# A Survey of the Digitisation of German Newspapers in interwar Lithuania (1918–1940)

**Lina Plaušinaitytė, Heike Zinsmeister**
Department of German Philology, Institute for German Studies
Vilnius University, University of Hamburg
lina.plausinaityte@flf.vu.lt, heike.zinsmeister@uni-hamburg.de

**Abstract**

This paper presents a survey of the preservation and digitisation status of the German-language press published in interwar Lithuania, which existed between 1918 and 1940. In the newly established and ethnically diverse Lithuanian Republic, which was operating in accordance with the European Minority Protection Regime, German-language newspapers and other periodicals formed a relevant part of the country's multilingual press. They represent an interesting yet underexplored resource for historical and linguistic research. The survey summarises bibliographic information and the results of earlier digitisation projects. Although systematic digitisation remains future work, this paper notes some challenges for optical character recognition (OCR) within this collection, particularly in relation to typographic variation, and outlines the envisaged format of the digital resource

**Keywords:** historical German newspapers, interwar Lithuania, digitisation, OCR

## 1. Introduction

Historical newspapers and other periodicals, such as calendars and almanacs, constitute valuable sources for the study of social, cultural, religious, economic and political life in earlier periods, as well as for tracing various development processes over time (Mills, 1981). They also represent an important resource for linguistic research. As publications addressed to a broad public readership, they generally adhere to conventional linguistic standards; however, subtle geographic and temporal influences can be detected comparing significant amounts of digitally accessible data stratified by place and time of publication. Newspaper-based corpora provide evidence for linguistic structures and their diachronic development. At the same time, newspapers' coverage of a wide range of topics provides a rich basis for analysing vocabulary use, including collocations and lexical change (e.g., Pedrazzini and McGillivray, 2022).

Digitisation of historical newspapers often results in facsimiles and low-quality automatic optical character recognition (OCR) that identifies (sub)strings in the image without actually capturing the semantic zones of the page. Further processing of structure and content of newspapers remains largely a matter of debate. OCR of historical papers in general, and newspapers in particular, presents a distinct set of challenges in comparison with modern papers (see, e.g., Springmann and Lüdeling, 2017 for OCR of historical German books published between 1487 and 1914). These problems are due to the material condition of the paper or ink, as well as to typography and layout complexities. The latter are particularly prominent in newspaper settings, where pages are partitioned into different articles with varying numbers of columns, and interspersed images, figures, etc. (e.g. Barman et al., 2021). Newspaper texts also include different font types and script variations. The historical blackletter Fraktur font (also referred to as 'Gothic' script) is particularly challenging for OCR (e.g. Généreux et al., 2014; Bjerring-Hansen et al., 2022). Advertisements in newspapers distinguish themselves from ordinary articles and often employ creative typographic layouts, including non-horizontal lines. Last but not least, historical newspapers of the diaspora exhibit not only linguistic variation due to diachronic change, but also code-switching and interference from local contact languages. These characteristics pose considerable challenges for OCR and further processing.

Despite these challenges, a number of projects provide search interfaces and download options for historical OCRed German-language newspapers, in particular the *German Newspaper Portal* (DZP), *Austrian Newspaper Online* (ANNO), and *Deutsches Textarchiv* (DTA). The *Digitales Forum Mittel- und Osteuropa* (DiFMOE) specializes in periodicals of Central and Eastern Europe. Recently, several projects have advanced research on the digitisation of historical newspapers and the application of natural language processing (NLP) tools to them; see the compilation in Ridge et al. (2019). The Swiss-based Impresso project (Ehrmann et al., 2020), for example, supports community building. In addition to a corpus collection, the project also hosts a DataLab platform with NLP resources accessible via Jupyter notebooks that can be applied to the user's own research data. Another more recent large international effort is the European project PressMint[1], which aims at compiling com-

[1] https://www.clarin.eu/pressmint



*Proceedings of the First Workshop on Creating Interoperable Corpora of Historical Newspapers (PressMint 2026)*, pages 65–71

parable, interoperable, and annotated corpora of European historical newspapers for about the last 125 years.

None of the aforementioned resources include historical German-language newspapers from interwar Lithuania, which are the focus of the present study.

## 2. Historical contextualisation

To contextualize the German-language press in interwar Lithuania between 1918 and 1940, it is necessary to give a brief historical outline of the political development of the region that had a strong influence on languages and publication efforts (see, e.g., Plaušinaitytė (2021) and the literature cited therein). While playing a prominent political role in Eastern Europe in the medieval and early modern periods, Lithuania lost its independence as part of the Russian empire at the end of the 18th century. Only after the end of World War I was its political independence restored, and the formerly suppressed Lithuanian language became the national language again. In accordance with the European Minority Protection Regime, minorities were allowed to use their own languages, such as German, Yiddish, or Polish in public life, churches, schools, and even state administration. This was revolutionary for a region in which schools were only taught in Russian for over a century.[2] Figure 1 shows the political setting of Lithuania in 1939-40 with Germany (former Prussia and German Reich) in the South-West and Polish territory (modern Belarus) in the South-East. The bright yellow area is core Lithuania with Kaunas as its interim capital. The black outline marks the modern state of Lithuania since 1990.

Two of the orange-coloured areas require additional explanation. The Klaipėda region in the West (former German *Memelland*) became part of Lithuania only in 1923. It was a partly German-speaking area, because it had belonged to Prussia (or, later, the German Reich) for many centuries and had never been part of the Russian Empire. The Vilnius region in the South-East including the historical (and modern) capital of Lithuania was disputed after 1918 and came under Polish rule in 1920 until 1939.

The democratic nation of Lithuania turned into a right-wing dictatorship in 1926 and ended in 1940.[3] This is also the end of the German-language publications in the area.

In this survey, we are mainly interested in the German-language press that was published by the German minority in core Lithuania. In contrast to the Klaipėda region, their ancestors were mostly artisans and merchants invited to migrate to Lithuania in the late medieval ages, or Lutheran Christians expelled from the Austrian Salzburg area and invited to stay in neighbouring East Prussia in the 18th century, settling on both sides of the border. In the 19th century, a number of German workers also arrived in Lithuania to work on railway construction and in metal processing factories.

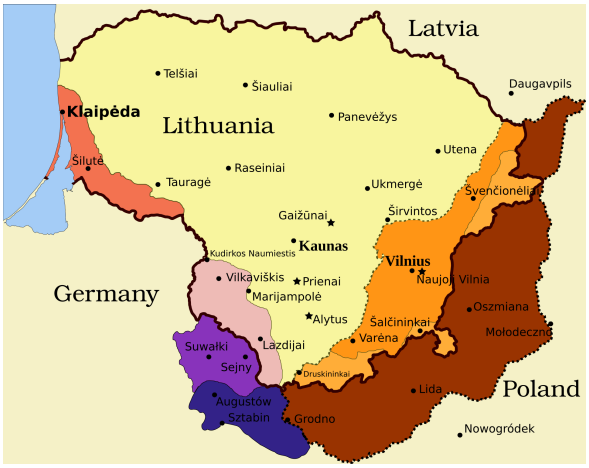


Figure 1: Map of territorial disputes and claims regarding Lithuania in 1939-1940. For the current survey the (yellow) core region is most relevant (image by Renata3, CC BY-SA 4.0 via Wikimedia Commons)

## 3. Bibliographical preservation

To explore the preservation of relevant newspapers, we consulted bibliographies and libraries' collections in Lithuania and Germany, as well as those of the Library of Congress in Washington, D.C. Table 1 summarizes our findings for German-language newspapers distributed in Kaunas in 1918–1940. Table 2 in the Appendix provides a more detailed report for one of the newspapers. The individual sources are briefly introduced in the rest of this section.

A specialised four-part bibliography of German-language periodicals from Eastern Europe, including Lithuania, has been compiled at the Regensburg *Leibniz Institute for East and Southeast European Studies*. The volumes contain bibliographic information on newspapers and journals (Weber, 2013a), popular calendars, almanacs, and yearbooks (Weber, 2013b), and also research publications on the German press in Eastern Europe (Weber, 2013c). Weber (2013a) mentions 55 German-language newspapers and journals published in Lithuania during the relevant years from 1918 to

[2] Even the printing of Lithuanian books in Latin script had been criminalized in the Russian empire.

[3] It was first under brief Soviet occupation, then occupied by Nazi Germany. From 1944 to 1990, it was part of the Soviet Union.

| Newspaper | Years | Source | Comment |
|---|---|---|---|
| *Litauische Rundschau / Lietuvos apžvalga* | 1920–1921, 1924–1929 | LNB Vilnius, VL Vilnius, ZDB (and Carlton 1965) | for details on preservation and digitisation, see Table 2 |
| *Deutsche Nachrichten für Litauen / Vokieciu Zinios Lietuvoje* | 1931—1940 | LNB, ZDB, Weber 2013a, Carlton 1965 (starting 1930[!]) | published by the German-Lithuanian Cultural Association; 151 issues digitized by VU Vilnius |
| *Kownoer Zeitung / Soldatenrat Kowno* | 1918 | LNB Vilnius, ZDB, Carlton 1965 | Kowno = Kaunas; initiated in 1916å |
| *Die neue Zeit: Organ des Soldatenrates Kowno* | 1918—1919 | LNB Vilnius, ZDB | |
| *Korrespondenz B* | 1918 | ZDB, Weber 2013a | 'Reports [...] from the administrative territory of the Oberbefehlshaber Ost', since 1916 |
| *Baltisch-Litauische Mitteilungen* | 1918 | ZDB, Weber 2013a | successor of *Korrespondenz B* |

Table 1: German-language newspapers published in Kaunas, the interim capital of Lithuania, between 2018 and 1940 (abbreviations are explained in the main text)

1940, only three of which were published in the interwar capital Kaunas (see Table 1), nine in Vilnius, and the remaining 41 in the Klaipėda region.

Copies of German newspapers have been systematically collected at least since 1912 when the *Deutsche Bücherei* was founded by the German Booksellers Society among others, which developed into the modern *Deutsche Nationalbibliothek* that hosts the German Union Catalogue of Serials (*Zeitschriftendatenbank* ZDB) which "is the largest dedicated database for serials of all kinds, particularly journals and newspapers." Its interface allows filtering according to dates, language, location of distribution, among others. The ZDB points to six relevant newspapers published in Kaunas as shown in Table 1. In addition, it mentions 14 newspapers published in Vilnius and 29 in the Klaipėda region. The ZDB aggregates information about the availability of paper copies and microfilm holdings of newspaper issues in German and Austrian libraries. However, it does not take into account resources in other countries, such as the extensive collection of German-language newspapers held by the *Martynas Mažvydas National Library of Lithuania* (LNB)[4] in Vilnius, which, for example, includes 1,167 issues of the newspaper *Litauische Rundschau* in print (as well as approximately 440 duplicate issues). A smaller collection of this newspaper is held by the Vrublevski Library (VL) of the Lithuanian Academy of Sciences in Vilnius, which also provides facsimiles for eight issues, see Table 2. We also consulted the bibliographic resources of the Library of Congress in Washington, D.C., the world's largest library, in particular Carlton et al. (1965).

While the German newspaper portal (DZP) does not host any of the newspapers published in interwar Lithuania, it offers interesting inter-textual insights into their reception in the German Reich, exemplified in Figure 2.

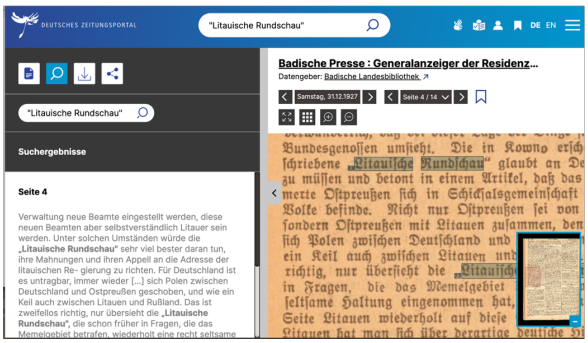


Figure 2: Mention of *Litauische Rundschau* in Badischer Presse in 1927 (Source: DZP, Visualisation: DFG Viewer)

## 4. Digitisation

The intended target representation of the digitised material comprises facsimile and OCRed text with positional coordinates, as well as article and zone segmentation, enabling users to switch between

[4] LNB: https://www.lnb.lt/

the text and its position in the facsimile. We intend to use the DFG Viewer tool[5], which requires METS/MODS[6] and structured METS/TEI[7] representations. In addition to plain text access via the DFG Viewer, we aim to compile the newspaper articles into a linguistically annotated corpus, at least enriched with lemmas and parts of speech, and make it searchable for linguists, for example using SpaCy[8] for annotation and ANNIS (Krause and Zeldes, 2016) for search, or the more recent Discourse Analysis Tool Suite (DATS) (Fischer and Biemann, 2025).

### 4.1. What has been digitised of interwar Lithuanian newspapers?

As mentioned in section 3 and shown in Table 2, the *Vrublevki Library* of the Lithuanian Academy of Sciences offers facsimiles of eight issues of *Litauische Rundschau*, each issue comprising about two to eight pages. A much larger collection of higher quality images is available at *Die Presse der deutschen Minderheit in Litauen 1918-1940* (Nareckaitė and Plaušinaitytė, 2020) that provides facsimiles for two newspapers and two further periodical resources, accompanied by neat introductions that contextualize the respective resources. In particular, these are: *Litauische Rundschau* with 24 issues from 1920 and 127 issues from 1912,[9] and *Deutsche Nachrichten für Litauen* with 151 issues unevenly distributed over six years. The other two resources are *Deutsche Genossenschaftnachrichten* (eight issues) and *Deutscher Kalender für Litauen* (ten annual editions with more than 100 pages on average). Several facsimiles of interwar German press publications have also been published on the Lithuanian digital heritage portal *epaveldas.lt*. Here you can find some issues of *Deutscher Kalender für Litauen* (1922, 1924, 1925, 1932, 1933).

### 4.2. OCR support

In the process of preparing our own digitisation project, we could rely on an active community concerned with questions of OCR, including information by the special interest group for newspapers and journals of the Association of Digital Humanities in the German-speaking area (DHd)[10], a monthly online OCR consulting hour of experts

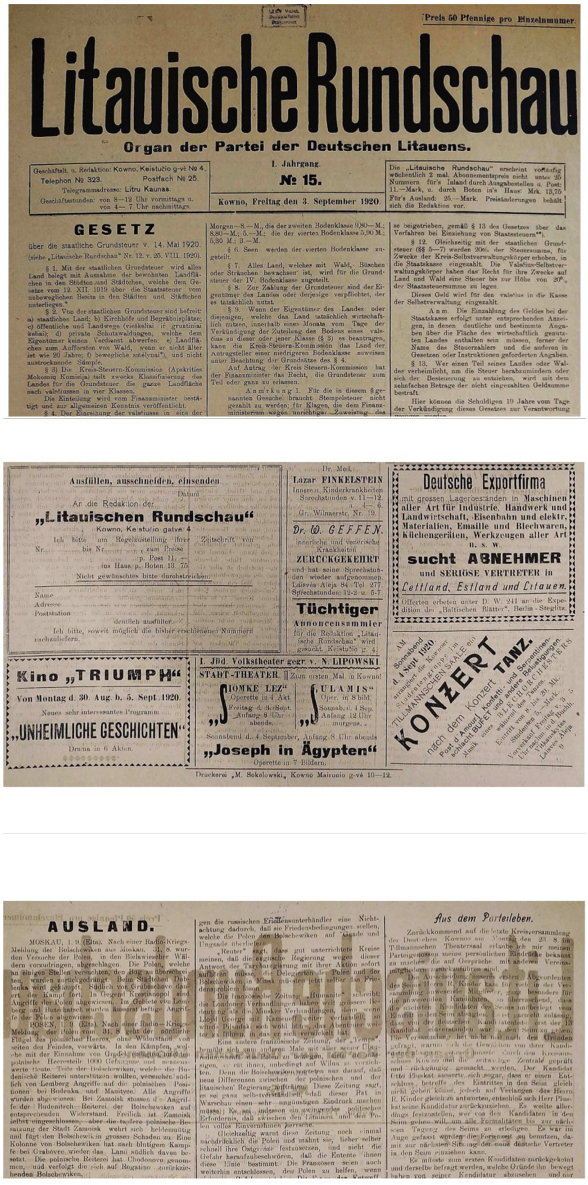


Figure 3: Facsimiles of *Litauische Rundschau* from year 1920 exemplifying some of the challenges for OCR: different font sizes, complex page layout, advertisements, bad paper quality with shining-through letters.

from German National Data Infrastructure (NDFI) consortia, and an online OCR recommender[11] that provides detailed recommendations for different OCR software and transcription tools. One of the reviewers also pointed out the extensive resources and community work of the project OCR-D.[12]

Even with professional support, the task of digitising historical newspaper is substantial. Figure 3

[5] https://dfg-viewer.de/en/the-project

[6] https://www.loc.gov/standards/mods/

[7] https://www.loc.gov/standards/mets/METSOverview.v2.html

[8] SpaCy:https://spacy.io/

[9] Litauische Rundschau: https://www.dpl.flf.vu.lt/litauische-rundschau/

[10] https://dhd-ag-zz.github.io/index.html.

[11] OCR recommender by BERD@NFDI: https://wiki.bib.uni-mannheim.de/limesurvey/index.php/996387

[12] https://ocr-d.de/

shows three example page sections of *Litauische Rundschau* from 1920 that exhibit challenging characteristics for OCR. Whiles these pages are printed in Antiqua script, issues from this newspaper published between 1924 and 1929 present the addition challenge of being printed in Fraktur script (not shown here).

## 5. Conclusion and future work

The paper provides a survey of historical German-language newspapers published in interwar Lithuania. By detailing the political history of the area and distinguishing three different historical regions in the area of the modern Lithuanian state, we try to motivate the selection of newspapers with which we want to work. Our bibliographic search identified relevant newspaper collections in Lithuania and Germany and also collections of facsimiles.

As far as we are aware, there is no machine-readable corpus of interwar Lithuanian German-language newspapers yet. The value of physically inspecting the available paper copies should not be underestimated either. When browsing through the collections, one comes across unexpected discoveries. For instance, whilst examining the collection of paper copies of the *Litauische Rundschau* from 1926 at the National Library of Lithuania (LNB), several issues of another German newspaper from Kaunas, Der *Wächter/Sargas*, were discovered. The title is neither listed in the library catalogue, nor mentioned in Weber's bibliography.

We performed pilot studies with different OCR tools in the context of a Digital Humanities and Linguistics seminar at the University of Hamburg, which resulted in very low-quality OCR texts, mainly because the semantic structures of the pages were not correctly identified. We are aware that this reflects not only the challenges of applying OCR to this type of image, but also our own limitations as novices in the field. In the next step, we will address the challenging task of applying OCR to the facsimiles in a more systematic manner.

## Acknowledgements

We would like to say thanks to three anonymous reviewers for their helpful comments and suggestions. We would also like to thank the participants of the workshop "Korpora und Editionen. Ansätze der Digital Humanities und didaktische Perspektiven" at Vilnius University for interesting discussions. Part of this work was funded by the German Academic Exchange Service (DAAD) with funds from the German Federal Foreign Office, project ID 57759025.

## Bibliographical References

Raphaël Barman, Maud Ehrmann, Simon Clematide, Sofia Ares Oliveira, and Frédéric Kaplan. 2021. Combining visual and textual features for semantic segmentation of historical newspapers. *Journal of Data Mining & Digital Humanities*, (HistoInformatics).

Jens Bjerring-Hansen, Ross Deans Kristensen-McLachlan, Philip Diderichsen, and Dorte Haltrup Hansen. 2022. Mending fractured texts. A heuristic procedure for correcting OCR data. In *Proceedings of the 6th Digital Humanities in the Nordic and Baltic Countries Conference, DHNB 2022*, volume 3232 of *CEURS Workshop proceedings*, pages 177–186.

Robert G. Carlton et al. 1965. Newspapers of East Central and Southeastern Europe in the Library of Congress. Slavic and Central European Division, Reference Department, Library of Congress, Washington.

Maud Ehrmann, Matteo Romanello, Simon Clematide, Phillip Benjamin Ströbel, and Raphaël Barman. 2020. Language Resources for Historical newspapers: the Impresso Collection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 958–968, Marseille, France. European Language Resources Association.

Tim Fischer and Chris Biemann. 2025. Semi-automatic Sequential Sentence Classification in the Discourse Analysis Tool Suite. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations)*, pages 151–162, Albuquerque, New Mexico. Association for Computational Linguistics.

Michel Généreux, Egon W Stemle, Verena Lyding, and Lionel Nicolas. 2014. Correcting OCR errors for German in Fraktur font. In *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & and of the Fourth International Workshop EVALITA 2014: 9-11 December 2014, Pisa*, pages 186–190. Pisa University Press.

Thomas Krause and Amir Zeldes. 2016. ANNIS3: A New Architecture for Generic Corpus Query and Visualization. *Literary and Linguistic Computing*, 31(1):118–139.

T. F. Mills. 1981. Preserving yesterday's news for today's historian: A brief history of news-

paper preservation, bibliography, and indexing. *The Journal of Library History (1974-1987)*, 16(3):463–487.

Nilo Pedrazzini and Barbara McGillivray. 2022. Machines in the media: semantic change in the lexicon of mechanization in 19th-century British newspapers. In *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*, pages 85–95, Taipei, Taiwan. Association for Computational Linguistics.

Lina Plaušinaitytė. 2021. Der Gebrauch der litauischen Sprache in der Presse der deutschen Minderheit im Litauen der Zwischenkriegszeit. In *Schnittstelle Germanistik: Forum für Deutsche Sprache, Literatur und Kultur des mittleren und östlichen Europas.*, volume 1, pages 31–55. Universitätsverlag WINTER GmbH Heidelberg.

Mia Ridge, Giovanni Colavizza, Laurel Brake, Maud Ehrmann, Jean-Phillipe Moreux, and Andrew Prescott. 2019. The past, present and future of digital scholarship with newspaper collections. *Multi-paper panel in DH 2019 book of abstracts*.

Uwe Springmann and Anke Lüdeling. 2017. OCR of historical printings with an application to building diachronic corpora: A case study using the RIDGES herbal corpus. *Digital Humanities Quarterly*, 11(2).

Albert Weber. 2013a. Teil 1: Zeitungen und Zeitschriften. In *Bibliographie deutschsprachiger Periodika aus dem östlichen Europa*. Institut für Ost- und Südosteuropaforschung.

Albert Weber. 2013b. Teil 2: Volkskalender, Almanache und Jahrbücher. In *Bibliographie deutschsprachiger Periodika aus dem östlichen Europa*. Institut für Ost- und Südosteuropaforschung.

Albert Weber. 2013c. Teil 3: Fachbibliographie deutschsprachiger Periodika. In *Bibliographie deutschsprachiger Periodika aus dem östlichen Europa*. Institut für Ost- und Südosteuropaforschung.

## Language Resource References

ANNO. *AustriaN Newspaper Online*. Austrian National Library (ÖNB).

DiFMOE. *Periodika der Digitalen Bibliothek*. Digitales Forum Mittel- und Osteuropa.

DTA. *Erweiterungskorpus des Deutschen Textarchivs, Genre Zeitung*. Digitalen Wörterbuchs der deutschen Sprache.

DZP. *Deutsches Zeitungsportal, German Digital Newspaper Portal*. German Digital Bibliothek (DDB).

Nareckaitė, Vidmantė and Plaušinaitytė, Lina. 2020. *Die Presse der deutschen Minderheit in Litauen 1918-1940*. Vilnius University.

VL. *Vrublevski Library of the Lithuanian Academy of Sciences*.

## Appendix

| Year | ID | Format | Issues | Location | Comment |
|---|---|---|---|---|---|
| 1920 | [1] | image (pdf) | 1–24 | VU Vilnius | based on [3] |
| | [2] | | 15, 18, 20–23 | VL Vilnius | based on [4] |
| | [3] | paper | 1–24 | LNB Vilnius | 07/16-10/08 |
| | [4] | | * | VL Vilnius | |
| | [5] | | 2–84 | DNB Leipzig | 07/20–12/30 |
| | [6] | microfilm | 2–84 | NOB Lüneburg, IFA Stuttgart | based on [5] |
| 1921 | [7] | image (pdf) | 1–127 | VU Vilnius | based on [9] |
| | [8] | | 80, 87 | VL Vilnius | based on [10] |
| | [9] | paper | 1–127 | LNB Vilnius | 01/01–06/29 |
| | [10] | | * | VL Vilnius | 01/01–07/17 |
| | [11] | | 2–146 | DNB Leipzig | 01/04–07/22 |
| | [12] | microfilm | 1–146 | NOB Lüneburg, IFA Stuttgart | based on [11] |
| 1922–1923 | | | not published | | |
| 1924 | [13] | paper | 1–157 | LNB Vilnius | 06/08–12/31 |
| | [14] | | 1–156 | LNB Vilnius | |
| | [15] | | * | VL Vilnius | |
| | [16] | | 2–157 | DNB Leipzig | 06/11–12/31 |
| | [17] | microfilm | 2–157 | NOB Lüneburg, IFA Stuttgart | based on [16] |
| 1925 | [18] | paper | 1–293 | LNB Vilnius | 01/01–12/31 |
| | [19] | | * | VL Vilnius | |
| | [20] | | 1–293 | DNB Leipzig | |
| | [21] | microfilm | 1–293 | NOB Lüneburg, IFA Stuttgart | based on [20] |
| 1926 | [22] | paper | 1–132,134-.279 | LNB Vilnius | |
| | [23] | paper | 27–28 | LNB Vilnius | |
| | [24] | | 1–286 | DNB Leipzig | 01/01–12/19 |
| | [25] | microfilm | 1–286 | NOB Lüneburg, IFA Stuttgart | based on [24] |
| 1927 | [26] | paper | 1–295 | DNB Leipzig | 01/01–12/31 |
| | [27] | microfilm | 1–295 | NOB Lüneburg, IFA Stuttgart | based on [26] |
| 1928 | [28] | paper | 1–215,217–289 | LNB Vilnius | |
| | [29] | paper | 1–222,224–249, 251–256,258–289 | LNB Vilnius | |
| | [30] | | 1–289 | DNB Leipzig | 01/01–12/30 |
| | [31] | microfilm | 1–289 | NOB Lüneburg, IFA Stuttgart | |
| 1929 | [32] | paper | 1–48 | | 01/01–02/27 |
| | [33] | | 1–145 | DNB Leipzig | 01/01–06/29 |
| | [34] | microfilm | 1–145 | NOB Lüneburg, IFA Stuttgart | based on [33] |

Table 2: Status of **Litauische Rundschau** = Lietuvos apžvalga ('Lithuanian Review'), published from 1920–1921 and 1924–1929; periodicity: initially two or three times weekly, later daily, irregularly; '*': availability of paper copies not verified; **VU Vilnius**: Nareckaitė and Plaušinaitytė (2020); **VL Vilnius**: Vrublevski Library of the Lithuanian Academy of Sciences; **LNB Vilnius**: Martynas Mažvydas National Library of Lithuania; **DNB Leipzig**: Deutsche Nationalbibliothek; **NOB Lüneburg**: Nordost Institut; **IFA Stuttgart**: Institut für Auslandsbeziehungen; other libraries in Berlin and Marburg also hold some issues of 1924 and 1925.