

Jörn Stegmeier/Kevin Kuck/Anna Christina Kupffer/Anne Christine Günther/
Angela Hammer/Dario Kampkaspar/Marcus Müller/Thomas Stäcker

Die Digitalisierung des Darmstädter Tagblatts (1740–1986)

Drei Jahrhunderte deutscher Sprachentwicklung

Abstract The Darmstädter Tagblatt (1740–1986) is one of the longest-running regional newspapers in the German-speaking world. Since 2019, it has been digitised, linguistically processed, and made freely accessible as part of a DFG-funded project conducted by the University and State Library Darmstadt (ULB) and Technical University of Darmstadt. The digitisation process poses both technical and legal challenges, particularly regarding layout and script recognition as well as the clarification of rights for individual contributions. The processed data are provided as TEI-XML full texts and made available via platforms such as TUDigit, WDB+, and CQPweb. This enables new applications in diachronic linguistic research. As demonstrated in this paper, the corpus supports analyses of morphosyntactic patterns and lexical developments spanning three centuries. Additional examples highlight the newspaper's transformation from a publication focused on classified advertisements to a modern daily, as well as the geographical reach of Darmstadt in the 18th century.

Keywords digitisation, corpus creation, rights clearance, newspaper, TEI, full texts



Abb. 1: Titelseite des Darmstädter Tagblatts vom 15.1.1750

1. Projektorganisation

Das hier vorgestellte Projekt¹ wird von der DFG gefördert² und ist eine Kooperation zwischen der Universitäts- und Landesbibliothek Darmstadt (ULB, Prof. Dr. Thomas Stäcker) und dem Fachgebiet Digitale Linguistik am Institut für Sprach- und Literaturwissenschaft der TU Darmstadt (Prof. Dr. Marcus Müller). Das Projekt umfasst zwei Phasen: Die erste Projektphase dauerte von 2019 bis 2022 und die zweite Phase begann 2023. In Phase 1 wurden die Jahrgänge 1740 bis 1941 digitalisiert. Die verbleibenden Jahrgänge bis 1986 folgen in Phase 2.

Die ULB ist dabei für die Erstellung der hochauflösenden Digitalisate und des maschinenlesbaren Volltextes zuständig. Der Kooperationspartner versieht zum einen den Volltext auf Token-Ebene mit Annotationen zu Wortart, Lemma und Named Entities (automatisierte Erkennung von z. B. Eigennamen, Ländernamen, Namen von Organisationen). Dabei entwickelt er zum anderen ein Verfahren, Urhebernennungen im Volltext zu erkennen, um damit die Lizenzanfragen im Rahmen der Rechtklärung, die in der zweiten Phase eine erhebliche Rolle spielt, zu unterstützen. Für die Veröffentlichung sind beide Kooperationspartner zuständig, da die Ergebnisse sowohl als durchsuchbarer Volltext als auch als Korpus zugänglich gemacht werden.³

2. Ausgangslage und Ziele

Zeitungen eröffnen auf mehreren Ebenen Zugänge zu einer Gesellschaft. Es sind dabei nicht nur die überregionalen Nachrichten, die in den Mittelpunkt des (Forschungs-)Interesses rücken. Regionale Nachrichten und Mitteilungen geben direkten Einblick in bestimmte Aspekte des öffentlichen Lebens, die bei überregionalen Nachrichten keine Rolle spielen. Dies gilt in besonderem Maße für Kleinanzeigen und Artikel über regionales Kultur- und Vereinsleben und zumindest teilweise auch für Werbeanzeigen. Unabhängig von den Ereignissen bieten Zeitungen – und insbesondere Regionalzeitungen – aber auch Zugang zum Sprachgebrauch einer bestimmten Region. Zeitungen, die über einen langen Zeitraum in derselben Region erscheinen, eröffnen all diese Zugänge auch diachron und erlauben somit vergleichende Studien.

Allerdings sind die Originalausgaben historischer Zeitungen nicht immer einfach zugänglich und eher umständlich im Gebrauch. Die in der Regel schlechte Papierqualität insbesondere von Zeitungen des 19. Jahrhunderts schränkt die Benutzbarkeit ebenfalls ein, bis dahin, dass Exemplare nicht mehr ausgehändigt werden können, weil sie durch den Gebrauch zu sehr strapaziert werden. Die Digitalisierung des Zeitungsbestands ist daher eine sinnvolle und auch notwendige Maßnahme, um sowohl interessierten Wissenschaftler:innen als auch Privatpersonen uneingeschränkten und bequemen Zugang bieten zu können. Im Zeitalter von automatischer Layout- (OLR, *optical layout recognition*), Buchstabenerkennung (OCR, *optical character recognition*) und internetbasiertem Zugriff sind frühere Digitalisate wie Filme oder Mikroformate jedoch oft nicht mehr ausreichend. Zum einen erfordert ihre Publikation über das Internet durchaus aufwendige Transformationsschritte und zum anderen reicht die Gesamtqualität meist nicht aus, um daraus maschinenlesbaren Volltext zu erstellen.

1 Eine ausführlichere Projektbeschreibung findet sich in Stegmeier et al. (2022).

2 Projektnummer 422840794.

3 Die weiterführenden Links sind über die Projektwebseite erreichbar: www.ulb.tu-darmstadt.de/forschen_publizieren/forschen/darmstaedter_tagblatt.de.jsp (Stand: 27.11.2025).

Das Darmstädter Tagblatt erschien in sich wandelnden Titeln und mit einer Unterbrechung während der 1940er Jahre von 1740 bis 1986, als es von der Echo Medien Gruppe aufgekauft und zugunsten des Lokalkonkurrenten, dem Darmstädter Echo, eingestellt wurde. Während seiner Existenz wandelte sich das Tagblatt von einem ‚Frag- und Anzeigblatt‘, das zunächst nur einige Male pro Monat erschien, zu einer modernen Tageszeitung. Während die Anfangsphase primär Anzeigen unter Nennung der Produktpreise, Listen von in Darmstadt ankommenden Reisenden und offizielle Verkündigungen enthielt, wandelte sich der Inhalt ab dem letzten Viertel des 19. Jahrhunderts. Ab diesem Zeitpunkt überwogen Nachrichten und Berichte, während Anzeigen zwar weiterhin existierten, nun aber zunehmend in den Hintergrund rückten. Im Jahr 1941 wurde das Darmstädter Tagblatt durch eine Anordnung der Reichspressekammer eingestellt und erschien nach Ende des Zweiten Weltkriegs ab 1949 wieder. Neben dem Inhalt wandelte sich auch das Format mehrmals und nach dem Krieg fand ebenso ein Schriftwechsel von Fraktur auf Antiqua statt. Insgesamt umfasst das Tagblattprojekt 598 Bände bzw. 600.000 Seiten.

Der Gegenstand des Digitalisierungsprojekts ist daher keineswegs homogen und stellt den Digitalisierungsworkflow des Projekts vor zahlreiche Herausforderungen. Offensichtlich ist zunächst die schiere Menge an Daten, die während des Digitalisierungsworkflows gespeichert, verarbeitet und bereitgestellt werden muss. Allein dies macht das Tagblatt zu einem der aufwändigsten Projekte, das am Zentrum für Digitale Editionen (ZEiD) an der ULB gehostet wird. Weiterhin hat die Heterogenität des Materials Auswirkungen auf den kompletten Digitalisierungsworkflow, etwa beim Scannen oder bei der OCR- und Layouterkennung. So sind beispielsweise Umfang (Zahl der Ausgaben und damit auch der Seitenzahlen), Format und Anzahl bei den einzelnen Bänden, in denen das Tagblatt im Bestand der ULB überliefert wurde, höchst unterschiedlich. Gleiches gilt für die einzelnen Ausgaben. Während das Tagblatt beispielsweise im 18. Jahrhundert einmal wöchentlich mit einem Umfang von bis zu sechs Seiten erschien, wurde es ab dem Ende des 19. Jahrhunderts als Tageszeitung herausgegeben. Je nach Umfang der eigentlichen Zeitungsausgabe sowie etwaigen Beilagen schwankt hier die Seitenzahl zwischen 8 und 36 Seiten. Mit Hilfe der Onlinepräsentation lassen sich zudem die für den eigenen Forschungsschwerpunkt relevanten Details zum Umfang im Sinne von Ausgaben und Seitenzahlen schnell erörtern. Der Arbeitsaufwand bei der Digitalisierung ist also durchaus vergleichbar mit der Digitalisierung mehrerer Zeitungen.

Neben der technischen Umsetzung stellen auch rechtliche Aspekte das Projekt vor große Herausforderungen, da die Ausgaben der zweiten Projektphase (in Ausnahmen sogar noch Ausgaben aus der ersten Phase) dem Urheberrecht unterliegen. Das Projekt konnte in dieser Form überhaupt nur zustande kommen, weil die Eigentümerin des Darmstädter Tagblatts, die Echo Medien GmbH, als Projektpartnerin gewonnen werden konnte und dem Projekt eingeräumt hat, alle Inhalte, für die sie die Rechte besitzt, veröffentlichen zu dürfen. Dennoch ist damit die Rechtklärung noch nicht abgeschlossen. Grundsätzlich ist jeder Beitrag, beispielsweise Berichte, Bilder, Zeichnungen oder Fortsetzungsromane, um nur eine Auswahl zu nennen, als eigenständiges Werk mit eigenem/eigener Urheber:in anzusehen, für den jeweils die Rechtslage mit der/dem betroffenen Verlag, Agentur, Fotograf:in etc. geklärt werden muss.

3. Digitalisierung und Bereitstellung

3.1 Technische Umsetzung

In beiden Projektphasen wurden im Digitalisierungszentrum der ULB hochqualitative Scans angefertigt. Während in der ersten Phase die OCR- und Layouterkennung im ZEI^D mittels der Plattform Transkribus (www.transkribus.org/de, Stand: 27.11.2025) anhand eigens trainierter Modelle selbst durchgeführt wurde, wurde in der Projektplanung für die zweite Projektphase zwecks besserer Planbarkeit entschieden, einen Dienstleister mit der automatischen Erstellung der Volltexte zu beauftragen. In jedem Fall steht am Ende die Bereitstellung der durchsuchbaren Volltexte in TEI-XML bzw. der Digitalisate über verschiedene Plattformen, etwa TUDigit (ULB), WDB+ (ULB/ZEiD) und CQPweb (Digitale Linguistik/TU).⁴

3.2 Rechteklärung Phase 1 und Phase 2

Die Vorarbeiten in der ersten Projektphase (1740–1941) haben bestätigt, dass selbst bei einer bis zu den ermittelbaren Einzelurheber:innen reichenden Rechteeinräumung die Möglichkeit verbleibt, dass einzelne Beiträge oder Abbildungen nachlizenziiert oder auf Unterlassungsaufforderung wieder aus dem Netz genommen werden müssen bzw. lediglich in den Räumen der Bibliothek gezeigt werden dürfen. Diese Unschärfe ist nicht heilbar, da das Urheberrecht grundsätzlich bei der Person verbleibt, die das Werk geschaffen hat. Wenn man also nicht gänzlich auf die Digitalisierung und Publikation von Zeitungen jüngerer Datums verzichten will, muss trotz Rechteeinräumung von Verlagen ein gewisses, wenn auch geringes Restrisiko in Kauf genommen werden.

Da die Digitalisierung und weltweite Bereitstellung im Internet als Vervielfältigung und Wiederveröffentlichung in neuer Form zu verstehen ist, hat die ULB den rechtlichen Rahmen des Projekts geprüft. Das Tagblatt ist sowohl ein vergriffenes als auch ein verwaistes Werk im Sinne des Urheberrechtsgesetzes (UrhG).⁵ Für die Veröffentlichung von vergriffenen Werken ist eine Lizenzvereinbarung zu treffen, für die Publikationsform „Zeitung“ erklärten sich die typischen Lizenzierungspartner wie die VG Wort zu Projektbeginn explizit jedoch als nicht zuständig. Der „Lizenzierungsservice Vergriffene Werke“ der Deutschen Nationalbibliothek (VW-LiS) war zum Zeitpunkt der Prüfung bzw. seit 7.6.2021 für unbestimmte Zeit ausgesetzt.⁶ Da die Digitalisate des Tagblatts von Forschenden aber zeitnah benötigt wurden bzw. werden, hatte sich die ULB entschieden, das Tagblatt weiterhin in seiner Eigenschaft als „verwaistes Werk“ zu behandeln.

4 Die verschiedenen Plattformen werden über die Projektwebseite verlinkt. www.ulb.tu-darmstadt.de/forschen_publizieren/forschen/darmstaedter_tagblatt.de.jsp (Stand: 27.11.2025).

5 Das Tagblatt wurde 1986 eingestellt. Rechte an der Marke und am Archiv wurden von der Konkurrenz-Zeitung Darmstädter Echo übernommen. Trotz gezielter Prüfung ist es rechtlich nicht belegt, dass das Darmstädter Echo der vollumfängliche Rechtsnachfolger des Darmstädter Tagblatts ist. Das Darmstädter Tagblatt musste wegen Unterschlagungsdelikten der damaligen Geschäftsführung eingestellt werden (www.spiegel.de/politik/vom-schlag-getroffen-a-41637d60-0002-0001-0000-000013520424, Stand: 27.11.2025). In diesem Zusammenhang sind mündlicher Aussage des Echo-Managements zufolge viele Geschäftsdokumente vernichtet worden, und es lassen sich keine Arbeitsverträge damaliger fester oder freier Mitarbeiter:innen nachweisen.

6 Der „Lizenzierungsservice Vergriffene Werke“ steht mittlerweile wieder zur Verfügung: www.dnb.de/DE/Professionell/Services/VW-LiS/vwliis_node.html (Stand: 28.1.2026).

Für „verwaiste Werke“ sind „besondere gesetzlich erlaubte Nutzungen“ vorgesehen, wenn diese „Bestandsinhalte von Sammlungen öffentlich zugänglicher Bibliotheken sind, [...] bereits veröffentlicht (wurden) und deren Rechtsinhaber auch durch eine sorgfältige Suche nicht festgestellt oder ausfindig gemacht werden konnten“ (UrhG §61,2). Die Grundvoraussetzung für die Vervielfältigung und öffentliche Zugänglichmachung in UrhG §61,5 ist insofern erfüllt: Die ULB handelt in Erfüllung ihrer im Gemeinwohl liegenden Aufgaben, indem sie Bestandsinhalte bewahrt und auf Dauer zugänglich macht, um kulturellen und bildungspolitischen Zwecken zu dienen.

Die ULB hat in der ersten Projektphase umfangreiche Vorarbeiten geleistet und zunächst die Expertise eines Fachanwalts für Urheber- und Medienrecht eingeholt. Dieser sieht in der o.g. „sorgfältigen Suche“ den Dreh- und Angelpunkt der Bestimmungen zu den verwaisten Werken. Nur wer die Suche in nachvollziehbarer Weise durchführe, dokumentiere und archiviere, könne sich auf die Privilegierung der Vorschriften des §61 ff. UrhG berufen. Jeder Bestandsinhalt müsse gesucht werden und zwar vor der Veröffentlichung im Internet. Dort, wo kein Name genannt ist oder wo einem Kürzel kein Klarnamen zugeordnet werden kann, seien die Rechtsverhältnisse an einem Werk nicht nachvollziehbar.

Es scheint auf dieser Grundlage sinnvoll, die Inhalte so zu kategorisieren und zu priorisieren, wie sie im Fall einer Rechtsstreitigkeit ins Gewicht fallen würden. In die oberste Kategorie fällt das Agentur- und Verlagsmaterial (heute noch aktiver Firmen), das im Tagblatt in großer Menge abgedruckt war. Die zweite Kategorie bilden Fotograf:innen und die dritte Gruppe die Textautor:innen. Durch die Vorarbeiten konnten ca. 40% der in den Jahrgängen der Phase II enthaltenen Werke kostenfrei lizenziert werden, vor allem durch die erzielten Vereinbarungen mit den größten Agenturen.

Um auch für den Rest der Werke möglichst vollständig Lizenzabkommen abschließen zu können, wurden bereits in der ersten Phase Listen mit Kürzeln (bestehend aus 1–4 Buchstaben) und Namen der Urheber:innen angelegt. Da die Vertragshistorie des Darmstädter Tagblatts unvollständig ist, sind diese Kürzel und Namen nicht anhand verfügbarer Dokumente auflösbar. Wegen des starken Regionalbezugs der Zeitung konnte die gezielte Internetsuche (Google) in vielen Fällen erste Ergebnisse wie Traueranzeigen Verstorbener oder aktuelle Anschriften liefern. Zudem wurde in Stadtlexika, in der Deutschen Nationalbibliothek (DNB), in HeBIS (Hessisches BibliotheksInformationssystem) und im jüngsten Darmstädter Adressbuch (2002) gesucht. Wenn mehrere Namensträger in Frage kamen, wurden alle Adressen mit in die Versandliste aufgenommen. Auf Anfrage teilte das Einwohnermeldeamt mit, dass für eine Adressermittlung mindestens drei Datenpunkte wie vollständiger Name, letzte Adresse und Geburtsdatum benötigt werden. Da nur die (Nach-)Namen bekannt sind, ist dieser Weg nicht gangbar. Ähnlich aussichtslos ist ein Ersuchen an das Nachlassgericht zur Ermittlung von Erben. Die Suche wurde detailliert dokumentiert (inklusive Suchdatum, Lebensdaten, Adresse, Notizen zur Art der Tagblatt-Mitarbeit). An alle ermittelten Adressen wurde ein Informationsbrief versendet, der den Urheber:innen das Digitalisierungsprojekt vorstellt und sie um eine kostenlose, einfache, weltweite und zeitlich unbeschränkte Lizenz für ihre im Tagblatt abgedruckten Werke bittet.

In der derzeit laufenden zweiten Phase werden mit fortschreitender Digitalisierung des Tagblatts technische Ansätze für eine (teil-)automatisierte Urhebersuche erprobt und durchgeführt. Diese zielen darauf ab, Kürzel und Namen von Urheber:innen automatisch zu finden, grob zu klassifizieren, zu extrahieren und in einer Datenbank zu speichern. Das Vorgehen ist dabei zweigleisig.

Zum einen erarbeiten die Mitarbeiter:innen des Projekts mögliche Kategorien und reguläre Ausdrücke auf der Grundlage der Lektüre ausgewählter Ausgaben. Die Kategorien dienen

dabei unter anderem der Priorisierung und orientieren sich an den oben genannten Beispielen. Die folgenden Ausführungen dienen der Illustration, sind daher nicht vollständig. Als mögliche Urheber:innen bzw. Lizenzgeber:innen kommen also hauptsächlich Agenturen, Autor:innen, Fotograf:innen und Verlage vor (vgl. Abb. 2). Für die automatische Urhebererkennung ist darüber hinaus noch wichtig zu vermerken, wo im Artikel die Nennung stattfindet und vor allem, wie diese Stellen in der XML-Struktur des Volltexts möglichst eindeutig erreichbar sind. Mögliche Stellen sind z.B. in einem Untertitel oder als vollständiger Untertitel. Dies trifft vor allem auf Nennungen eines vollständigen Namens zu, die oft, aber nicht immer mit „Von“ eingeleitet werden. Sind sie Teil eines Untertitels, stehen sie nach dem Untertitel und sind oft mit einem Schrägstrich vom Untertitel abgetrennt. Autorenkürzel können am Anfang oder am Ende eines Artikels stehen. In der Regel sind sie vor bzw. nach dem Text zu finden und können mit einem Punkt abgeschlossen werden oder Bindestriche enthalten. Agenturkürzel stehen in der Regel nach einem Ortsnamen in der ersten Zeile in runden Klammern. Es können mehrere Agenturen in einer Klammer genannt werden, die dann normalerweise mit Schrägstrichen voneinander abgetrennt sind.⁷

Abkürzung einer Presseagentur

Autorenkürzel

XML-Datei

Eintrag in Datenbank

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<tei:div type="jArticle">
  <tei:head>Die Waffe der Furcht</tei:head>
  <tei:p>Kps. „Wir müssen der kommunistischen Waffe der Furcht entgegenzutreten“, betonte Präsident Truman in seiner letzten Ansprache auf der ...
```

urheber	heber_or	urheberart	stelle	durchsuchter_text	kontext	muster
Kps.	Kps.	Zb_autor_abk	Artikelanfang	Kps. „Wir müssen der kommunistischen Waffe der Furcht entgegenzutreten“, betonte Präsident Truman in seiner letzten Ansprache auf der ...	Kps. „Wir müssen der kommunistischen Waffe der Furcht entgegenzutreten“, betonte Präsident Truman in seiner letzten Ansprache auf der ...	^(A-Za-z)+\.

Abb. 2: Verschiedene Formen der Urhebernennung und Extraktionsergebnis

Zum anderen arbeiten zwei Mitarbeiter:innen des Projektes daran, einen Goldstandard der Urhebernennungen auf der Grundlage der bisher gescannten Jahrgänge (1950–1970) zu erarbeiten.⁸ In 18 Ausgaben (ca. 4.500 Artikel⁹) annotieren sie manuell alle Urhebernennungen mit den oben beschriebenen Kategorien und vermerken, ob auf eine Agentur, eine:n Autor:in etc. mit Kürzel oder vollständigem Namen am Anfang oder am Ende eines Artikels verwiesen wird. Während der Annotationsarbeiten können neue Kategorien eingefügt werden, die dann von beiden Annotierenden verwendet werden. Nach Abschluss der Annotierungen werden Zweifelsfälle vereinheitlicht und die Daten dienen als Referenz für die automatisiert erkannten und extrahierten Stellen. Sobald die Extraktionen per Skript eine akzeptable Trefferquote (angestrebt sind 96%) erreicht haben, werden sie produktiv zur Extraktion eingesetzt und dabei kontinuierlich geprüft und verbessert.

⁷ Siehe Abbildung 2.

⁸ Hierfür wird die Software INCEpTION verwendet, mit der beliebige Annotationskategorien erstellt und zugewiesen werden können (vgl. Klie et al. 2018 und <https://inception-project.github.io/>).

⁹ In den zu annotierenden Texten wurden 4.224 Texteinheiten von der automatischen Layouterkennung als Artikel und 4.789 als Überschrift ausgezeichnet. Aufgrund des Seitenlayouts werden manche Texteinheiten jedoch fälschlich als Teil eines Artikels erkannt, obwohl es sich inhaltlich um mehrere Artikel handelt. Vorläufige Ergebnisse zeigen ca. 100 Urhebernennungen in einer Ausgabe von 1950 bei 175 als Artikel ausgezeichneten Texteinheiten (184 Überschriften).

Die Einträge in der Datenbank werden dann mit den bereits bekannten Kürzeln und Namen abgeglichen. Noch nicht bekannte Kürzel werden nach Möglichkeit aufgelöst, indem die Impresen und noch vorliegenden Dokumentationen eingesehen werden. Für die aufgelösten Kürzel und eventuell neue vollständige Namen führen die Projektmitarbeiter:innen den oben beschriebenen Prozess durch.

3.3 Linguistische Aufbereitung

Nach der Buchstabenerkennung wurden die resultierenden Texte in einzelne Wortformen zerlegt (Tokenisierung) und für jede Wortform wurden die Grundform (Lemmatisierung) und die Wortart (Part-of-Speech-Tagging) annotiert. Darüber hinaus wurden sogenannte Named Entities, also z. B. Eigennamen und Namen von Institutionen und morphologische Informationen, auf Tokenebene ermittelt. Der Text wurde in Sätze zerlegt und jeder Satz wurde auf seine Struktur hin analysiert (Dependency Parsing).

Die folgenden Werkzeuge kamen für die oben genannten Aufgaben zum Einsatz, wobei die Tagging-Ergebnisse ohne weitere Korrekturen parallel genutzt werden:

- Tokenisierung und Satzsegmentierung: SoMaJo (Proisl/Uhrig 2016)
- Lemmatisierung: TreeTagger (Schmid 1995)
- Part-of-Speech-Tagging: spaCy (de_dep_news_trf; Honnibal et al. 2021) und TreeTagger
- NER: spaCy (de_core_news_lg), Flair (dbmdz/flair-clef-hipe-german-base; Schweter/März 2020)
- Morphologische Informationen und Dependency Parsing: Stanford Stanza (Qi et al. 2020)

3.4 Bereitstellung und Nutzungsmöglichkeiten

Die Daten werden auf unterschiedlichen Wegen und mit unterschiedlichen Nutzungsschwerpunkten zugänglich gemacht.

Die linguistisch aufbereiteten Daten wurden in die Corpus Workbench (CWB) (vgl. Evert/Hardie 2011) importiert und sind über die Plattform DiscourseLab zugänglich.¹⁰ Der Zugang erfolgt über CQPweb (vgl. Hardie 2012), das sowohl eine browserbasierte Nutzeroberfläche als auch Analysefunktionen für Korpora, die in der CWB verwaltet werden, bereitstellt. Eine Veröffentlichung der Faksimiles erfolgt ebenso über die Plattform TUDigit der ULB Darmstadt, die digitale Sammlungen wie die des Darmstädter Tagblatts bereitstellt. Die ebenfalls an der ULB Darmstadt entwickelte Plattform WDB+ ermöglicht neben der Anzeige von Digitalisaten auch textbasierte Recherchen, da zusätzlich digitale Volltexte im TEI-XML-Format bereitgestellt werden.¹¹ Durch die Bereitstellung dieser Volltexte sind weiterführende linguistische Analysen durch Forschende möglich, etwa zur diachronen Sprachentwicklung, zur Untersuchung lexikalischer oder morphosyntaktischer Muster sowie zur Erforschung sprachlicher Variation.

10 Der Zugang findet sich auf www.discourselab.de/cqpweb/ (Stand: 27.11.2025). Für die Nutzung ist eine Registrierung notwendig.

11 Die jeweils aktuellen Links zu den Plattformen sind auf der Projektwebseite aufgelistet: www.ulb-tu-darmstadt.de/forschen/publizieren/forschen/darmstaedter_tagblatt.de.jsp (Stand: 27.11.2025).

4. Beispiele für Anwendungsmöglichkeiten

4.1 Nominalphrasen zeigen den Wandel vom Anzeigenblatt zur Zeitung

Wie oben in der Beschreibung des Darmstädter Tagblatts dargestellt, entwickelte sich dieses von einem Anzeigenblatt zu einer regionalen Tageszeitung mit wachsendem redaktionellen Teil und verschiedenen Ressorts. Wie in Regionalzeitungen üblich, erschienen weiterhin Klein- und Werbeanzeigen. Es ist zu erwarten, dass die zusätzlichen Themen und Inhalte sich auf das Vokabular auswirken und diese Auswirkungen durch Abfragen des CQPweb-Korpus zu belegen sind.

Listen der häufigsten Substantive sind hierfür nur bedingt geeignet. Hilfreicher präsentieren sich die Ergebnisse, wenn statt einzelner Substantive ganze Nominalphrasen (NP) untersucht werden. In die Abfrage werden die minimal erweiterte NP aus Artikel gefolgt von einer beliebigen Anzahl an attributiv gebrauchten Adjektiven gefolgt von einem Nomen aufgenommen. Die folgende Suchanfrage findet beides: „[pos=“ART“] [pos=“ADJA“]+ [pos=“NN“]“. Der Ausdruck „[pos=“ADJA“]“ findet alle als attributives Adjektiv getaggten Wörter und das Pluszeichen dahinter steht für „einmal bis beliebig oft“.

Nr.	1750	1800	1850	1880	1900
1	der armen Seele	dem hiesigen Rathhaus	ein kleines Logis	eine ruhige Familie	ein möbl. Zimmer
2	Ein Maaßjunges ¹² Bier	eine ledige Person	Ein vollständiges Logis	der mittlere Stock	eine schöne 3=Zimmerwohnung
3	das gantze Jahr	der langen Gasse	den evangelischen Gemeinden	ein möbl. Zimmer	eine kleine Wohnung
4	ein ungenannter Freund	der neuen Vorstadt	Ein freundliches Logis	ein schönes Logis	der Technischen Hochschule
5	der Herrnhutischen Secte	der großen Ochsen-gasse	der mittlere Stock	eine schöne Wohnung	eine schöne 4=Zimmerwohnung

Tab. 1: Die häufigsten Nominalphrasen in Fünf-Jahres-Abschnitten ab dem im Spaltenkopf vermerkten Jahr. 1750–1900

Wie die Einträge in Tabelle 1 zeigen, sind in den ausgewählten Zeiträumen keine offensichtlichen Hinweise auf nachrichtliche oder redaktionelle Inhalte enthalten. Die meisten der aufgeführten Nominalphrasen zeigen stattdessen deutlich den Anzeigencharakter. Stichproben belegen dies. So ist z. B. *der armen Seele* Teil eines Buchtitels, der in einer Verkaufs-

12 „Maaßjunges“ ist ein OCR-Fehler, der auf sehr eng gedruckte Tabellen zurückgeht. Richtig heißt es „Ein Maaß junges Bier“.

anzeige genannt wird: „Betrachtung über die Bräutigams=Liebe des HErrn Iesu zu der armen Seele“ (vgl. z.B. Darmstädter Tagblatt vom 4.1.1753, S. 2, <https://exist.ulb.tu-darmstadt.de/3/v/pa000059-0001>¹³). Dasselbe gilt für den Eintrag der *Herrnhutischen Secte*, dieser ist Teil des Titels „Die Schule der Verständigen, oder das auf Sinnlichkeit gegründete Atheistische Lehr=Gebäude, der heutiges Tages in der Herrnhutischen Secte wieder auflebenden sogenannten Homium Intelligentiæ, in einem Gespräch zwischen Cautorius und Berrhoisius ausfindig gemacht und ans Licht gestellt, von Johann Christoph Venator, zweiten Evangelis. Predigern in Friedberg.“ (vgl. z.B. Darmstädter Tagblatt vom 30.4.1750, S. 3, <https://exist.ulb.tu-darmstadt.de/3/v/pa000059-0001>, Stand: 27.11.2025). Keiner Verkaufsanzeige, sondern einer Art Werbeanzeige gehört *ein ungenannter Freund* an: „AVERTISSEMENT Bei dem HErrn finden unsere Waysen Hülfe , und Errettung aus aller ihrer Noth. Daher schickte ein ungenannter Freund den 20.ten Martii 1.fl. 15 Alb. mit Verlangen vor ihn in seiner Krankheit zu beten“ (vgl. z.B. Darmstädter Tagblatt vom 26.3.1750, Titelseite, <https://exist.ulb.tu-darmstadt.de/3/v/pa000056-0013>, Stand: 27.11.2025).

Wenn auch die NP *dem hiesigen Rathhaus* ein Hinweis auf Berichterstattung über regionale Politik sein könnte, so zeigen die Konkordanzen doch recht deutlich, dass es sich auch dabei um Teile von Anzeigen handelt, wie in Tabelle 2 gezeigt wird:

Context before	Query item	Context after	year
Mittwoch den 29ten die= ses Nachmittags um 2 Uhr auf	dem hiesigen Rathhaus	unter denen in dem Termin bekannt zu machenden Bedingungen versteigt	1803 ¹⁴
den 13ten künftigen Monats Ottober Nachmittags Um 2 Uhr auf	dem hiesigen Rathhaus	nochmalen versteigt und dem Meistbietenden unwiederruflich zugeschlagen werden. Darmstadt	1802 ¹⁵
künftigen Monats August , Nachmittags 2 Uhr , fen auf	dem hiesigen Rathhaus	der dem Feldwebel Fenner im 2ten Bat . Leibregiments gehörige	1801 ¹⁶
Monats Februarii , Nachmittags um 2 Uhr , sollen auf	dem hiesigen Rathhaus	nachbeschriebene zur Verlassenschaft des Burger und Metzgermeister Ludwig Haxt gehörige	1802 ¹⁷
Erben und Christian Engel, vorbehaltlich der Ratification , auf	dem hiesigen Rathhaus	öffentlich versteigt werden. Darmstadt den 28ten März 1800 .	1800 ¹⁸

Tab. 2: Konkordanzen zu „dem hiesigen Rathhaus“ 1800–1805 (zufällige Auswahl)

13 Die digitale Edition des Darmstädter Tagblatts befindet sich noch im Aufbau. Die hier genannten Links auf <https://exist.ulb.tu-darmstadt.de> (Stand: 27.11.2025) verweisen auf eine Vorschau und sind nicht permanent, also nicht zitierfähig.

14 Vgl. <https://exist.ulb.tu-darmstadt.de/3/v/pa000109-0024>.

15 Vgl. <https://exist.ulb.tu-darmstadt.de/3/v/pa000108-0040>.

16 Vgl. <https://exist.ulb.tu-darmstadt.de/3/v/pa000107-0031>.

17 Vgl. <https://exist.ulb.tu-darmstadt.de/3/v/pa000108-0004>.

18 Vgl. <https://exist.ulb.tu-darmstadt.de/3/v/pa000106-0013>.

Die Einträge in Tabelle 3 ab 1910 zeigen jedoch deutlich, dass mehr nachrichtliche Texte abgedruckt werden.

Nr.	1910	1920	1925	1930	1935
1	die höchsten Preise	Die deutsche Regierung	der Deutschen Volkspartei	des deutschen Volkes	Die 22 mm
2	der andern Behörden	der Deutschen Volkspartei	den letzten Tagen	Das deutsche Volk	des deutschen Volkes
3	die trauernden Hinterbliebenen	das deutsche Volk	der letzten Zeit	den letzten Jahren	Das deutsche Volk
4	den politischen Teil	der deutschen Regierung	den letzten Jahren	den letzten Tagen	den letzten Jahren
5	den redaktionellen Teil	des deutschen Volkes	der anderen Seite	der letzten Zeit	den letzten Tagen

Tab. 3: Die häufigsten Nominalphrasen in Fünf-Jahres-Abschnitten ab dem im Spaltenkopf vermerkten Jahr. 1910–1935

Die Phrasen „den redaktionellen Teil“ bzw. „den politischen Teil“ in den fünf Jahrgängen ab 1910 referieren direkt auf diesen Wandel, während in den weiteren Fünf-Jahres-Ausschnitten Phrasen wie „das deutsche Volk“, „des deutschen Volkes“, „die deutsche Regierung“ etc. ihn auf inhaltlicher Ebene zeigen.

4.2 Kein Datenschutz im 18. Jahrhundert:
Reisende nach Darmstadt

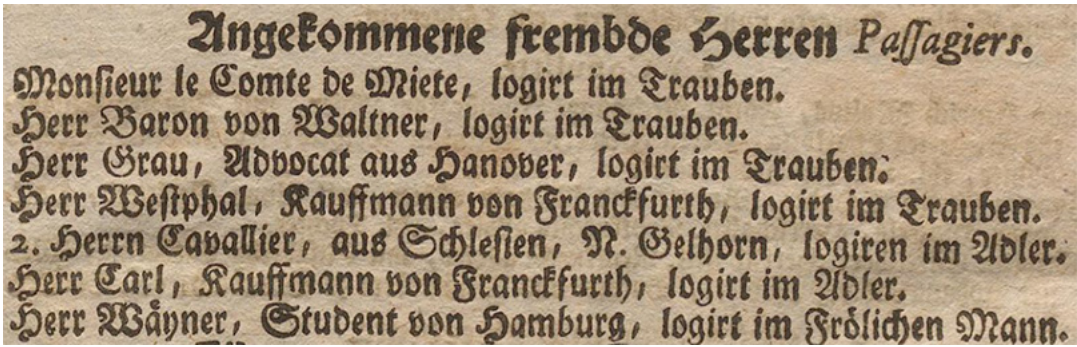


Abb. 3: Liste angekommener Besucher. Darmstädter Tagblatt 15.1.1750, S. 3 unten

Ein Blick in vergangene Ausgaben bringt nicht nur erwartbare Ergebnisse wie z.B. Verkaufsangebote und amtliche Bekanntmachungen. Das Darmstädter Tagblatt veröffentlichte in den frühen Ausgaben bis in das 19. Jahrhundert die Namen, Berufe und Unterkünfte von Reisenden, die nach Darmstadt kamen (vgl. Abb. 3). Eine Auswertung dieser Daten ermöglicht es, einen Eindruck vom Einzugsgebiet Darmstadts im 18. Jahrhundert zu bekommen, wie in Stegmeier/Müller (2024) näher ausgeführt wird. Nach einer semi-automatischen Aufbereitung der Passagierdaten wurden diese als CSV-Dateien gespeichert und über den Geo-Browser (vgl. Kollatz 2016) von DARIAH-DE (<https://geobrowser.de.dariah.eu/>, Stand: 27.11.2025) verfügbar gemacht. Hierfür wurden die Ortsnamen nor-

malisiert. Im Geo-Browser selbst wurden den Einträgen dann anhand der Ortsnamen Längen- und Breitengrade zugeordnet. In Zweifelsfällen wurde der Darmstadt am nächsten liegende Ort gewählt. Alle so erstellten Datenblätter können über einen persistenten Link von **DARIAH-DE** heruntergeladen werden.¹⁹ Die folgende Abbildung 4 vergleicht die Ergebnisse für drei Zehn-Jahres-Schritte:

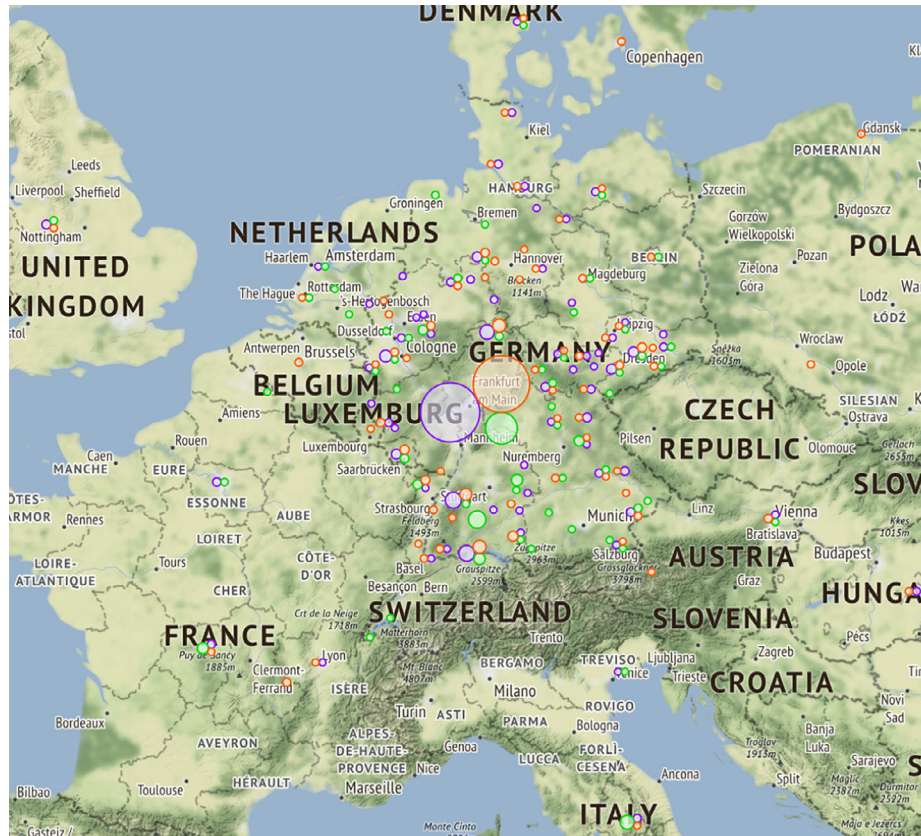


Abb. 4: Zehn-Jahres-Schritte im Vergleich: 1745–1754 (orangefarben), 1755–1764 (violett), 1785–1794 (grün)

Wenig überraschend kommen die meisten Besucher aus der Umgebung. Das Einzugsgebiet insgesamt ist jedoch angesichts der eher beschwerlichen Reisen in der zweiten Hälfte des 18. Jahrhunderts beeindruckend: Vom heutigen Frankreich, England, Italien, Österreich, Ungarn, Polen und Dänemark kamen Besucher nach Darmstadt. Den Berufsbezeichnungen nach waren die meisten davon Kaufleute, aber auch Geistliche und Kulturschaffende zählten zu den Gästen in Darmstadt. Welchem Zweck genau diese Besuche dienten, wird nicht immer vermerkt. Es gibt jedoch Fälle, in denen eindeutig auf den geschäftlichen Charakter des Aufenthalts verwiesen wird:

Mercier, Kaufmann aus Frankreich, ist hier mit einem Sortiment von ver=schiednen Waaren angekommen, und bestehet solches in Pariser Parfümerien, Confituren und feinen Liqueurs, Dragees von Verdün, Punsch=Cappillair= und Mandelmilch=Sirop, Himbeeren=Essig, Englischen Tabletten von Menthe, Pastillen a la Mamboise von verschiedenem Geschmack, feinem Pariser Senft, Italieni=schen Blumen und mehreren dergleichen. Er wird sich 8 Tage dahier im Gasthaus zum Engel aufhalten, wo man diese Waaren in seinem Zimmer, eine Stiege hoch No. 3, aufgestellt findet. (Darmstädter Tagblatt, 21.3.1791)

19 Beispiel: Die Daten für den Zeitraum 1745–1750 finden sich hier: <https://cdstar.de.dariah.eu/dariah/EAEA0-CDE2-C0B0-1BED-0> (Stand: 27.11.2025). Eine vollständige Liste der Links ist in Stegmeier/Müller (2024, S. 75) verfügbar.

5. Resümee

Im vorliegenden Beitrag haben wir das Editionsprojekt des Darmstädter Tagblatts (1740–1986) vorgestellt, das an der Universitäts- und Landesbibliothek Darmstadt (ULB) und der Technischen Universität Darmstadt seit 2019 in einem DFG-geförderten Vorhaben realisiert wird. Wir haben über die Aufbereitung des Datensatzes als Korpus berichtet, die OCR, Layoutanalyse, POS-Tagging, Lemmatisierung, Named Entity recognition und die Erfassung morphologischer Informationen umfasst. Ein besonderer Fokus liegt auf der halbautomatisierten Rechteklärung für Einzelbeiträge. Dabei handelt es sich um einen rechtlich wie technisch komplexen Aspekt, der neue Ideen und Verfahren zur Urhebersuche erfordert.

Das dargestellte Projekt will einen Baustein zur wachsenden Sammlung linguistisch aufbereiteter Ressourcen zur Sprachgeschichte des Neuhochdeutschen bieten, die der empirischen Sprachgeschichtsschreibung zahlreiche neue Perspektiven bietet. Durch die Kombination linguistischer Annotation mit detaillierten Metadaten sind diachrone Analysen über Jahrhunderte hinweg möglich. Die Untersuchung lexikalischer oder morphosyntaktischer Muster – etwa der Wandel von Nominalphrasen im Zuge der Transformation des Tagblatts von einem Anzeigenblatt zur Tageszeitung – bietet exemplarische Einblicke in Medien-, Kultur- und Sprachgeschichte. Die Erfassung von Herkunftsorten reisender Personen im 18. Jahrhundert erlaubt raumzeitliche Rekonstruktionen von Mobilität, Reichweite und Kommunikationsräumen – ein bislang kaum erschlossener Zugang zur historischen Soziolinguistik. Insgesamt bietet das Projekt ein Beispiel dafür, wie digitale Editions- und Korpusarbeit mit linguistischer Analyse, rechtlicher Reflexion und kulturwissenschaftlicher Interpretation verknüpft werden kann.

Das Korpus des Darmstädter Tagblatts eignet sich aufgrund des langen Zeitraums, den es kontinuierlich abdeckt, sehr gut für die Untersuchung semantischen und morphosyntaktischen Wandels, wie Müller am Beispiel von *Risiko* gezeigt hat: Graf/Müller (2025) nutzen das Darmstädter Tagblatt als Quelle für eine begriffsgeschichtliche Darstellung von *Risiko*. Müller (2025) untersucht u. a. am Darmstädter Tagblatt die Dynamik der Morphosyntax von *Risiko* und *Gefahr*. Anhand empirischer Korpusdaten wird dort die Beziehung zwischen Flexion, Wortbildung, Phrasensyntax und Diskursfunktion beim Bedeutungswandel analysiert.

Literatur

Evert, Stefan/Hardie, Andrew (2011): Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. Birmingham: University of Birmingham. www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf (Stand: 26.11.2025).

Graf, Rüdiger/Müller, Marcus (2025): Risiko. In: Müller, Ernst/Picht, Barbara/Schmieder, Falko (Hg.): Das 20. Jahrhundert in Grundbegriffen. Lexikon zur historischen Semantik in Deutschland. Basel/Berlin: Schwabe. https://doi.org/10.31267/Grundbegriffe_91070823.

Hardie, Andrew (2012): CQPweb – combining power, flexibility and usability in a corpus analysis tool. In: International Journal of Corpus Linguistics 17, 3, S. 380–409. DOI: [10.1075/ijcl.17.3.04har](https://doi.org/10.1075/ijcl.17.3.04har).

Honnibal, Matthew/Montani, Ines/Boyd, Adriane/Van Landeghem, Sofie/Peters, Henning (2020): spaCy: Industrial-strength natural language processing in Python. <https://explosion.ai/blog/spacy-v3> (Stand: 26.11.2025). [Modell: https://spacy.io/models/de#de_dep_news_trf (Stand: 26.11.2025)].

Klie, Jan-Christoph/Bugert, Michael/Boullosa, Beto/de Castilho, Richard E./Gurevych, Iryna (2018): The INCEpTION Platform: Machine-assisted and knowledge-oriented interactive annotation. In:

- Zhao, Dongyan (Hg.): Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations. Santa Fe, NM: Association for Computational Linguistics, S. 5–9.
- Kollatz, Thomas (2016): Raum-Zeit-Analysen mit Geo-Browser und Datasheet-Editor. In: Bibliothek Forschung und Praxis 40, 2, S. 229–233. DOI: [10.1515/bfp-2016-0032](https://doi.org/10.1515/bfp-2016-0032).
- Müller, Marcus (2025): ‚Morphologie aus Sicht der Diskursgrammatik‘. Am Beispiel der zeithistorischen Morphosyntax von Risiko. In: Michel, Sascha (Hg.): Diskursmorphologie: Ansätze und Fallstudien zur Schnittstelle zwischen Morphologie und Diskurslinguistik. (= Diskursmuster/Discourse Patterns 36). Berlin/Boston: De Gruyter, S. 55–84. <https://doi.org/10.1515/9783111363936-003>.
- Proisl, Thomas/Uhrig, Peter (2016): SoMaJo: State-of-the-art tokenization for German web and social media texts. In: Cook, Paul/Evert, Stefan/Schäfer, Roland/Stemle, Egon (Hg.): Proceedings of the 10th Web as Corpus Workshop. Berlin: Association for Computational Linguistics, S. 57–62. <https://aclanthology.org/W16-2607/> (Stand: 26.11.2025).
- Qi, Peng/Zhang, Yuhao/Zhang, Yuhui/Bolton, Jason/Manning, Christopher D. (2020): Stanza: A Python natural language processing toolkit for many human languages. <https://nlp.stanford.edu/pubs/qi2020stanza.pdf> (Stand: 26.11.2025).
- Schmid, Helmut (1995): Improvements in part-of-speech tagging with an application to German. www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/ (Stand: 26.11.2025).
- Schweter, Stefan/März, Luisa (2020): Triple E - Effective Ensembling of Embeddings and Language Models for NER of Historical German. In: Cappellato, Linda/Eickhoff, Carsten/Ferro, Nicola/Névéol, Aurélie (Hg.): Proceedings of the Working Notes of CLEF 2020 – Conference and Labs of the Evaluation Forum. Thessaloniki, Greece, September 22–25, 2020. Padua: Universität Padua. https://ceur-ws.org/Vol-2696/#paper_173 (Stand: 26.11.2025).
- Stegmeier, Jörn/Müller, Marcus (2024): „Angekommene frembde Herren Passagiers“: Besucher in Darmstadt in drei Dekaden des 18. Jhs.: Ein Werkstattbericht. In: Bartsch, Sabine/Borek, Luise/Hegel, Philipp (Hg.): Computer im Musenhain. Von träumenden Büchern und der Aura des Digitalen. (= Festschriften der Technischen Universität Darmstadt 1.2023). Darmstadt: Technische Universität Darmstadt. DOI: [10.26083/tuprints-00027908](https://doi.org/10.26083/tuprints-00027908).
- Stegmeier, Jörn/Günther, Anne-Christine/Hammer, Angela/Müller, Marcus/Stäcker, Thomas (2022): Eine Zeitung in drei Jahrhunderten: Digitalisierung des Darmstädter Tagblatts. In: Information – Wissenschaft & Praxis 73, 2–3, S. 89–96. DOI: [10.1515/iwp-2022-2210](https://doi.org/10.1515/iwp-2022-2210).
- Urheberrechtsgesetz (UrhG) = Bundesamt für Justiz (2024): Gesetz über Urheberrecht (UrhG) und verwandte Schutzrechte. www.gesetze-im-internet.de/urhg/ (Stand: 8.9.2025).

Kontaktinformationen

Anne Christine Günther
Technische Universität Darmstadt
Magdalenenstraße 8
64289 Darmstadt
E-Mail: anne-christine.guenther@tu-darmstadt.de

Angela Hammer
Technische Universität Darmstadt
Magdalenenstraße 8
64289 Darmstadt
E-Mail: angela.hammer@tu-darmstadt.de

Dario Kampkaspar
Technische Universität Darmstadt
Residenzschloss 1
64283 Darmstadt
E-Mail: dario.kampkaspar@tu-darmstadt.de

Kevin Kuck
Technische Universität Darmstadt
Residenzschloss 1
64283 Darmstadt
E-Mail: Kevin.kuck@tu-darmstadt.de

Anna Christina Kupffer
Technische Universität Darmstadt
Residenzschloss 1
64283 Darmstadt
E-Mail: anna.kupffer@tu-darmstadt.de

Prof. Dr. Marcus Müller
Technische Universität Darmstadt
Residenzschloss 1
64283 Darmstadt
E-Mail: marcus.mueller@tu-darmstadt.de

Prof. Dr. Thomas Stäcker
Technische Universität Darmstadt
Magdalenenstraße 8
64289 Darmstadt
E-Mail: direktion@ulb.tu-darmstadt.de

Dr. Jörn Stegmeier
Universität Trier
54286 Trier
E-Mail: stegmeier@uni-trier.de

Bibliografische Angaben

Dieser Text ist Teil der Publikation: Brunner, Annelen/Hansen, Sandra/Lang, Christian/Tu, Ngoc Duyen Tanja/Wolfer, Sascha (Hg.) (2026): Deutsch im Wandel – Tools, Methoden, Ressourcen. Beiträge zur Methodenmesse der IDS-Jahrestagung 1. (= *IDSopen* 16). Mannheim: IDS-Verlag. <https://doi.org/10.21248/idsopen.16.2026.63>.