

Ressourcenbeitrag: Lyrikline

Jan Horstmann  ¹

1. Universität Münster

forTEXT

Thema:	Korpusbildung	DOI:	10.48694/fortext.3814
Jahrgang:	1	Ausgabe:	2
Erscheinungsdatum:	12-06-2024	Erstveröffentlichung:	2019-07-29 auf forttext.net
Lizenz:			open  access

Allgemeiner Hinweis: Rot dargestellte *Begriffe* werden im Glossar am Ende des Beitrags erläutert. Alle externen Links sind auch am Ende des Beitrags aufgeführt.

1. Kurzbeschreibung

Lyrikline ist eine multimodale, mehrsprachige Plattform für Lyrik. Die Gedichte werden frei online zugänglich gemacht (vgl. [Open Access](#)), mit einem Bild der Autor*innen kombiniert und können zudem von diesen vorgelesen angehört werden. Die übersichtlich gestaltete Webseite ermöglicht u. a. eine gezielte und schnelle Suche nach Kategorien, Autor*innen oder Gedichten und die automatische Auswahl eines Zufallsgedichts.

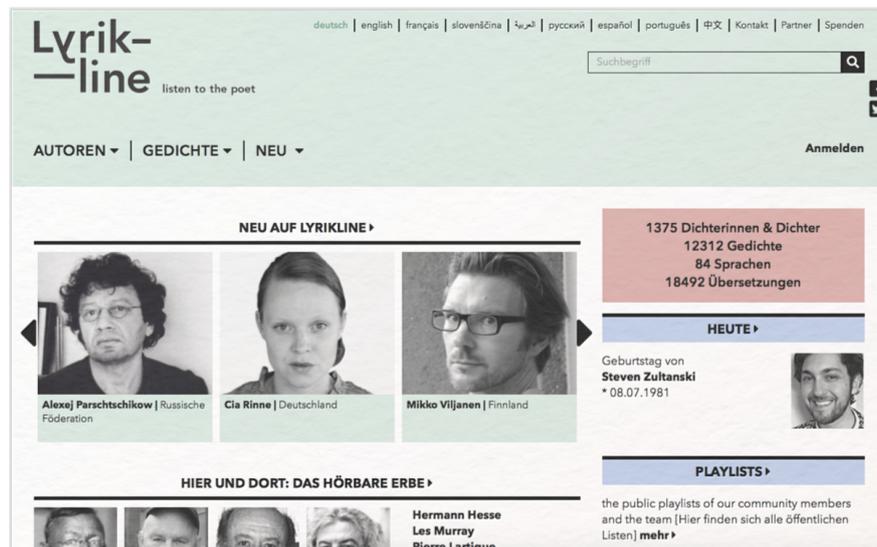


Abb. 1: Startseite von Lyrikline vom 8. Juli 2019

Steckbrief

- <https://www.lyrikline.org>
- 1999 vom Haus für Poesie Berlin initiiert; seit 2003 zahlreiche nationale und internationale Partner, darunter das Goethe-Institut München, die Zentral- und Landesbibliothek Berlin, die Lyrikzeitschrift *Zwischen den Zeilen*, die Anthologie *Jahrbuch der Lyrik*, die Schweizerische Kulturstiftung Pro Helvetia etc.
- Ziel: dem Schrumpfen des lyrischen Segments im Buchhandel sowie dem „Auseinanderfallen von Schrift und Ton auf dem Feld der Lyrik“ (s. Lyrikline: About) entgegenwirken durch eine pluralistische, nicht eurozentrische Sammlung bekannter und unbekannter Autor*innen
- derzeit 1375 vertretene Dichter*innen; 12312 Gedichte in 84 Sprachen; 18492 Übersetzungen (Stand Juli 2019)
- verzeichnete **Metadaten**: Titel, Autor*innen (Geburts- und Sterbedaten, weitere biografische Angaben, weitere Publikationen, Preise, Links), Sprache, Herausgeber, Ort, Verlag, Audioproduktionsdaten, teilweise ISBN-Nummer, Übersetzungen in verschiedene Sprachen
- Schwerpunkte: internationale Lyrik in Schrift und Ton (Originalsprache und ausgewählte Übersetzungen), teilweise auch mit Videos

2. Anwendungsbeispiel

Sie wollen ein thematisches **Korpus** zusammenstellen, das Gedichte enthält, die sich dem Thema Tiere widmen. Mit einem Klick liefert Lyrikline Ihnen 358 Gedichte in unterschiedlichen Sprachen zu diesem Thema, die Sie mit zwei weiteren Klicks auf 144 deutschsprachige und 125 aus Deutschland stammende Gedichte reduzieren können. Sämtliche Gedichte können Sie nicht nur lesen, sondern auch anhören (häufig von den Autor*innen selbst vorgetragen). Zudem werden Ihnen zu etlichen Gedichten diverse Übersetzungen angeboten.

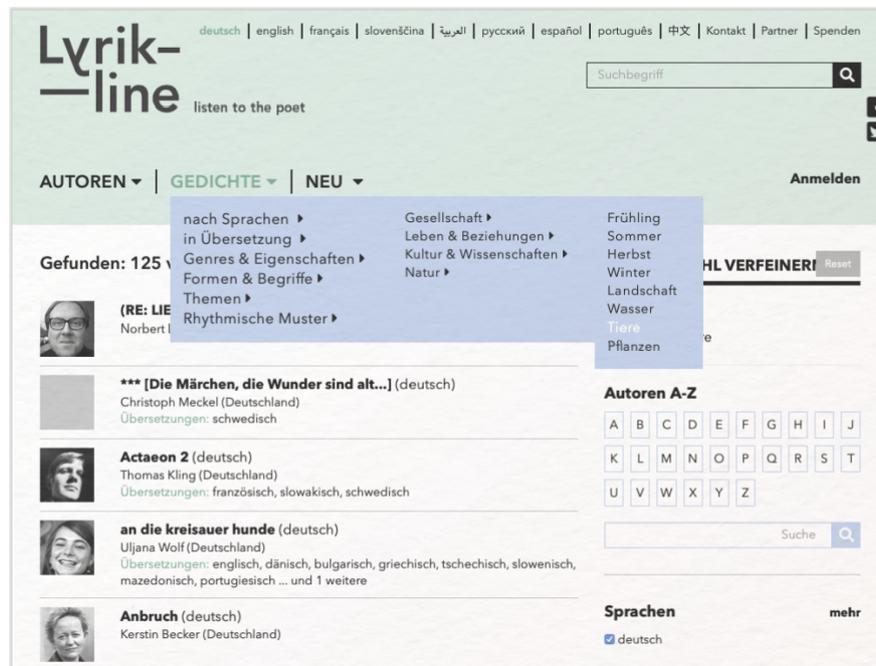


Abb. 2: Filteroptionen für die Gedichte auf Lyrikline

3. Diskussion

3.1 Kann ich Lyrikline für wissenschaftliche Arbeiten nutzen?

Ja, das unter der deutschen UNESCO-Kommission stehende Projekt wird als „exzellente Form einer zeitgemäßen, grenzüberschreitenden Kulturarbeit“ bezeichnet und wurde u. a. 2005 mit dem Grimme-Online-Award in der Sparte „Kultur und Unterhaltung“ ausgezeichnet. Die „verlässliche Plattform“ wird für ihre Sorgfalt gerühmt (Braun 2009). Die Auswahl und Bereitstellung der Gedichte sowie Übersetzungen obliegt den oben genannten Verbundpartnerschaften der jeweiligen Länder, wobei für die Kuratierung Expert*innen für aktuelle Poesie eingesetzt werden und auch für die Übersetzungen auf renommierte Lyrikübersetzer*innen zurückgegriffen wird, um Qualitätsstandards sicherzustellen. Das Projekt vereinheitlicht zudem die Audioaufnahmen in technischer Hinsicht, um eine einheitlich hohe Qualität und Vergleichbarkeit zu gewährleisten. Jedes zur Verfügung gestellte Gedicht ist vollständig, Lyrikline verfolgt jedoch nicht den Anspruch, sämtliche Texte der vertretenen Autor*innen zu verzeichnen. Für Gedichte, bei denen dies wichtig ist (wie beispielsweise der visuellen Poesie), wird ebenfalls eine dokumentnahe Nachempfingung des originalen Seitenlayouts angestrebt, um die Gedichte in ihrer Flächigkeit rezipieren zu können.

Die Quelle des jeweiligen Gedichtes ist angegeben, allerdings fehlen hierbei die Seitenangaben, sodass eine genaue Zitation für eine wissenschaftliche Arbeit einer weiteren Recherche in der jeweils angegebenen Ausgabe bedarf. Lyrikline bietet außerdem keine Downloadfunktionen (etwa in unterschiedlichen Dateiformaten) an, sodass die Texte nur per copy & paste (als Reintext (vgl. **Reintext-Version**) ohne Metadaten) gespeichert werden können.

Hervorzuheben ist die Kooperation von Lyrikline mit dem von der Volkswagenstiftung geförderten **Rhythmicalizer-Projekt** der Freien Universität Berlin. Das Projekt hat - auf Grundlage der amerikanischen Theorien der freien Versprosodie - die in Lyrikline enthaltenen deutsch- und englischsprachigen Audioaufnahmen verwendet und eine Methode zur digitalen Erkennung von Prosodie entwickelt. Die auf der Projektseite besprochenen **Muster** verweisen bereits auf Beispieltexthe aus dem Lyrikline-Korpus. Auf Lyrikline selbst finden sich die Projektergebnisse insbesondere in Form der Filtermöglichkeiten implementiert.

3.2 Wie benutzerfreundlich ist die Arbeit mit Lyrikline?

Lyrikline ist sehr benutzerfreundlich gestaltet und die Inhalte sind frei zugänglich. Die zahlreichen Such- und Filteroptionen (vgl. Query) sind übersichtlich, intuitiv bedienbar und beschleunigen die Textsuche erheblich. Unter der jeweiligen Gedichtüberschrift gibt es immer ein Pull-down-Menü, das weitere Gedichte der jeweiligen Person verzeichnet. Eine kurze Autor*innenbiografie wird immer rechts neben dem Gedicht angezeigt. Die rechte Spalte der Startseite verzeichnet aktuelle Meldungen wie ausgewählte Autor*innengeburtstage des jeweiligen Tages, diverse von Mitgliedern erstellte sog. „Playlists“, den Push-Button, der Ihnen ein Zufallsgedicht anzeigt, ein ausgewähltes Gedicht, das man direkt in der Sidebar anhören kann, die Möglichkeit, der Lyrikline Community beizutreten, und diverse Blogbeiträge. Ein Handbuch zu Lyrikline ist nicht vorhanden, die gesamte Seite ist neben Deutsch jedoch in acht weiteren Sprachen einsehbar, die per Klick in der oberen Zeile ausgewählt werden können.

The screenshot shows the Lyrikline website interface. At the top, there is a navigation bar with the Lyrikline logo and the tagline 'listen to the poet'. A search bar is located on the right. Below the navigation bar, there are tabs for 'AUTOREN', 'GEDICHTE', and 'NEU', along with an 'Anmelden' button. The main content area features the author's name 'ULJANA WOLF' and the title of the poem 'an die kreisauer hunde'. A dropdown menu for 'Gedichte' is visible. To the right of the poem title is a profile picture of Uljana Wolf and her biographical information: '* 06.04.1979, Berlin, Deutschland' and 'lebt in: Berlin, Deutschland'. Below this, there is a short biography: 'Uljana Wolf wurde 1979 in Berlin geboren, wo sie auch heute noch lebt.' and a paragraph about her education and work: 'Sie studiert Germanistik, Anglistik und Kulturwissenschaften an der Humboldt-Universität zu Berlin und arbeitete nebenbei im Buchhandel.' A play button and a progress bar are shown below the poem title. The poem's text is displayed at the bottom: 'o der dorfhunde kleingescheckte schar: schummel schwänze stummelbeine zähe schnauzen am zaun euch gehört die straße der staub am asphaltsaum euch die widerhallende nacht im schlafenden tal'. On the right side, there are sections for 'Publikationen' and 'Preise'.

Abb. 3: Ansicht eines Gedichtes und der Autorinnenseite von Uljana Wolf auf Lyrikline

Es werden zudem weitere Ressourcen zur Verfügung gestellt, die in Zukunft noch ausgebaut werden: So finden sich über die Startseite erreichbar z. B. „Poetische Handreichungen zu zeitgenössischer Lyrik“, in denen derzeit für drei ausgewählte Gedichte didaktische Handreichungen für die Vermittlung von Lyrik im (Schul-)Unterricht bereitgestellt werden. Außerdem gibt es Videoportraits zu ausgewählten Dichter*innen und eine Sammlung an verfilmten Gedichten.

4. Wie funktioniert die Textsuche in Lyrikline?

Gedichte können direkt auf der Startseite nach Autor*innen (alphabetisch, nach Sprachen, nach Ländern) oder nach Gedichten (nach Sprachen, Übersetzungen, Genres & Eigenschaften, Formen & Begriffen, Themen [mit diversen Unterkategorien] sowie rhythmischen Mustern) gefiltert werden (siehe Abb. 2). Zudem verzeichnet der Reiter „NEU“ jüngst hinzugefügte Gedichte, Autor*innen und Übersetzungen. Eine Eingabe im Suchfeld oben rechts auf der Startseite zeigt unmittelbar die Ergebnisse nach verschiedenen Kategorien gegliedert an (s. Abb. 4).

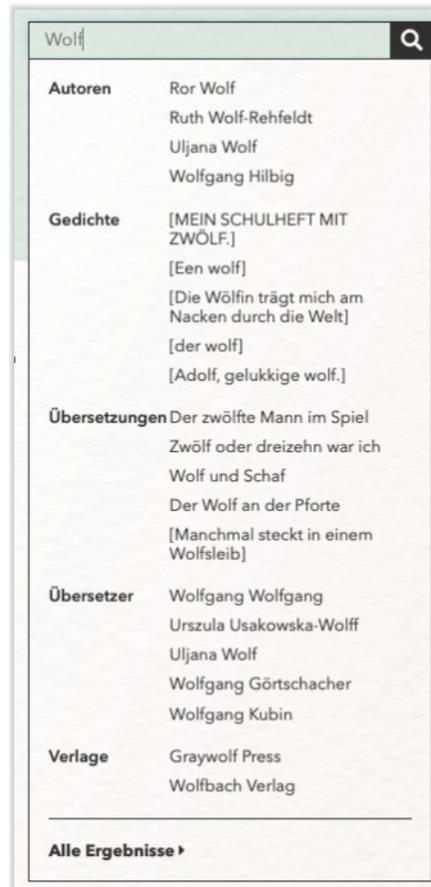


Abb. 4: Freitextsuche auf Lyrikline

- Haus für Poesie Berlin: <https://web.archive.org/save/https://www.haus-fuer-poesie.org/de/literaturwerks-tatt-berlin/home/> (Letzter Zugriff: 04.06.2024)
- Lyrikline: <https://web.archive.org/save/https://www.lyrikline.org/de/about/> (Letzter Zugriff: 04.06.2024)
- Partner Lyrikline: <https://web.archive.org/save/https://www.lyrikline.org/de/partner/> (Letzter Zugriff: 04.06.2024)
- Rhythmicalizer-Projekt: <https://web.archive.org/save/https://www.geisteswissenschaften.fu-berlin.de/v/rhythmicalizer/> (Letzter Zugriff: 04.06.2024)
- Muster Rhythmicalizer-Projekt: <https://web.archive.org/save/https://www.geisteswissenschaften.fu-berlin.de/v/rhythmicalizer/Muster/index.html> (Letzter Zugriff: 04.06.2024)

Bibliographie

- Braun, Michael. 2009. Lyrikline.org. Internationale der Dichter. *Tagesspiegel online*. <https://www.tagesspiegel.de/kultur/literatur/lyrikline-org-internationale-der-dichter/1621806.html> (zugegriffen: 17. Juli 2019).
- Brendel, Gerd. 2009. Hier spricht der Dichter. Das Portal ‚Lyrikline.org‘ feiert zehnjähriges Bestehen. *Deutschlandfunk Kultur*. https://www.deutschlandfunkkultur.de/hier-spricht-der-dichter.1013.de.html?dram:article_id=169718 (zugegriffen: 17. Juli 2019).
- Hardick, Stefanie. 2017. Gedicht-Rhythmik. Dichtung und Daten. *Tagesspiegel online*. <https://www.tagesspiegel.de/themen/freie-universitaet-berlin/gedicht-rhythmik-dichtung-und-daten/19626986.html> (zugegriffen: 17. Juli 2019).
- Seidel, Änne. 2013. Poesie in 58 Sprachen. Relaunch der Gedichte-Datenbank ‚Lyrikline‘. *Deutschlandfunk*. https://www.deutschlandfunk.de/poesie-in-58-sprachen.691.de.html?dram:article_id=260013 (zugegriffen: 17. Juli 2019).

Glossar

Annotation Annotation beschreibt die manuelle oder automatische Hinzufügung von Zusatzinformationen zu einem Text. Die manuelle Annotation wird händisch durchgeführt, während die (teil-)automatisierte

Annotation durch **Machine-Learning-Verfahren** durchgeführt wird. Ein klassisches Beispiel ist das automatisierte **PoS-Tagging** (Part-of-Speech-Tagging), welches oftmals als Grundlage (**Preprocessing**) für weitere Analysen wie Named Entity Recognition (NER) nötig ist. Annotationen können zudem deskriptiv oder analytisch sein.

Browser Mit Browser ist in der Regel ein Webbrowser gemeint, also ein Computerprogramm, mit dem das Anschauen, Navigieren auf, und Interagieren mit Webseiten möglich wird. Am häufigsten genutzt werden dafür Chrome, Firefox, Safari oder der Internet Explorer.

CSV CSV ist die englische Abkürzung für *Comma Separated Values*. Es handelt sich um ein Dateiformat zur einheitlichen Darstellung und Speicherung von einfach strukturierten Daten mit dem Kürzel `.csv`, sodass diese problemlos zwischen IT-Systemen ausgetauscht werden können. Dabei sind alle Daten zeilenweise angeordnet. Alle Zeilen wiederum sind in einzelne Datenfelder aufgeteilt, welche durch Trennzeichen wie Semikola oder Kommata getrennt werden können. In Programmen wie Excel können solche Textdateien als Tabelle angezeigt werden.

HTML HTML steht für *Hypertext Markup Language* und ist eine textbasierte Auszeichnungssprache zur Strukturierung elektronischer Dokumente. HTML-Dokumente werden von **Webbrowsern** dargestellt und geben die Struktur und Online-Darstellung eines Textes vor. HTML-Dateien können außerdem zusätzliche **Metainformationen** enthalten, die auf einer Webseite selbst nicht ersichtlich sind.

Korpus Ein Textkorpus ist eine Sammlung von Texten. Korpora (Plural für „das Korpus“) sind typischerweise nach Textsorte, Epoche, Sprache oder Autor*in zusammengestellt.

Lemmatisieren Die Lemmatisierung von Textdaten gehört zu den wichtigen **Preprocessing**-Schritten in der Textverarbeitung. Dabei werden alle Wörter (**Token**) eines Textes auf ihre Grundform zurückgeführt. So werden beispielsweise Flexionsformen wie „schneller“ und „schnelle“ dem Lemma „schnell“ zugeordnet.

Machine Learning Machine Learning, bzw. maschinelles Lernen im Deutschen, ist ein Teilbereich der künstlichen Intelligenz. Auf Grundlage möglichst vieler (Text-)Daten erkennt und erlernt ein Computer die häufig sehr komplexen Muster und Gesetzmäßigkeiten bestimmter Phänomene. Daraufhin können die aus den Daten gewonnen Erkenntnisse verallgemeinert werden und für neue Problemlösungen oder für die Analyse von bisher unbekanntem Daten verwendet werden.

Markup Language Markup Language bezeichnet eine maschinenlesbare Auszeichnungssprache, wie z.B. **HTML**, zur Formatierung und Gliederung von Texten und anderen Daten. So werden beispielsweise auch **Annotationen** durch ihre Digitalisierung oder ihre digitale Erstellung zu Markup, indem sie den Inhalt eines Dokumentes strukturieren.

Metadaten Metadaten oder Metainformationen sind strukturierte Daten, die andere Daten beschreiben. Dabei kann zwischen administrativen (z. B. Zugriffsrechte, Lizenzierung), deskriptiven (z. B. Textsorte), strukturellen (z. B. Absätze oder Kapitel eines Textes) und technischen (z. B. digitale Auflösung, Material) Metadaten unterschieden werden. Auch **Annotationen** bzw. **Markup** sind Metadaten, da sie Daten/Informationen sind, die den eigentlichen Textdaten hinzugefügt werden und Informationen über die Merkmale der beschriebenen Daten liefern.

Named Entities Eine Named Entity (NE) ist eine Entität, oft ein Eigenname, die meist in Form einer Nominalphrase zu identifizieren ist. Named Entities können beispielsweise Personen wie „Nils Holgerson“, Organisationen wie „WHO“ oder Orte wie „New York“ sein. Named Entities können durch das Verfahren der Named Entity Recognition (NER) automatisiert ermittelt werden.

Open Access Open Access bezeichnet den freien Zugang zu wissenschaftlicher Literatur und anderen Materialien im Internet.

POS PoS steht für *Part of Speech*, oder „Wortart“ auf Deutsch. Das **PoS-Tagging** beschreibt die (automatische) Erfassung und Kennzeichnung von Wortarten in einem Text und ist ein wichtiger **Preprocessing**-Schritt, beispielsweise für die Analyse von **Named Entities**.

Preprocessing Für viele digitale Methoden müssen die zu analysierenden Texte vorab „bereinigt“ oder „vorbereitet“ werden. Für statistische Zwecke werden Texte bspw. häufig in gleich große Segmente unterteilt (*chunking*), Großbuchstaben werden in Kleinbuchstaben verwandelt oder Wörter werden **lemmatisiert**.

Query *Query* bedeutet „Abfrage“ oder „Frage“ und bezeichnet eine computergestützte Abfrage zur Analyse eines Textes. Um Datenbestände zu durchsuchen, werden Abfragesprachen eingesetzt, die *Queries* (Anfragen) an den Datenbestand senden. So bilden alle möglichen Queries zusammen die *Query Language* eines Tools.

Reintext-Version Die Reintext-Version ist die Version eines digitalen Textes oder einer Tabelle, in der keinerlei Formatierungen (Kursivierung, Metadatenauszeichnung etc.) enthalten sind. Reintext-Formate sind beispielsweise TXT, RTF und **CSV**.

Type/Token Das Begriffspaar „Type/Token“ wird grundsätzlich zur Unterscheidung von einzelnen Vorkommnissen (Token) und Typen (Types) von Wörtern oder Äußerungen in Texten genutzt. Ein Token ist also ein konkretes Exemplar eines bestimmten Typs, während ein Typ eine im Prinzip unbegrenzte Menge von

Exemplaren (Token) umfasst.

Es gibt allerdings etwas divergierende Definitionen zur Type-Token-Unterscheidung. Eine präzise Definition ist daher immer erstrebenswert. Der Satz „Ein Bär ist ein Bär.“ beinhaltet beispielsweise fünf Worttoken („Ein“, „Bär“, „ist“, „ein“, „Bär“) und drei Types, nämlich: „ein“, „Bär“, „ist“. Allerdings könnten auch vier Types, „Ein“, „ein“, „Bär“ und „ist“, als solche identifiziert werden, wenn Großbuchstaben beachtet werden.