RUNNING AUTOMATIC TRANSCRIPTION FOR THE MOST CHALLENGING SOURCES: TABULAR PARISH REGISTERS FROM TRANSYLVANIA

Luminița DUMĂNESCU*
Angela LUMEZEANU**
Nicoleta HEGEDÜS***

Abstract:
The paper explores the challenges and adaptations required for applying Transkribus to the parish registers from Transylvania, used as main sources for the HPDT and the complexity introduced by linguistic, alphabetic, and structural diversity inherent to the sources. The denominational and ethnic mosaic of Transylvania has resulted in parish registers written in a variety of languages—Latin, Hungarian, German, and Romanian—and using diverse alphabets, including Latin, Cyrillic, and Kurrentschrift. In many instances, these multilingual registers coexist within records from the same village, further complicating the recognition process. The structure of these tables, which can vary significantly, adds another layer of complexity.
The discussion is focused on adapting and developing the Transkribus platform to accurately recognize and convert the handwritten text into a format compatible with database structure. The adaptation process involves fine-tuning recognition models for each specific language, alphabet, and tabular configuration.

Keywords: Transkribus, automatic handwriting recognition, HPDT

Since 2014, the Centre for Population Studies has developed and maintained the Historical Population Database of Transylvania (HPDT), a research tool initially funded through the EEA Grants mechanism (2014–2017). [1] The core sources of the HPDT are parish registers preserved by the churches, covering the period between 1850 and 1914. These records document the key events of an individual's life course from a demographic

* Babeș-Bolyai University, Centre for Population Studies, Romania (luminita.dumanescu@ubbcluj.ro)
** Babeș-Bolyai University, Centre for Population Studies, Romania (angela.lumezeanu@ubbcluj.ro)
*** Babeș-Bolyai University, Centre for Population Studies, Romania (nicoleta.hegedus@ubbcluj.ro)

[1] Luminița Dumănescu et al., "Historical Population Database of Transylvania. Sources, Particularities, Challenges, and Early Findings," *Historical Lifecourse Studies,* 12, (2022): 133-150.

DOI: 10.2478/sucsh-2025-0009

perspective. Until 1895, when Hungarian legislation introduced compulsory civil registration, parish registers functioned as the official records of vital events. Despite their limitations and occasional fragmentary character, they remain highly reliable sources for historical demography.

The database was designed as a faithful representation of its sources, namely the parish registers. A major challenge in constructing the HPDT was the integration of diverse record structures and fields across denominations. The registers incorporated into the database originate from Orthodox, Roman Catholic, Greek Catholic, Lutheran, Reformed, and Jewish communities, and were written in Latin, Cyrillic, and Kurrentschrift scripts. This linguistic and denominational diversity is reflected in the heterogeneous data recorded, since each church employed distinct formats and categories of information. Following best practices in historical demography, the HPDT sought to incorporate all available fields. Throughout the transcription process, the database structure was continuously refined, either by adding new fields to existing tables or by creating auxiliary tables, in order to facilitate data organization and enhance research usability.

After the conclusion of the EEA funding in 2017, the research team at the Centre for Population Studies continued to expand the HPDT, adding new communities, standardizing the information, and implementing data linkage. At the same time, substantial research based on the database has been carried out, with results published in leading journals in the field[2]. By providing access to a large and diverse set of demographic records, the HPDT has opened new perspectives for Romanian historical demography and contributed to making patterns and trends in the population history of Transylvania accessible to the international scholarly community. In addition, through its genealogical interface (hpdt.ro), the database serves as a valuable resource for the wider public interested in family history reconstruction. However, despite sustained efforts to extract high-quality data from historical parish registers, the extraordinary diversity of ethnicities and confessions that characterizes Transylvania made data entry

[2] Elena Crinela Holom, Oana Sorescu-Iudean, Mihaela Hărăguș, "Beyond the Visible Pattern: Historical Particularities, Development, and Age at First Marriage in Transylvania, 1850-1914," *History of the Family*, 2018, 23 (2), 329-358; Luminița Dumănescu, Ioan Bolovan**,** "From the cradle to the grave I am my father's daughter!' Women and their married names in Transylvania in the second half of 19th century," *The History of the Family,* 2021, 26(3), 466-481; Elena Crinela Holom, Luminița Dumănescu, Daniela Mârza, "Occupations and Social Class Transformations in Two Mining Areas in Transylvania (1850-1910)," *Romanian Journal of Population Studies*, 2023, 17 (2), 5-24.

an extremely painstaking task. The process relied heavily on the expertise of individual transcribers and required a considerable investment of time, often disproportionate to the amount of material that could subsequently be used for analysis. Since the level of financial and institutional investment available during the initial development of the Historical Population Database could no longer be maintained, the team turned to Handwritten Text Recognition (HTR) solutions, now widely employed to convert historical records into structured data.

In 2025, we started to implement a new project[3] intend to adapt and develop Transkribus in order to be able to recognize the handwriting text from parish registers and to transfer it in a specific format, compatible with the database structure. Transkribus is a platform for computer-aided transcription, recognition and retrieval of digitized historical documents Transkribus operates on the basis of artificial neural networks, most prominently recurrent neural networks (RNN) in combination with state-of-the-art deep learning techniques for Handwritten Text Recognition (HTR). At its core, the system employs Long Short-Term Memory (LSTM) networks, specifically designed to capture sequential dependencies within textual data, a capability essential for historical sources in which context strongly influences meaning. The models are trained on extensive datasets of manuscripts and printed documents, consisting of images paired with verified transcriptions. Through this process, Transkribus develops the capacity to associate visual patterns—such as letter shapes, ligatures, and stylistic variations of handwriting—with corresponding characters and sequences. More recent iterations of the system integrate Convolutional Neural Networks (CNN) for feature extraction, combined with LSTM architectures for sequence modeling, resulting in a hybrid CNN-LSTM framework comparable to those employed in speech recognition and machine translation. Within the context of digital history, such supervised machine learning approaches are particularly valuable: the more accurate and extensive the manually transcribed corpus available, the more effectively the models learn, thereby enabling historians to process large collections of historical documents and opening new pathways for quantitative and qualitative research. In order to understand the processes of text recognition it is important to mention that the technology behind Transkribus are artificial neural networks, more

[3] The project *From parish registers to digital infrastructures: the development a HTR solution for automatic transcription of civil status church books* (https://htr-hpdt.granturi.ubbcluj.ro/).

specifically recurrent neural networks (RNN), combined with modern deep learning techniques specialized in Handwritten Text Recognition (HTR)[4].

Transkribus provides models for both specific types of handwriting and alphabets, as well as for specific languages. What mattered to us when working with our sources was not just the ability to recognize script and language: there is also needed to have the data in a format fully compatible with our database. This was a much harder task than it sounds, and much more difficult than the recognition of flowing, narrative text. Transkribus launched this option in 2024 and it is now completely usable. This paper addresses the challenges that automatic transcription may entail when the sources themselves inherently present a range of difficulties.

In the process of project implementation two specific challenges occurred: the first one – which have considered as being the most important during the application phase – is to develop or adapt new language and alphabet models that will enable the automatic recognition of parish records written in the main languages of modern Transylvania. This task is complicated by the variety of alphabets, and their evolution through the time. For example, the transition from writing in Cyrillic to writing in the Latin alphabet was not sudden and included a phase called *transition,* which represents a mixture of the two alphabets, and which will give us a lot of headaches in the coming months. The registers written in Cyrillic characters themselves will require a further stage of processing with AI for transliteration into the Latin alphabet.

The second one, is concerning the transfer of the new data into the population database. This is also derived from the very nature of the source. These records were filled in at the discretion of parish priests, even though, at least after 1850, the tabular format prescribed a specific structure. In practice, priests completed them in whatever way they found most convenient or appropriate—employing abbreviations and contractions, or, conversely, turning the register into a kind of parish chronicle. While the latter is undoubtedly valuable for social history, it poses significant challenges for the transfer of information into a structured database. It should nevertheless be emphasized that the urban registers are markedly superior to those originating from rural areas, a distinction evident from the outset in the much higher qualifications of their compilers: graduates of higher education, proficient in the languages spoken in Transylvania, and, in some cases, true scholars. Moreover, many of them held positions within the ecclesiastical hierarchy, which likely

[4] Guenter Muehlberger et. al, ”Transforming scholarship in the archives through handwritten text recognition,” *Journal of Documentation* 75 No. 5 (2019), 954-976.

fostered a form of what we might now describe as good practice: if the archpriest himself failed to complete the registers properly, how could he have expected a subordinate priest to do so? To create structured data from (almost) narrative text - that's a challenge! The second part of the paper shows that our current solutions for generating structured, database-compatible data with minimal manual intervention remain inadequate. Further work is therefore required to identify more effective approaches.

The "Transylvanian" models developed in Transkribus

The project started its pilot phase in January 2025 and focuses its first efforts on the city of Cluj. The reasons for this choice are related mainly to the fact that the "manual transcription phase" does not include any big cities of Transylvania. The communities of such cities were large and the possibility of covering their registers was limited by the time-cost algorithms. At the other hand, the city of Cluj displays a variety of features that maximize future social history research: during the Middle Ages, the city of Cluj ("the treasure city") was inhabited primarily by Saxons and Hungarians, gradually evolving into the principal urban center of Transylvania. From the mid-16th century, Cluj emerged as a leading center of the Protestant Reformation in Transylvania. Many inhabitants adopted the Calvinist or Lutheran faith, while native-born Ferenc Dávid founded Unitarianism, a denomination distinctive to the region. Between the 16th and 18th centuries, Cluj intermittently served as the de facto capital of Transylvania; in 1790, the provincial government relocated there permanently, a status it retained until 1867.

Economic, cultural, and demographic growth characterized the city from the 16th to the 19th centuries. In 1872, it became home to Hungary's second university, the Franz Joseph University. The population grew from approximately 12,000 in the mid-18th century to 32,831 in 1880 and 67,733 by 1910, the majority (over 51,000) being ethnic Hungarians, alongside 8,886 Romanians and 1,673 Saxons. The 1920 census, when the city's population exceeded 85,000, recorded around 11,000 Jews[5].

Religious affiliation in Transylvania often reflected ethnic identity: Hungarians were largely Reformed (Calvinist or Unitarian) or Roman Catholic; Romanians were Orthodox or Greek Catholic; Saxons were Lutheran. In 1910, Cluj's population comprised 21,045 Reformed, 19,129

[5] Varga E. Árpád (ed.), *Erdély etnikai és felekezeti statisztikája. IV. Fehér, Beszterce-Naszód és Kolozs megye. Népszámlálási adatok 1850 - 1992 között* (Budapest – Csíkszereda: Múltunk könyvek, 2001), 667.

Roman Catholics, 9,136 Greek Catholics, 2,318 Orthodox, 1,930 Unitarians, 2,026 Lutherans, and 7,076 Israelites[6].

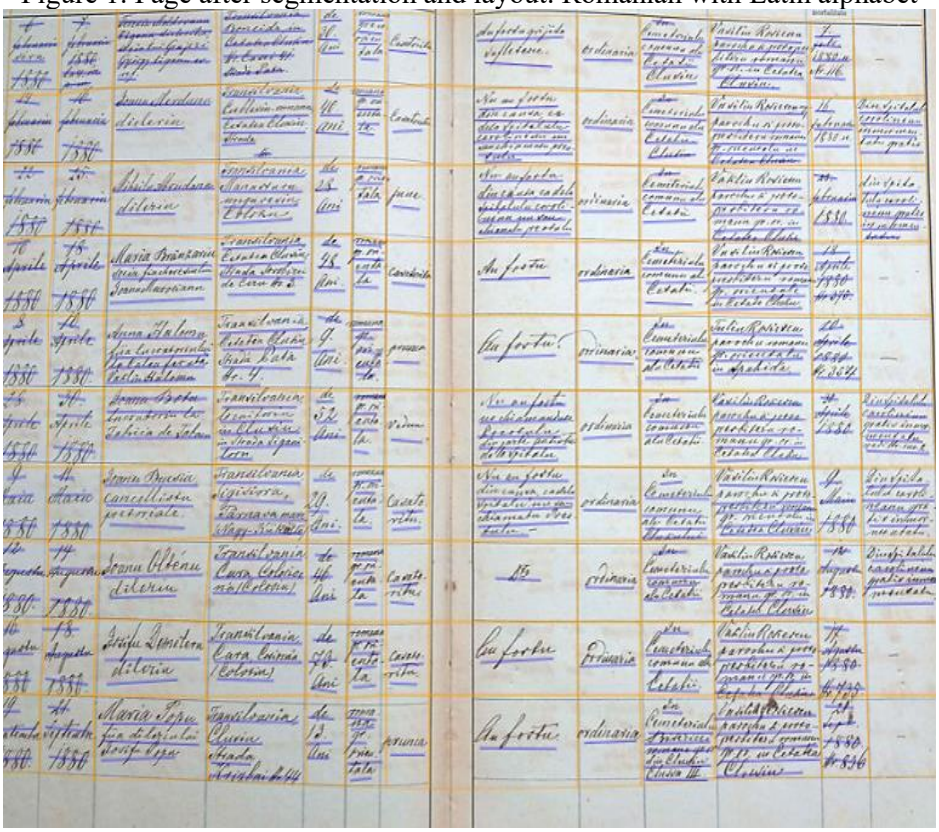
The registers kept into the National Archive of Cluj totalize a number of 15.000 pages, coming from 80 registers, belonging to six denominations and written in 4 languages – basically a representative Transylvanian town. Considering the expected impact of the project we chose Cluj as being the most representative city for this pilot phase. Once those 80 registers will be transcribed, we'll extend the process to the cities of Brașov and Sibiu.

Steps and results. Once the financial flow was assured, the first important step was to secure the subscription to Transkribus for the entire period of implementation. We agreed upon a specific plan for the team which entailed the acquisition of 50.000 credits and the entire package involved in table recognition/transcription/retrieving information. We benefited from a one-month free trial (now reduced to only a few days and without all the features), for each member of the team so we were quite familiar with the platform before subscription. The online webinars and the tutorials uploaded on their YouTube page, as well as the private meeting with one of the Transkribus representatives allowed us to learn as quickly as possible how to use it. But soon it became obvious that the complexity of the sources will create unexpected problems: for instance, we realized that there was no use to train a table model for each event since the probability of applying the model was reduced at most to one register. In most of the cases (it seems that for the Lutherans and the registers filled in German there is possible to apply the same model for each event) no model could be used from one register to another, for multiple reasons, some attributed to the physical state of the source or to the (low) quality of the scans, other to the contents itself: even where the table structure is the same, the incapacity of the priest to fill the information in dedicated fields compromise any attempt to use a model. This means that we are spending valuable time on preliminary steps, like segmentation, which basically consist in marking the regions and drawing tables.

The text recognition was based, at first, on the existing models: the super models like "Titan I", for Latin and Romanian with Latin alphabet, "German Giant" for *Kurrentschrift*, Hungarian model developed by the Szechenyi National Library from Budapest. Currently, the text recognition is based on private models developed for Romanian with Latin alphabet and Hungarian and publicly available models for German and Latin.

[6] Varga, *Erdély etnikai és felekezeti statisztikája*, 667.

Figure 1. Page after segmentation and layout. Romanian with Latin alphabet



The Romanian model. Without any doubt, to develop a model for Romanian was the most difficult task by now, since there was no previous model for this language to base on. We start from ground 0 in what concerns text recognition, and the journey starts from almost 26% CER (Character Error Rate, which expresses the share of characters that are recognized erroneously and need to be corrected). In six months, the model (which involved 12 training cycles) was constantly improved, and we are currently working with a version with 5,67% CER - still much to correct, but satisfactory. The current version is based on 270.000 words (including 10% validation data), but we must acknowledge that an important proportion of the words are repetitive (specific for serial sources) like those for denomination, locality, name of the priests and of the midwives, civil status and so on.

Figure 2. The evolution of training runs for Romanian alphabet

MODEL TITLE	LANGUAGES	WORDS	LABELS	CER
HTR_HPDT_4	Romanian; Moldavian; Moldovan	252 346		5.67%
HTR_HPDT_3	Romanian; Moldavian; Moldovan	182 141		8.77%
Model NameHTR_HPDT_2.1	Romanian; Moldavian; Moldovan	156 383		12.97%
HTR_HPDT_2	Romanian; Moldavian; Moldovan	96 114		13.27%
HTR_HPDT_1	Romanian; Moldavian; Moldovan	70 405		14.02%
HTR_HPDT_Ro_5	Romanian; Moldavian; Moldovan	48 328		13.04%
ROGrCat_4	Romanian; Moldavian; Moldovan	28 124		15.63%
Ro_Gr_Cat_v3	Romanian; Moldavian; Moldovan	13 524		24.53%
RoGr_Cat_V2	Romanian; Moldavian; Moldovan	7 698		19.40%
RoGrCat_v1	Romanian; Moldavian; Moldovan	4 990		25.44%

Figure 3. HTR_HPDT_4

Although it still takes quite some time for correction after text recognition, the current model is faster than its predecessors. The biggest and most time-consuming problems continue to be caused by segmentation

and layout operations, which often need to be adjusted manually. Due to the nature of the registers, there is no hope that these issues will be resolved in the future, so we can assume that most of the process in the future will be dedicated to these two preliminary operations.

The Hungarian model. Like we already mentioned, at the turn of the 20th Century the Hungarian community represented over 50% of the population of Cluj, and they were Roman Catholics and Reformed (Calvinist and Unitarians). The Roman-Catholic records were generally kept in Latin, while Calvinist and Unitarian registers were in Hungarian—making Hungarian the dominant language in Cluj's parish registers between 1850 and 1918. Since for the Latin we can use, hopefully, one of the super models (Titan for instance) for the Hungarian it was necessary to train a specific model. To improve transcription accuracy, a Hungarian-language recognition model is being developed in Transkribus, trained on a Unitarian protocol (1885–1895) containing baptisms, marriages, and deaths, as well as a Reformed marriage register (1885–1903) and a Reformed death register (1875-1880). Unlike narrative Hungarian texts (there is a Hungarian handwriting recognition model based on around 280.000 words, made public by the Szechenyi National Library from Budapest), parish registers consist largely of recurring data—names, places, occupations, causes of death, dates, and ages—requiring the creation of a specialized recognition model.

The Unitarian protocol, although not very extensive (124 pages), contains, in addition to anthroponyms and toponyms, terminology specific to each type of event (for example, in the death register a range of causes of death are recorded). The Reformed marriage register (given that in a city with about 20,000 parishioners there were multiple parishes) is a compiled copy bringing together the marriages from all Reformed parishes in Cluj, and the handwriting varies within the same document.

Figure 4. Header - Marriage register-Unitarian-1890

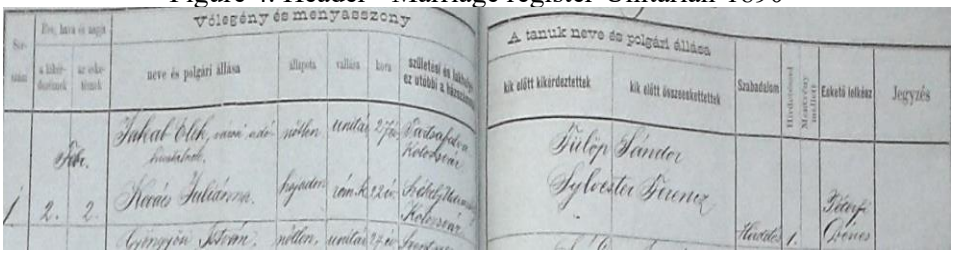
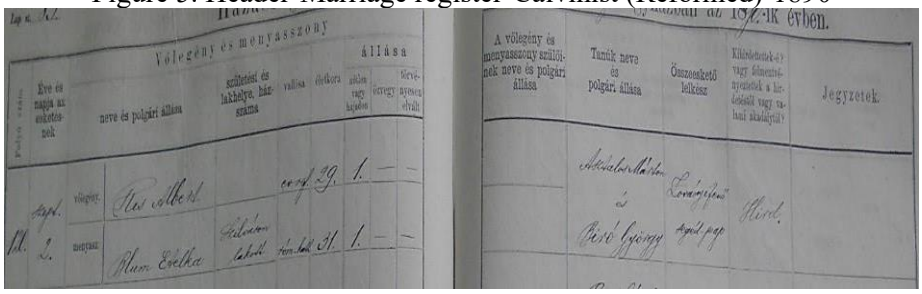


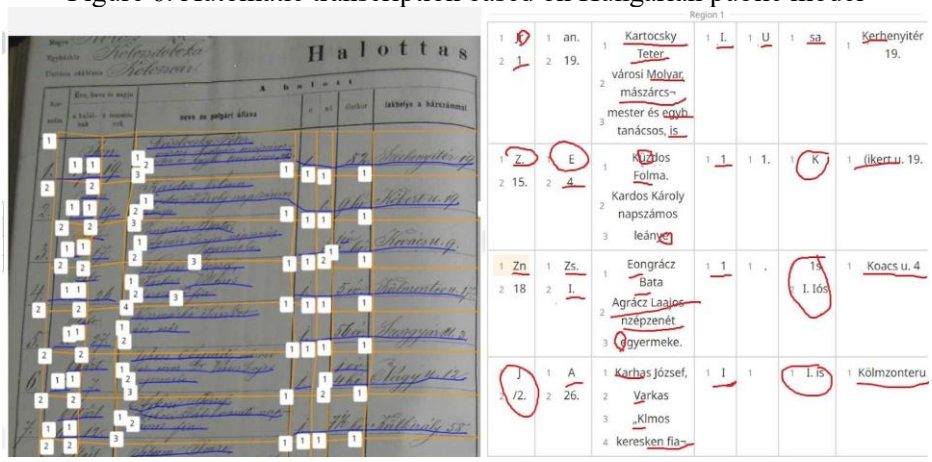
Figure 5. Header-Marriage register-Calvinist (Reformed)-1890



As we can see from this example, the headers from the Unitarian and Calvinist registers differ, although they record the same events, marriages, from the same year, 1890. This is why a universal table model for each type of event cannot be developed.

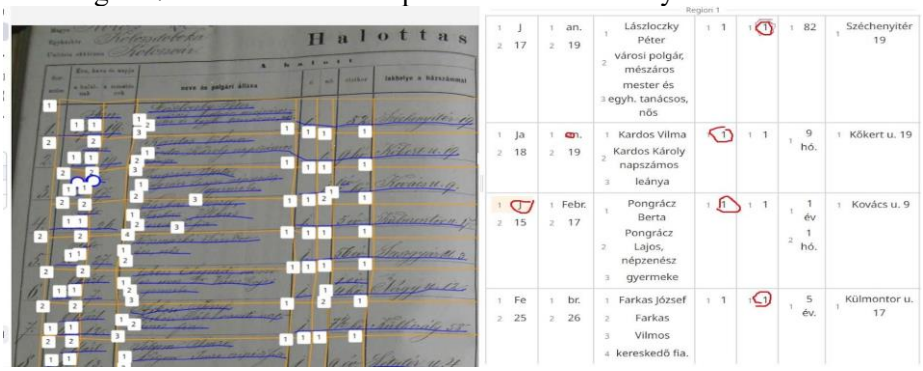
Using the public Hungarian handwritten text recognition model, we can see that most of the words need correction, and the numbers are rarely transcribed correctly:

Figure 6. Automatic transcription based on Hungarian public model



Using the latest version of the text recognition model developed based on Hungarian parish registers from Transylvania, on the same text, the results are much better.

Figure 7. Automatic transcription based on “Transylvanian” model



This model, trained on 100.000 words, has a 2,15% CER and combines handwriting from seven different hands and probably it will be the first model made public as the result of this project.

Figure 8. The Hungarian model developed on HTR-HPDT

The German ”model”. Although the records belonging to the German community in Cluj are very well suited to the public models on

the Transkribus platform, and there is no need to develop a language model, there are nevertheless a few specific issues.

Within the same register entry, one often encounters multiple languages, as recording practices evolved alongside shifts in parish language use. For instance, the Evangelical community in Cluj gradually switched to Hungarian, while also updating older records over time. This circumstance may require either a mixed German–Hungarian model or the application of the Hungarian model to a specific region within the table. A telling example comes from a marriage record of 1848 (fol. 53r), where in 1941 the Reformed parish in the city added information to an earlier entry concerning the father of a baptized child. These complexities also explain why Hungarian names, with their distinctive characters, are often misrecognized by the German Genius model, which lacks them in its vocabulary. Such cases highlight the need for a more accurate tool—one adapted to the linguistic realities of Transylvania.

The further it advances with the table recognition within the source, the more errors it encounters, even though the layout and image quality are identical from beginning to end of source. Parish records have a specific vocabulary, for instance in terms of occupations covered. German records particularly have a wide range of occupations, in a lot of institutional contexts. Because the German supermodels are not usually trained on serial sources that have occupational titles, sometimes when the writing is difficult to decipher, the model has the tendency to recognize something it might deem plausible based on its training data.

In the top of challenges: processing data after Transkribus

As we already mentioned, the most challenging part of the HTR process consists in post processing data to be compatible with an existing population database. After the data extraction from Transkribus, the goal of this process is to create structured data that are compatible with HPDT from semi-organized historical records. The objective is to automate as much as possible and then make manual correction easy for the remaining cases.

For this we have tried several methods:

1. *Using the n8n automation platform and a GPT-4.1-mini AI model to transform semi-structured data into JSON objects conforming to a predefined header structure*.

The base is a Google sheets file with three components:

- original worksheet – the contains the raw data that need processing
- headers worksheet – 2 sets of identical rows that represent the exact structure of the desired fields for the output
- parsed data worksheets – that will contain the structured data after parsing.

The workflow will read the headers first and read all the rows from the original file. It will split data into batches of 10 rows. AI agent sends each batch to GPT that parses the JSON and transforms it into a format compatible with Google sheets and then writes the resulting data into the parsed data worksheet.

2*. Using a Python script.* The core of the script contains several functions, each designed to parse a specific column of a piece of information. They operate on a "heuristic" basis, meaning they use rules of thumb and pattern matching rather than a strict, guaranteed formula. The script follows a Read -> Process -> Write pattern. It reads a semi-organized file, applies a series of custom, rule-based functions to clean and structure the data from key columns, and then writes the result into a well-organized Excel file.

Issues with organizing data. After running the script several times and applying the n8n workflow we identified that there are several problems that need to be addressed to have a well-organized flow of data compatible with HPDT structure.

1. The original data contains text that spans multiple lines with inconsistent formatting. The position of the fields is not consistent. The

number of lines, the order of information, and the keywords used all change from entry to entry.

"Georgiu Tioldosiu /first name, family name
fiulu nelegiuitu /illegitimate child
alu Mariei Tioldosiu, Dileritia" / [son] mother first name/ mother last name /mother's occupation

"Vasiliu Nastanu, /first name, family name
studinte in 1a Clasa normala /deceased occupation
fiul lui Constantinu Nastanu / son of father first name/father last name
Pardositor" /father's occupation

Exemples from Hungarian registers:

"Frankkis Gyula (a nevet a bába / family name, first name (baptized by the midwife)
adta neki), Frankkis Sándor, mészáros /son of: last name, first name of the father/father's occupation
fia kereszteletlen" / not baptized officially
"Elcser Rózália, /last name (maiden name), first name
özv. Losonczi Györgyné /widowed wife of: husband's first and last name
női szabóné" /husband occupation

2. Inconsistent Delimiters - commas and line breaks are used but not regularly. Commas also appear *within* the data fields which is confusing. The inconsistent use of delimiters is a classic problem for and creates ambiguity.

"Anna Rosiescu,
prunca Protopresbiterlui romanu
gr. o. Vasiliu Rosiescu din Clusiu"

"Johanna Camilla, 22.
J. alt, Tochter des Friedrich
Müller, Goldarbeiter"

3. There is an overload of data within a cell that should be split and the text within the columns is complex and often ambiguous. In the column

designated for the deceased name we often find information about occupation, residence, birthplace, marital status. If the deceased is a woman, we also find information about the spouse and sometimes maiden name or if the deceased is a child the father and sometimes mother are also mentioned.

“Szász Anna, toroczkói sz. /Deceased last name, first name, mother’s birthplace
Kolozsvárt szolgáló cseléd Szász /Mother’s occupation and workplace, mother’s last name
Anna törvénytelen leánya” /Mother’s first name, illegitimate child

“Georg, 2 Monat alt
unehel Sohn der Rósa
Schuster, Köchin.”

4. The variation in name formats is a great example of the data's complexity. The program function assumes a "First Name + Last Name" structure, but the actual data isn't nearly that consistent. Especially for infants there are cases where only one name was recorded but is not a rule. More problematic is that some names contain more than two words, and it's not always clear which part is the first name, and which is the last.

“Enlaki Sala Károlina, /Deceased last name, first name and nobiliary particle
néhai Gálfalvi Imre, főgondnok, özvegye” /Widow of: last name, first name, occupation of the deceased husband

“Mari, 5 Wochen alt
Tochter des Daniel
Andraschofsky junior
Kupferschmiedmeist. Allhier”

“Iuliana Eugenia Nasta
prunca lui Constantinu Nasta
pardositoriu”

5. Sometimes the titles are mixed with the names that require a separate rule for splitting the name correctly. And sometimes the nickname is present.

"Ádámosi /Nobiliary particle
Végh Mihály /Last name, first name of the deceased
magyarnádasi földbirtokos özv." /occupation, marital status

"Dr Josef Engel, aus
Wien, geb. 63. J. alt."

6. The handwritten records are a major source of inconsistency. The different spellings for the same word are a classic sign of this, reflecting both the lack of standardized spelling at the time and the difficulty in finding the variations. Exemples: "Wife of" (socia lui / soția lui), "Widow of" (văduva lui / vedua), "Son of" / "Daughter of" (fiul lui / fiulu / fiia /fia). The core issue with these spelling variations is that a script only knows what we tell it to look for, so we need to find all the variations in order to feed the script.

Besides the various spellings there is also the problem of using the abbreviations.

"n. Marosujvar (born)
l. Clusiu Nr 285" (lives in)

Like inconsistent spellings, abbreviations create a "dictionary" problem. The script must be pre-loaded with a list of all possible abbreviations and their full meanings to make sense of the text. If it encounters a new abbreviation (e.g., Spit. for Spitalul (Hospital), it won't understand it and might misinterpret it as part of a name or location, leading to silent data corruption.

7. One of the most challenging problems for any automated script because it requires contextual understanding that is natural for humans but difficult to program is the text that can be interpreted in different ways. Ambiguous text forces the script to assume, and that assumption can sometimes be wrong.

"Kovács János, / last name, first name of the deceased

Kovács János főgymnáziumi /father's last and first name, occupation
igazgató tanár fia"

Kovács is also an occupation and that confuses the script that must make a choice.

"Teodoru Bibolariu
plugariu
cà arestante comitatensu"

8. For Hungarian language the gendered suffixes like -né (the wife of) create further ambiguity in the parsing logic.
"Fekete Albert /last name, first name of the deceased
Özv. Fekete Istvánné fia/mother's married name (widowed wife of: husband's name)
kereskedő segéd s közhonvéd" /occupation of the deceased

Instead of conclusions – hopes.

After 10 months of project implementation, with several training courses on text recognition on the specific languages used in Transylvania and despite the problems exposed in this paper we are enthusiasts in what concerns the outcomes of the project and the benefits for the future of automatic transcription. The only thing is to discover the proper solution for automatic splitting of the data and transferring in HPDT. But the advance of the AI in the last few years, the development of the Transkribus accompanying this development (let's remember that before autumn 2024 there was no solution for tabular documents and now this is fully available) and the perseverance of the team will lead us to the solution sooner or later in this 2.4 years left and, of course, for the years to come, as long as we will have money to buy credits for using the platform. Any other HTR solution, except the fact that it is much beyond our financial capacities, is far from having the millions of characters recognized by this platform behind it, which greatly facilitates the work of any developers of new HTR solutions for languages and alphabets belonging to smaller spaces.

Acknowledgement: This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS-UEFISCDI, project number PN-IV-P1-PCE-2023-0339, within PNCDI IV.

Restaurare / Restoration