# Einleitung

Schreiben und Textproduktion sind zweifellos von einer breiten Palette unterschiedlicher Faktoren beeinflusst, die weit über die bloße Beherrschung der Sprache hinausgehen. Diese Faktoren sind vielschichtig und komplex und schließen soziokulturelle Variablen wie Herkunft, Erziehung, Enkulturation, Alter, Geschlecht und viele andere ein. Die Formulierung dieser Aussage mag zunächst absichtlich diffus erscheinen, doch genau hierin liegt der Ausgangspunkt für eine tiefergehende Analyse, die nur durch sorgfältige Untersuchungen und wissenschaftlich-methodische Präzision in kleinen Schritten bestätigt oder widerlegt werden kann.

Das Hauptziel dieser Forschungsarbeit besteht darin, Theorien im Kontext von Schreiben und Textproduktion zu dekonstruieren und sie mithilfe präziser wissenschaftlicher Methoden, einschließlich statistischer Tests, genauer zu untersuchen. Diese Methodik erfordert nicht nur ein Verständnis der vorhandenen Literatur und Forschung, sondern auch die Fähigkeit, diese Erkenntnisse in strukturierte Hypothesen umzuwandeln. Der Forschungsprozess beginnt damit, bestehende Theorien und wissenschaftliche Erkenntnisse in diesem Bereich zu analysieren und zu synthetisieren. Aus dieser Analyse werden dann spezifische Theorien abgeleitet, die als Grundlage für die weiteren Untersuchungen dienen. Diese Theorien werden in präzise Hypothesen umgewandelt, die klar formuliert und testbar sind. Diese Hypothesen werden dann anhand eines sorgfältig ausgewählten Korpus getestet, der für die Forschungsfrage von Relevanz ist. Die Datenerhebung und -analyse erfolgen mit wissenschaftlicher Präzision, um valide Ergebnisse zu gewährleisten. Die Bedeutung dieser Arbeit erstreckt sich weit über die einfache Bestätigung oder Widerlegung von Theorien hinaus. Sie trägt dazu bei, das Verständnis darüber zu vertiefen, wie sich wissenschaftliche Methodik gestaltet und beschäftigt sich mit dem Prozess wissenschaftlicher Erkenntnisfindung. Darüber hinaus fördert sie das Verständnis für die Bedeutung soziokultureller Einflüsse auf sprachliche Prozesse. Insgesamt wird diese Forschungsarbeit dazu beitragen, die Komplexität von Schreiben und Textproduktion in der vorliegenden Stichprobe und den hierbei verwendeten Metriken aufzuschlüsseln. FormularbeginnFormularende

## Das Korpus

Das vorliegende Korpus stellt eine Sammlung von 2429 zeitgenössischen Gedichten deutscher Sprache dar, die von verschiedenen Autorinnen und Autoren im 20. und 21. Jahrhundert verfasst wurden. Diese Gedichte wurden von der öffentlich zugänglichen Website lyrikline.org extrahiert. Neben den eigentlichen Gedichttexten bietet die Website auch wertvolle biographische Informationen über die Autor\*innen. Es ist interessant festzustellen, dass nur wenige Autor\*innen auf dieser Plattform mehr als fünf Gedichte veröffentlicht haben, was die Vielfalt der Schreibenden und ihrer Werke unterstreicht.

### Datenakquise

Die Datenvorbereitung für dieses Korpus war keine einfache Aufgabe und erforderte den Einsatz von Web-Scraping-Tools wie *BeautifulSoup* und *Selenium*, um die Gedichte und zugehörigen Informationen von der Website zu extrahieren. Die anschließende Tokenisierung der Gedichte sowie das POS-Tagging wurden mithilfe von *Spacy* 3.5.3 und dem Modell *de-core-news-md* 3.5.0 durchgeführt. Allerdings gestaltete sich die Ermittlung der Veröffentlichungsdaten der Gedichte als problematisch. Oft waren diese Daten unvollständig oder lediglich mit dem Datum der Audioproduktion auf lyrikline.org versehen. Hier konnte keine zufriedenstellende Lösung gefunden werden, um die Veröffentlichungsdaten präzise und ohne erheblichen Aufwand zu extrahieren, ohne die Gedichte einzeln auf Veröffentlichungsdaten in verschiedensten Publikationen zu prüfen. Aus diesem Grund wurde das Geburtsjahr der Autor\*innen als Ersatz für die Veröffentlichungsdaten verwendet. Diese Herangehensweise ist zweifellos fehleranfällig und kann keine genauen Rückschlüsse auf stilistische Veränderungen in den Gedichten im Laufe der Jahre und Jahrzehnte bieten. Sie dient jedoch als Ausgangspunkt für eine Analyse im Hinblick auf die pädagogische und akademische Erziehung der einzelnen Autor\*innen.

Die vorliegende Datenbank mit den zeitgenössischen Gedichten und den zugehörigen Informationen zu den Autor\*innen ist eine wertvolle Ressource für die Literaturforschung und bietet die Möglichkeit, tiefgehende Analysen im Bereich der Literaturwissenschaft und der Poesie durchzuführen. Sie ermöglicht Einblicke in die Vielfalt literarischer Ausdrucksformen und die Art und Weise, wie verschiedene Autor\*innenverschiedene Themen in ihren Werken interpretieren. Trotz der Herausforderungen bei der Datenvorbereitung ist dieses Korpus ein wichtiges Instrument zur Vertiefung des Verständnisses für die Beziehung zwischen Literatur und Zeitgeist. Es eröffnet zahlreiche Forschungsmöglichkeiten und fördert das Interesse an zeitgenössischer Dichtung.

In dem dieser Arbeit vorangegangenen Gruppenprojekt wurde bereits eine grundlegende Analyse der Struktur sowie formellen und inhaltlichen Metriken eines Teils dieser Gedichte durchgeführt. Die Ergebnisse dieser Untersuchung sollen hier noch einmal kurz zusammengefasst werden. Es ist jedoch wichtig zu betonen, dass zum Zeitpunkt der Projektbearbeitung keine Informationen zur zeitlichen Einordnung der Gedichte verfügbar waren, wie beispielsweise das Geburtsjahr der Autor\*innen oder das Veröffentlichungsjahr der Gedichte. Daher wird diese Arbeit nun einen tieferen Blick auf das Thema Zeit werfen und untersuchen, wie sich die bereits erarbeiteten Ergebnisse mit der bislang unbehandelten Variablen Zeit verknüpfen lassen.

Die Grundtheorie, die zu Beginn der Arbeit erwähnt wurde, scheint im aktuellen wissenschaftlichen Diskurs durchaus relevant zu sein. Soziokulturelle und insbesondere erziehungsspezifische Faktoren, die das Schreiben und die Textproduktion beeinflussen können, sind eng mit der jeweiligen Zeitperiode verbunden. In diesem Zusammenhang wird besonders auf die Unterschiede in der Erziehung von Jungen und Mädchen bzw. Männern und Frauen eingegangen, ohne dabei extreme Präzision anzustreben. Dies kann allein anhand der rechtlichen Entwicklungen im 20. Jahrhundert verdeutlicht werden. Die Emanzipation der Frau im 20. Jahrhundert sowie die kontinuierliche Entwicklung des deutschen Bildungssystems sind Themen, die umfangreiche Studien erfordern und in dieser Arbeit nicht im Fokus stehen. Dennoch soll dieser geschichtliche Hintergrund als Aufhänger für die hier angestrebte Hypothesenbildung dienen.

Die vorliegende Untersuchung wird sich stattdessen darauf konzentrieren, wie sich diese soziokulturellen und erziehungsspezifischen Faktoren im Kontext der Zeitveränderungen manifestieren. Wie haben sich Schreibstile in Gedichten im Laufe der Jahrzehnte geändert? Gibt es Trends oder Muster, die auf Veränderungen in der Gesellschaft, der Bildung oder der Rolle von Frauen und Männern hinweisen? Diese Fragen werden im Zentrum der weiteren Forschungsarbeit stehen und sollen dazu beitragen, das Verständnis für die Wechselwirkungen zwischen Literatur, Zeit und Gesellschaft zu vertiefen.