs, in addition to the previously mentioned *árpa*-healing texts, we see the prominence of keywords in texts related to the new moon. We can also relate this to the collections of the MNA, MNT and especially the SZMNA, and to the fact that throughout her research career (also for the creation of the typology of Hungarian charms) she considered it important to study the simpler forms.

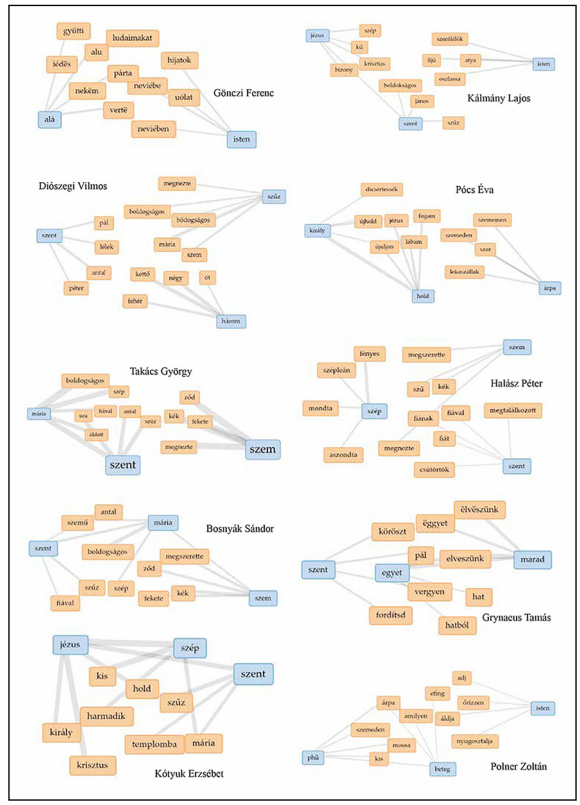

Fig. 19. Collocational network graphs of collectors' text corpora

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

György Takács and the data from Csík, Gyimes, Moldavia and Úz-valley

The corpus of György Takács stands out among the collectors' corpora. Since Takács only collected in Csík, Gyimes, Moldavia and Úz Valley settlements, the question arises: To what degree do Takács' collections give us a different picture of the region's textual tradition? Is the picture the same? In the following, my analyses primarily examine how the data collected by others in the same settlements compare to the data collected by György Takács.

First, I compare the entire corpus of György Takács with a corpus that contains all the texts from the settlements from which György Takács also collected texts.

Regarding the relative frequency of the most common words, there is no significant difference between the two corpora (Table 3). If, however, we examine the descriptions of rituals in the corpora, we can already see a significant difference in that canonical prayers occur in several instances. In the corpus of György Takács's corpus we find many calls for the saying of the Lord's Prayer and the Hail Mary (Fig. 20).[21]

It is also clear from the comparative summary of the descriptions of rituals, that one of the characteristic words of György Takács's corpus is *fejelfúvás* ('head-blowing'). There are no data for *fejelfúvás* at all in the corpus recorded by any of the other collectors. As a matter of fact, the ritual of *fejelfújás* means nothing more than spitting associated with the evil eye (Pócs 2014:610). For this reason, I examined the following possible forms of spitting in the two corpora: *köp**, *elköp**, *eltöp** ('spit'), *elfú** ('blow'), *fejel** ('head-blowing') (Fig. 21).

If we examine the appearances of *fejelfújás* with the help of Context, we can see that in each case they are the informants' own, emic names for the ritual (Fig. 22).

Table 3. Word frequency comparison between the corpus of György Takács and others collected at his field sites

György Takács's corpus	Occurrences	At György Takács's collection sites	Occurrences
szent ('holy' / 'saint')	536	*szent* ('holy' / 'saint')	318
szem ('eye')	457	*szem* ('eye')	196
Mária ('Mary')	263	*jó* ('good')	161
jó ('good')	247	*fekete* ('black')	151
szűz ('virgin')	216	*Mária* ('Mary')	136
igézet ('evil eye')	215	*szűz* ('virgin')	130
boldogság ('blessed')	215	*boldogságos* ('blessed')	111
szentlélek ('holy spirit')	199	*hetvenhétféle* ('seventyseven kinds')	109
gonosz ('evil')	194	*gonosz* ('evil')	106

[21]These are the initial syllables of the most common canonical prayers: the *Miatyánk* (Our Father)), the *Hiszekegy* (Creed), the *Üdvözlégy* (Hail Mary), etc.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

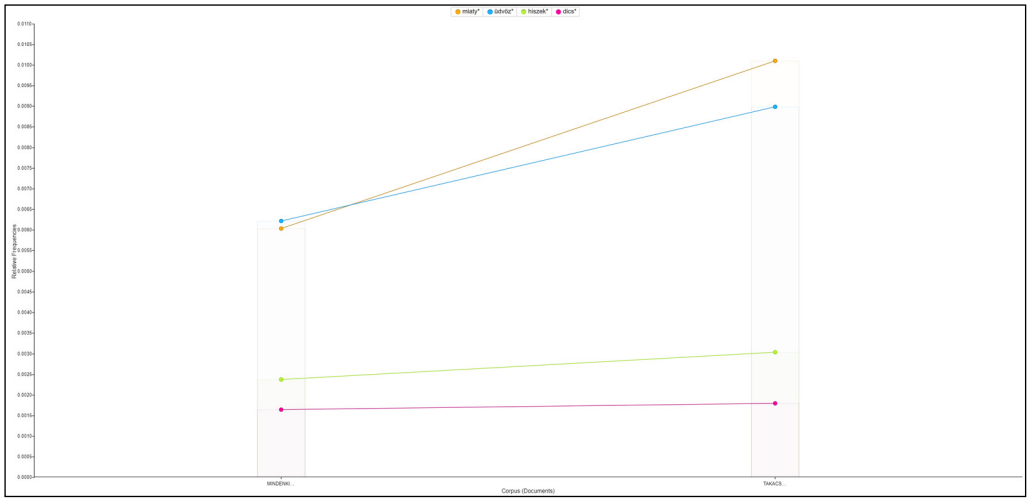

Fig. 20. Comparison of the forms *miaty**, *üdvöz**, *hiszek** and *dics** in the two corpora

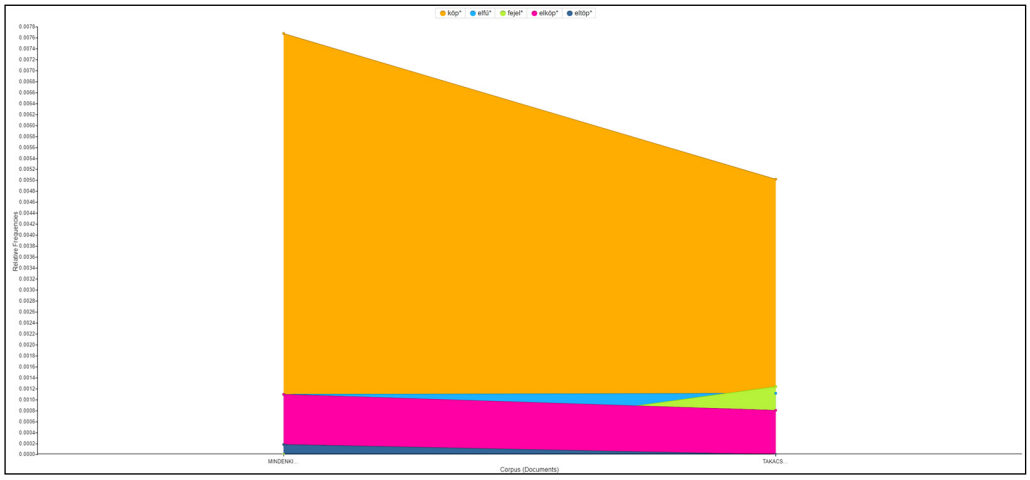

Fig. 21. Distribution of the forms *köp**, *elköp**, *eltöp**, *elfú**, *fejel** forms across the two corpora

On the one hand, it can be deduced from the above that our metadata series containing the descriptions of rituals must be further unified – in such a way that, for example, searching for spitting, the researcher will also find the texts of *fejelfújás*. On the other hand, we can also draw the conclusion that Takács must have paid more attention than his predecessors to noting down and retaining emic terms when providing descriptions of rituals as well.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

Left	Term	Right
megkapja a szemem, hogy itt	fejel	van ez a közkert ki
csiporba kilencszer ereszteni vizet, féligen	fejel	. S akkor a csipor tetejire
nagymamámtol tanóltam. * <Igézet gyógyítása> / * A	fejelfuvás	ugy van, hogy: * / {Kék szem
mondják. […] Csak ezt mondod a	fejelfuvásra	es. […] Az igizet es biztoson
vígasztald meg!} / * Ezt igy mondták	fejelfuvásra	. Keresztet huzott a homlokára, elfujta
nem vettünk vizet. […] S a	fejelfuvást	pedig ugy vettük, hogy keresztet
feje nagyon. Például ezt a	fejelfújást	, ezt kisbabáknál szokták leginkább. Me
János vót, ő ittegyen született. * <	fejelfúvás	igézettől> / * Egyszer szépen vizet merek
fejét!} / * „Na, ennyi. Ez a	fejelfúvás	." * <Zivatar, jégeső elhárítása> / {Az Atyának
tudomesén melyik fejét!} / * Ez a	fejelfúvás	. Vizet es csak igy vetnek
nem tartják ezeket. * <Igizet gyógyítása> / *	fejelfúvással	: * / {Szem megnezte, / szív megszerette, / egye

Fig. 22. Occurrences of *fejelf** in György Takács's corpus

WHAT CAN WE LEARN FROM COMPARISON?

In my paper, I analyzed the digital Hungarian charms corpus with computational textual analysis tools. In the textual analyses, I primarily formulated questions for the digital corpus that were not possible in previous, printed editions. For example, it was possible to run a separate analysis on the speech act texts and the descriptions of rituals associated with them.

Computer analyses offer the opportunity to test and quantify previous scientific conjectures and findings, i.e. to verify or refute statements that (in many cases intuitively) have already been formulated by the discipline but could not be supported by (numerical) data.[22]

The first, exploratory analyses carried out in the paper mostly confirmed the previous findings of research. At the same time, it was also possible to show a number of correlations that we would not have had the opportunity to do without the digital database and computer analysis and reading, and even the raising of certain questions was prompted by the digital tool. We could, for example, show that although texts related to the evil eye and the curing of styes were both included in the Atlas collections of the 1960s and 1970s, this only causes a serious disproportion in the corpus in the case of texts relating to curing styes. Statements and analyses are still rudimentary due to the half-finished state of the database. However, publishing all the data online, completing the typology work, and using the function list, keywords, and digital map will offer opportunities for much more in-depth analysis in the future.

To make the text corpus more analysable, we also face the task of normalizing and lemmatizing the texts. Once that is accomplished, more accurate modeling and corpus analysis will be possible with the help of advanced computational tools. The aspiration of folklore studies – and within it, of charms research – that computer programs should reveal connection points and

[22] According to Balázs Borsos, one of the greatest results of the computer cluster analysis performed on the data of the digitized Hungarian Ethnographic Atlas is, for example, the fact that it has become possible to prove with mathematical methods that Hungarian folk culture is divided into four larger territorial units, and these are roughly the same as previously assumed. As Borsos explains, "the fifth large region defined by the computer, covering the northeastern part of the Great Plain in a geographical sense and the areas bordering Transylvania, called Transitional, also separates from the Central Great Region along a line that was also known in ethnographic research, but due to the geographical unity of the Great Hungarian Plain, it was not accorded the emphasis its significance would have demanded" Borsos 2011:501–502. Supporting/refuting previous research results with computer tools: Wittek 2016:131; Karsdorp 2016:2–3; Labádi 2018:17–20.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

similarities between individual texts, instead of relying on "prefabricated" or "rigid" research categories, can therefore only be realized through future research.

I believe that my analysis of the digital Hungarian charms corpus supports one of the central insights of digital humanities: that distant and close reading are not mutually exclusive, nor do they substitute for each other, but rather function as complementary modes of interpretation. Moreover, digital tools do not primarily provide answers – they prompt us to ask new and previously unimagined questions. The tool used in this study, Voyant Tools, enables what N. Katherine Hayles (Hayles 2012:12) calls hyper reading – a form of enhanced reading that facilitates a continuous shift in focus between micro and macro perspectives.

ACKNOWLEDGMENT

This paper is a revised and expanded version of section 6.2 of my doctoral dissertation entitled Shifting paradigms in Folkloristics: Theoretical, Methodological and Technical Questions of Computational Folkloristics through the Example of the Digital Database of Hungarian Verbal Charms (ELTE BTK Doctoral School of Literature) which I defended in January 2020, cf. Ilyefalvi 2019. Supported by the ÚNKP–23–4 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development, and Innovation Fund and NRDI (National Research, Development and Innovation) Fund, project number 132535, project title: "Folk belief, vernacular religion, mentality, 16th–21st centuries: digital databases, encyclopedic summaries."

REFERENCES CITED

Archer, Dawn
2009 Does Frequency Really Matter? In Archer, Dawn (ed.) *What's in a word-list? investigating word frequency and keyword extraction*, 1–15. Farnham: Ashgate. (Digital Research in the Arts and Humanities).

Barabás, Jenő
1958 *A Magyar Néprajzi Atlasz kérdőíve* I–IV. [The Questionnaire of the Atlas of Hungarian Folk Culture I–IV.]. Budapest: Magyar Tudományos Akadémia.

Borsos, Balázs
2011 *A magyar népi kultúra regionális struktúrája: a Magyar Néprajzi Atlasz számítógépes feldolgozása fényében* [The Regional Structure of Hungarian Folk Culture: in the light of the computer analysis of the Hungarian Ethnographic Atlas.]. Budapest: MTA Néprajzi Kutatóintézet.

Bosnyák, Sándor
1982 A gyimesvölgyi magyarok hitvilága [Belief system of Hungarians in Gyimes]. In Hoppál, Mihály (ed.) *Folklór Archívum* 14. 68–157. Budapest: MTA Néprajzi Kutató Csoport.

Daczó, Árpád
2003 *Hosszú utak megszomorodának…: archaikus népi imádságok, ráolvasók, szentes énekek Erdélyből és Moldvából* [Long Roads have become Sad…: Archaic Folk Prayers, Charms, Holy Songs from Transylvania and Moldva]. Budapest: Magyar Napló.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

Diószegi, Vilmos
1957 Visszafelé számláló ráolvasások [Counting back charms.]- *Néprajzi Közlemények* 2(3–4): 162–171.
Ferenczi, Imre
1977 *Ráolvasások, archaikus népi gyógymódok a Maros mentéről* [Charms and archaic folk medicine from Maros region]. Szeged: Szegedi Ny.
Hayles, N. Katherine
2012 *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago Press.
Ilyefalvi, Emese
2014 *Ráolvasások. Gyűjtemény a történeti forrásokból 1488–1850* [Incantations. Collection from the Historical Sources (1488–1850)]. Budapest: Balassi.
2017 Textualization Strategies, Typological Attempts, Digital Databases: What is the Future of Comparative Charm Scholarship? *Incantatio* 6: 37–77. https://doi.org/10.7592/Incantatio2017_6_Ilyefalvi.
2019 *Léptékváltás a folklorisztikában: A számítógépes folklorisztika elméleti, módszertani és technikai kérdései a digitális magyar ráolvasás-adatbázis példáján keresztül* [Shifting paradigms in Folkloristics: Theoretical, Methodological and Technical Questions of Computational Folkloristics through the Example of the Digital Database of Hungarian Verbal Charms]. Phd Thesis. Submitted: 2019, defended: 2020. https://edit.elte.hu/xmlui/handle/10831/50760
2021 Distant Reading of the Metadata of the Digitized Hungarian Charm Corpus. *Incantatio* 9: 113–132. https://doi.org/10.7592/Incantatio2020_9_Ilyefalvi.
Karsdorp, Folgert
2016 *Retelling Stories: A Computational-Evolutionary Perspective.* PhD thesis, Radboud Universiteit, Nijmegen.
Labádi, Gergely
2018 Az olvasó gép: Berzsenyi Dániel versei távolról [The Reading Machine: A Distant Reading of Dániel Berzsenyi's Poems]. *Digital Bölcsészet* (1):17–34. https://doi.org/10.31400/dh-hun.2018.1.126.
Pócs, Éva
2014 *Ráolvasások. Gyűjtemény a legújabb korból, 1851–2012* [Incantations. Collection from the Modern Age (1851–2012)]. Budapest: Balassi.
2019 A Magyar Néphit Archívum keletkezéstörténetéhez [Data on the Genesis of the Hungarian Folk Belief Archive]. *Ethnographia* 130(1):114–123.
Ramsay, Stephen
2014 The Hermeneutics of Screwing Around; or What You Do with a Million Books. In Kee, K. (ed.) *Pastplay: Teaching and Learning History with Technology*, 111–120. Ann Arbor: University of Michigan Press.
Rockwell, Geoffrey – Sinclair, Stéfan
2016 *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. N. p.: The MIT Press.
Salamon, Anikó
1975 Gyimesi mondák [Belief legends from Gyimes]. In Kákosy, László (ed.) *Mai népi hiedelmek, ELTE Ókortörténeti Tanszékeinek Kiadványai*, 65–156. Budapest: ELTE.

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC

Takács, György (ed.)
2015 *Elindula boldogságos szép Szűz Mária: ráolvasók a régi Csíkszékről* [Blessed Virgin Mary has set out: charms from old Csíkszék]. Budapest: L'Harmattan.
Wittek, Stephen
2016 Distant Reading Conversion in Early Modernity. *Journal of Cognitive Historiography* 3(1–2): 119–33. https://doi.org/10.1558/jch.31658.

Emese Ilyefalvi earned her PhD in 2020 as part of the Hungarian and Comparative Folkloristics Program at the Doctoral School of Literature, Eötvös Loránd University (Budapest, Hungary). She is currently an assistant professor in the Department of Folkloristics at ELTE. She is the author of the book Ráolvasások a keresőben. A folklórszövegek digitalizálásának elméleti, módszertani és technikai kérdései [Charms in the Search Engine: Theoretical, Methodological and Technical Issues in the Digitization of Folklore Texts], published in Budapest in 2023.

Open Access statement. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated. (SID_1)

Brought to you by MTA Könyvtár és Információs Központ olvasók | Unauthenticated | Downloaded 12/17/25 08:29 AM UTC