

Master Angewandte Linguistik  
Vertiefung Organisationskommunikation  
Frühlingssemester 2023

Masterarbeit

**Musterschüler\*innen. Die Entdeckung der Schweizer Gesellschaft im statistischen  
Bildungsdiskurs des 19. Jahrhunderts**

vorgelegt am  
IAM Institut für Angewandte Medienwissenschaften  
Departement Angewandte Linguistik  
ZHAW Zürcher Hochschule für Angewandte Wissenschaften  
am  
9. Juni 2023

Betreuer: Prof. Dr. Philipp Dreesen

Diplomand:  
Michael Kläger  
Heiniweg 8  
CH-8404 Winterthur  
[michael@klaegers.ch](mailto:michael@klaegers.ch)  
+41 76 237 18 27

# Selbständigkeitserklärung

Ich versichere hiermit, dass ich

- die wesentlichen Kernaussagen und Resultate der vorliegenden Arbeit selbst hergeleitet habe,
- die Fachliteratur, auf die sich die Arbeit stützt, selbst recherchiert und rezipiert habe,
- allfällige Daten, die ich in der Arbeit verwende, selbst erhoben oder deren Herkunft im Text klar deklariert habe,
- wissenschaftliche und andere Texte, die ich in der Arbeit wörtlich oder sinngemäss integral oder in Ausschnitten übernehme, im Text gemäss den wissenschaftlichen Standards nachgewiesen und im Literaturverzeichnis aufgeführt habe,
- alle Personen und Institutionen, die mich bei der Arbeit substanzell unterstützt haben, gemäss den Vorgaben der «Richtlinie Wissenschaftliche Qualifikationsarbeiten» (Version vom Nov. 2022, Abschnitt 3) in der vorliegenden Arbeit aufgeführt habe.

Ich bestätige, dass ich das «Merkblatt zur Vermeidung von Plagiaten» der ZHAW vom 19.9.2012 zur Kenntnis genommen habe. Ich bin mir bewusst, dass ein Verstoss gegen die dort aufgeführten Richtlinien eine nachträgliche Aberkennung eines verliehenen Mastertitels zur Folge haben kann.

Ich verpflichte mich, bei einer Publikation der vorliegenden Masterarbeit oder einer Veröffentlichung von Ausschnitten daraus immer klar zu deklarieren, dass es sich um eine Qualifizierungsarbeit von Studierenden handelt. Allgemeine Verweise (z. B. «eine an der ZHAW durchgeführte Studie») reichen dazu nicht aus.

Ort, Datum: Winterthur, 9. Juni 2023

Unterschrift: M. Cläger

# Inhaltsverzeichnis

<b>Abstract</b>	<b>5</b>
<b>1 Einleitung</b>	<b>6</b>
1.1 Hypothese: Digitalisierung im 19. Jahrhundert? . . . . .	6
1.2 Untersuchungsgegenstand und Studiendesign . . . . .	7
1.3 Fragestellung . . . . .	8
1.4 Relevanz für die Angewandte Linguistik . . . . .	8
1.5 Abgrenzung . . . . .	8
1.6 Struktur der Arbeit . . . . .	9
<b>2 Die Entdeckung der Gesellschaft: Statistik, Bildungsreform und Nationsbildung im 19. Jahrhundert</b>	<b>10</b>
2.1 Die Schweiz in den transnationalen Diskursen um Bildung und Statistik . . . . .	10
2.1.1 Die Bildungsexpansion in der Schweiz, 1770–1830 . . . . .	10
2.1.2 Konfliktlinien des schweizerischen Volksschulwesens, 1830–1870 . . . . .	12
2.2 «The avalanche of numbers»: Statistikgeschichte . . . . .	13
2.2.1 Öffentliche Statistik als Funktion einer Digitalisierung . . . . .	14
2.2.2 Schweizer Statistik . . . . .	14
2.2.3 Staat, Schule und Statistik . . . . .	15
2.3 Zwischenfazit . . . . .	16
<b>3 Methodik: Historische Diskursanalyse mit korpuslinguistischen Methoden</b>	<b>17</b>
3.1 Korpuslinguistik . . . . .	17
3.1.1 Was ist ein Korpus? . . . . .	17
3.1.2 «Corpus-based» und «corpus-driven» . . . . .	18
3.2 Diskurs- und Korpuslinguistik . . . . .	18
3.2.1 Was ist ein Diskurs? . . . . .	18
3.2.2 Korpuslinguistische Untersuchung von Diskursen . . . . .	19
3.2.3 Konzepte der Diskursanalyse in der Geschichtswissenschaft . . . . .	19
3.2.4 Begriffsgeschichte und Historische Semantik . . . . .	20
3.2.5 Historische Diskursanalyse . . . . .	21
3.2.6 Digital Humanities und Interesse für quantitative Methoden . . . . .	21
3.3 Korpuslinguistik für kulturanalytische Fragestellungen . . . . .	22
3.3.1 Modellierung des Diskurses in einem Korpus . . . . .	22
3.3.2 Operationalisierung: Frequenzanalyse und Kollokationen . . . . .	23

## Inhaltsverzeichnis

<b>4 Explorative Annäherung</b>	<b>24</b>
4.1 Das Korpus . . . . .	24
4.2 Entwicklung von Begriffen im Zusammenhang mit Statistik . . . . .	25
4.2.1 Diachronische Keywords . . . . .	25
4.2.2 Wörterbuchansatz . . . . .	28
4.2.2.1 <i>tabelle, durchschnitt und prozent</i> . . . . .	31
4.2.2.2 <i>Statistik</i> . . . . .	32
4.3 Zwischenfazit . . . . .	35
<b>5 Hypothesengestützte Annäherung</b>	<b>36</b>
5.1 Verwendungszusammenhänge statistischer Termini . . . . .	36
5.2 Was Statistiken zählen . . . . .	36
5.2.1 Schülerinnen und Schüler . . . . .	36
5.3 Was Statistiken machen . . . . .	39
5.4 Wie Statistiken Kontrolle ausüben . . . . .	42
5.5 Kontrollüberschuss von Statistiken . . . . .	49
5.6 Zwischenfazit . . . . .	50
<b>6 Fazit</b>	<b>51</b>
6.1 Die Digitalisierung der Schweizer Gesellschaft anhand des Schuldiskurses . . . . .	51
6.1.1 Statistikgeschichte . . . . .	51
6.1.2 Statistik im Schuldiskurs . . . . .	51
6.1.3 Doch keine Digitalisierung im 19. Jahrhundert? . . . . .	52
6.2 Das Potential der korpuslinguistischen Methode für die Sozialgeschichte . . . . .	53
6.3 Limitationen . . . . .	53
6.4 Desiderate . . . . .	54
<b>Bibliografie</b>	<b>56</b>
<b>Abbildungsverzeichnis</b>	<b>62</b>
<b>Tabellenverzeichnis</b>	<b>63</b>

# Abstract

The thesis examines the possibilities of writing social history using the methods of corpus linguistics. It will do so with a case study of the Swiss-German educational discourse from around 1800 to the 1870s.

The main hypothesis is that the digitalisation of Switzerland began in the 19th century. After the French Revolution, the idea gained hold that people lived as equal citizens in a nation state. This population could only be described, recorded and controlled by ‘digital’ methods. ‘Rich’ social phenomena were reduced to numbers that were discrete and freely recombinable. These numbers could be combined in new ways to reveal patterns in social behaviour which had been invisible before. The method of generating these descriptions was statistics, which came to be the dominant mode how societies observed themselves. The expansion of elementary education was an important driver of this discovery of latent patterns of social behaviour.

Unlike hermeneutic procedures of text and discourse analysis, this study attempts to pursue the research question from an inductive perspective. It is therefore also a contribution to the methodological discussion of how applied linguistics can provide insights into social history.

Die Arbeit untersucht die Möglichkeiten einer korpusgestützen Sozialgeschichte unter Verwendung korpuslinguistischer Methoden. Grundlegend ist die Hypothese, dass die Digitalisierung der Schweiz bereits im frühen 19. Jahrhundert einsetzt. Diese These wird anhand einer Untersuchung des deutschschweizerischen Bildungsdiskurses von etwa 1800 bis in die 1870er-Jahre verfolgt.

Das Ende der Ständestaaten erzeugte die Vorstellung von territorial definierten Staaten, in denen gleichberechtigte Bürger zusammenlebten. Diese Bevölkerung war nur mit «digitalen» Methoden beschreib-, erfass- und steuerbar: Diskreten, frei re-kombinierbaren Signifikatoren, die «reiche» soziale Phänomene auf Zahlen reduzierten. Diese konnten auf vielfältige Weise in Bezug zueinander gesetzt werden. Die Methode, diese Beschreibungen zu erzeugen, war die Statistik, die sich in diesem Zeitraum als Modus der Selbstbeobachtung moderner Nationen etablierte. Sie erlaubte es, Gesellschaft als Abstraktum zu denken und Muster in scheinbar ungeregelten sozialen Phänomenen zu entdecken.

Anders als hermeneutischen Verfahren der Text- und Diskursanalyse wird versucht, die Fragestellung aus einer induktiven Perspektive heraus zu verfolgen. Sie ist daher, neben der inhaltlichen Untersuchung, ein Beitrag zur Methodendiskussion, wie die Angewandte Linguistik Erkenntnisse zur Sozialgeschichte liefern kann.

# 1 Einleitung

## 1.1 Hypothese: Digitalisierung im 19. Jahrhundert?

Die Digitalisierung der Schweiz beginnt im frühen 19. Jahrhundert. Diese These werde ich in der Arbeit mit einer korpuslinguistischen Untersuchung des deutschschweizerischen Bildungsdiskurses der 1830er bis in die 1870er-Jahre untersuchen.<sup>1</sup>

Ab den 1830er-Jahren intensiviert sich in der Schweiz, wie auch in anderen Ländern Europas, ein Diskurs um die Organisation und Verbesserung der Schulbildung, der in der Spätaufklärung wurzelt. Privatpersonen, Lehrer, Geistliche, liberale und konservative Politiker<sup>2</sup> diskutieren die Rolle der Schule, vom Verhältnis religiöser und beruflicher Bildung über die Lehrerbildung bis zur Mädchenbildung. Dieser Diskurs findet zuvorderst in Zeitschriften und Publikationen statt und er wird vermehrt geprägt von Statistiken. Denn der Untersuchungszeitraum fällt zusammen mit dem Aufkommen der Sozialstatistik als zentralem Mittel gesellschaftlicher Selbstbeschreibung. Eine «avalanche of printed numbers» setzte um 1820 ein, während die Anzahl an veröffentlichten Statistiken rapide anstieg (Hacking, 1982). Die Akteure, die diese Flut an gedruckten Zahlen verursachten, überschneiden sich mit den Akteuren des bildungsreformerischen Diskurses. So war der spätere Bundesrat Stefano Franscini sowohl Autor einer der ersten statistischen Beschreibungen der Schweiz wie auch Lehrer und als Politiker der Motor kantonaler und föderaler Bildungsreformen (Marcacci, 2022).

Dass dieser Diskurs öffentlich geführt wird und Statistiken darin eine wichtige Rolle spielen, ist eine Funktion der Digitalisierung der Schweizer Gesellschaft. Die enge Verbindung von Schulreform und Statistik ist kein Zufall, sondern hat gemeinsame Wurzeln in den politisch-sozialen Umbrüchen des frühen 19. Jahrhunderts, zuvorderst der Durchsetzung des «modernen» Staates mit dem liberalen Ideal rechtlich gleichgestellter Staatsbürger und einer Ausdifferenzierung der Gesellschaft infolge der Frühindustrialisierung. Mit Armin Nassehi gehe ich davon aus, dass die Entwicklung der sozialkundlichen Statistik seit dem frühen 19. Jahrhundert als Funktion einer Digitalisierung der Gesellschaft beschrieben werden kann. Ihm zufolge stehen die aufkommenden Nationalstaaten vor einem Problem: Das Ende der Ständestaaten erzeugte die Vorstellung von territorial definierten Staaten, in denen gleichberechtigte Bürger zusammenlebten. Diese Bevölkerung war nur mit «digitalen» Methoden beschreib-, erfass- und steuerbar: diskreten, frei re-kombinierbaren Signifikatoren, die «reiche» soziale Phänomene auf Zahlen reduzierten, die miteinander auf vielfältige Weise in Bezug gesetzt werden konnten (Nassehi, 2019, S. 63).

1 Für die Recherche des Forschungsstandes zu dieser Arbeit wurde die generative KI «ChatGPT» eingesetzt. Es wurden keine sinngemäßen oder wörtlichen Zitate von ChatGPT entnommen.

2 Hier liegt kein generisches Maskulinum vor. Die Diskursteilnehmer sind, soweit ich es überblicken kann, fast ausschließlich männlich.

Digitalisierung wird also funktionalistisch verstanden. Sie ist eine Art und Weise der gesellschaftlichen Selbstbeschreibung, die auf die Anforderungen der modernen Gesellschaft reagiert. Konkret zeigt sich Digitalisierung der Gesellschaft am Prozess der Herstellung solcher abstrakter Signifikatoren. Diese Signifikatoren werden in öffentlichen Diskursen verhandelt und zur Selbstbeschreibung einer Gesellschaft herangezogen: Die unübersichtliche Gesellschaft wird beschreibbar als Zusammenspiel der aggregierten Kategorien, die Regelmässigkeiten entdecken lässt. Für den vorliegenden Fall lässt sich das anschaulich machen. Schulen, Schülerinnen und Schüler, Lehrerinnen und Lehrer werden gezählt. Dazu müssen sie in Kategorien eingeteilt werden, die zueinander in Bezug gesetzt werden können. Das erlaubt, gemeinsame Muster im Verhalten der Individuen zu beschreiben. Die unübersehbare Vielzahl der Individuen wird intellektuell und sprachlich handhabbar und als Gesellschaft beschreibbar. Dieser Prozess schlägt sich diskursiv nieder, was diese Arbeit anhand einer quantitativ-qualitativen Analyse der bildungsreformerischen Diskurse der deutschsprachigen Schweiz im 19. Jahrhundert versuchen wird zu zeigen.

## 1.2 Untersuchungsgegenstand und Studiendesign

Untersucht wird ein bereits bestehendes Korpus aus gedruckten Quellen, die zwischen dem 1. Januar 1800 und dem 31. Dezember 1870 publiziert wurden und das mit den Suchbegriffen *schulbildung*, *schulwesen*, *kantonalschul*, *volksschul*, *primarschul* aus Digitalisaten der Plattformen e-rara.ch und e-periodica.ch zusammengestellt wurde.

Die Arbeit wird in einem ersten Schritt dieses Korpus mit quantitativen Methoden der Korpuslinguistik auf statistisch signifikante Phänomene auf der «Textoberfläche» untersuchen. Das heisst, es wird nach signifikanten Zeichenfolgen gesucht und zunächst versucht zu vermeiden, diese semantisch zu bestimmen. Die Ergebnisse dieser ersten statistischen Exploration sollen dazu dienen, die Abfragen zielgerichteter auf bestimmte sprachliche Muster einzuschränken. Die explorative Untersuchung liefert darüber hinaus Belegstellen für qualitative Betrachtungen, ob sie auch auf semantischer und inhaltlicher Ebene die angenommenen diskursiven Entwicklungen bestätigen. Anschliessend werden die Ergebnisse der Exploration ergänzt um hypothesesgestützte Analysen, die versuchen aufzuzeigen, wie Schulstatistiken eine Vorstellung von Gesellschaft erzeugen.

Dieser Ansatz entspricht einem *mixed-methods*-Verfahren mit zwei parallelen Strängen, einem eher quantitativen und *corpus-driven* Schritt und einem parallelen, aber textlich anschliessenden *corpus-based* Schritt. Eine Triangulation findet dadurch sowohl auf erkenntnistheoretischer Ebene (Erklären vs. Verstehen) als auch auf methodischer Ebene (statistische Analyse vs. Hermeneutik) statt. Durch die multiperspektivische Analysen, die weitere Analysen informieren, soll die Aussagekraft der Ergebnisse verbessert werden (Dreesen und Stücheli-Herlach, 2019, S. 139).

## 1.3 Fragestellung

Der besprochene Korpus wird im Hinblick auf folgende Fragen untersucht:

1. Gibt es beobachtbare Veränderungen auf der sprachlichen Oberfläche des Diskurses, die darauf hinweisen, dass die sozialkundliche Statistik Einfluss auf die gesellschaftliche Selbstbeschreibung gewinnt?
2. Werden Fragen und Probleme des bildungskundlichen Diskurses quantifiziert?
3. Welche Kategorien werden für die Quantifizierung eingesetzt?
4. Wie werden diese Kategorien sprachlich zueinander in Bezug gesetzt?

## 1.4 Relevanz für die Angewandte Linguistik

Diese Arbeit ist auf den ersten Blick nicht der Angewandten Linguistik im engeren Sinn zuzuordnen. Sie entspricht in Anlage und Erkenntnisinteresse mehr den Digital Humanities. Der Bezug zur Linguistik liegt in der verwendeten Methodik: Die Arbeit setzt auf Methoden der Korpuslinguistik und der Untersuchung kultur- und sozialgeschichtlicher Phänomene, die sich in den letzten Jahren etabliert haben (Bubenhofer, 2009). Die im Folgenden aufgezeigten und umgesetzten Arbeitsschritte der Diskursmodellierung und -analyse stammen aus der Angewandten Linguistik und lassen sich auch auf Fragestellungen anwenden, wie sie beispielsweise in der Organisationskommunikation vorkommen, auch wenn sie hier an einem historischen Thema vorgeführt werden. Das theoretische Wissen über den Sprachgebrauch in Diskursen ermöglicht einen reflektierteren Sprachgebrauch. Im Berufsalltag ist Wissen über die vorherrschenden Formulierungen und Themen eines Diskurses relevant, denn es ermöglicht eine anschlussfähige Kommunikation. Die theoretischen Überlegungen zu Diskursen und ihre Untersuchungsmöglichkeiten mit korpuslinguistischen Diskursen sind daher berufsrelevant. Diskurse lassen sich zwar nicht zielgerichtet manipulieren, aber eine empirisch gestützte Untersuchung liefert mögliche Optionen: Welche sprachliche Äusserung ist innerhalb eines Diskurses anschlussfähig, welche Akteure können eher als andere erreicht werden (Dreesen und Stücheli-Herlach, 2019)?

## 1.5 Abgrenzung

Das Erkenntnisinteresse der Arbeit ist primär ein sozialhistorischer Blick auf den Diskurs über die Statistik, wie er in den bildungskundlichen Medien der Deutschschweiz im 19. Jahrhundert geführt wurde und welche Effekte er auf gesellschaftliche Selbstbeschreibungen hatte. Auch wenn dafür die Geschichte des schweizerischen Volksschulwesens und Kenntnisse der Geschichte der Statistik unabdingbar sind, stehen sie nicht im Fokus. Da hier Analysemethoden der Angewandten Linguistik auf einen bisher kaum untersuchten Gegenstand angewendet werden, stehen inhaltliche Erkenntnisse zum historischen Phänomen im Vordergrund. Dagegen geht es weniger um eine methodisch-theoretische Weiterentwicklung dieser Forschungsansätze oder um Probleme der Angewandten Linguistik.

## 1.6 Struktur der Arbeit

Im folgenden Kapitel 2 wird zunächst der historische Kontext des Untersuchungsgegenstandes aufgearbeitet, die Geschichte des Schweizer Volksschulwesens und ihr enger Zusammenhang mit der Entstehung des modernen Staates, dessen Bewohner als einer Gesellschaft zugehörig verstanden werden und welche Bedeutung der Statistik darin zukam. Kapitel 3 skizziert den theoretischen Hintergrund der Diskursanalyse und Korpuslinguistik. Daran schliesst der Versuch einer induktiven Analyse des Korpus in Kapitel 4 an. In Kapitel 5 werden die Ansätze der induktiven Analyse weiterverfolgt und mit einem hypothesen-basierten Verfahren ergänzt. Ein kritisches Fazit 6 sowohl der erzielten Ergebnisse als auch der Methodik schliesst die Arbeit ab.

# **2 Die Entdeckung der Gesellschaft: Statistik, Bildungsreform und Nationsbildung im 19. Jahrhundert**

## **2.1 Die Schweiz in den transnationalen Diskursen um Bildung und Statistik**

### **2.1.1 Die Bildungsexpansion in der Schweiz, 1770–1830**

Im letzten Drittel des 18. Jahrhunderts lässt sich in Europa eine Transformation des Bildungswesens beobachten. Nicht nur nimmt die private Nachfrage nach Bildungsangeboten deutlich zu, sondern gleichzeitig interessieren sich auch die Regierungen dafür, was in den bislang religiös geführten Bildungseinrichtungen geschieht. Heinrich Busse sieht hier einen doppelten Prozess: eine Revolution von unten, bei der Bildung einen sozialen Aufstieg ermöglicht und deshalb von den Akteur\*innen immer mehr nachgefragt wird sowie eine Kulturrevolution von oben, in der die aufgeklärte Bürokratie das Schulwesen rational vereinheitlichen und «zukunftsfähig» machen will (Bosse, 2012, S. 50, 59).

Auch in der Schweiz lässt sich dieser Doppelprozess nachweisen. Um 1770 intensiviert sich das Interesse an der Schulbildung, besonders an den Landschulen. Auch hier gibt es spätaufklärerische Stimmen, die ihre Ideen in der Landbevölkerung verbreiten möchten und eine Verbesserung der Bildung fordern. Ebenso regulieren die Obrigkeiten nun das Schulwesen stärker. Im Kanton Zürich wird 1771/1772 erstmals eine Schulumfrage aller Landschulen durch Private organisiert, 1778 erlässt die Zürcher Kantonsregierung eine Landschulordnung (Bloch Pfister, 2007, S. 48–63). In der Stadt Zürich entstehen parallel dazu neue staatliche und private Bildungseinrichtungen. War höhere Bildung bisher vor allem religiös verstanden und zielte auf die Ausbildung von Pastoren gibt es nun neue Lehranstalten, die juristische, handwerkliche oder kaufmännische Fähigkeiten vermitteln (Tröhler, 2011, S. 156–157).

Träger dieses Prozesses von unten waren Anhänger der Aufklärung, die sich in sozial- und moralreformerischen Vereinen wie der Helvetischen Gesellschaft versammeln. Einflussreich waren etwa Heinrich Zschokke oder Johann Heinrich Pestalozzi, der seine Pädagogik in mehreren Reformschulen zu vermitteln suchte (Graf, 2022, Bütkofer, 2006, S. 68).

Dieser Ausbau des schweizerischen Bildungswesens war massgeblich geprägt vom politischen Grosskonflikt der Zeit zwischen, vereinfacht zusammengefasst, liberalen Anhängern eines säkularisierten, demokratischen Bundesstaats mit Volkssouveränität und Volksrechten auf der einen Seite und auf der anderen Seite konservativen Vertretern, die auf der Souveränität der Kantone und dem Erhalt ständischer Privilegien und einer verankerten Rolle der Religion im Staatswesen beharrten.

In der Helvetischen Republik von 1798 bis 1803 erlebte die Bildungs- und Schulreformen in der Schweiz einen ersten Durchbruch. Die nach dem Einmarsch französischer Revolutionstruppen 1798 geschaffene Republik ersetzt die vorher nur lose verbundenen Kantone durch einen Einheitsstaat. Nach französischem Vorbild geht das Erziehungswesen in die Hand eines zentralen staatlichen Erziehungsdepartements, dessen Vorsteher Johann Heinrich Stämpfli eine Schulreform anstrebt. Erstmals wird Schweizweit eine allgemeine Schulpflicht eingeführt. Das Schulwesen wird ganz unter die Aufsicht des Staates und nicht mehr der Kirchen gestellt. Die Schulformen und die Lehrerausbildung sollen vereinheitlicht und professionalisiert werden (Bütikofer, 2006).

Die Helvetische Republik organisiert zu diesem Zweck die erste landesweite öffentliche Schulumfrage der Schweiz, die heute als Stämpfli-Enquête bekannt ist. Im Januar 1799 verschickt das Departement einen Fragebogen an sämtliche Lehrer der Schweiz, von dem heute 2400 handschriftliche Antwortbögen überliefert sind. In 16 Fragen mit mehreren Unterfragen werden Angaben zu den Schulen erfragt, von der Schulträgerschaft über die Grösse der Schule und den Schulbesuch, die Lehrinhalte, die Unterrichtsdauer, die Situation der Lehrerschaft und die wirtschaftliche Situation der Schule (Schmidt et al., 2015, Tröhler, 2014).

Die Stämpfli-Enquête und weitere kantonale Schulumfragen der Epoche verweisen auf das frühe Interesse in der Schweiz nach dem Zustand der Volksbildung und die Bedeutung, die einer systematischen empirischen Datenerhebung zugemessen wurde. Sie zeigen, dass auch in der Schweiz die sich entwickelnden Methoden der Statistik rezipiert und erprobt wurden, bereits vor Beginn der Helvetik (Rothen und Ruloff, 2014). Die vorwiegend qualitativen und differenzierten Antworten vieler beantworteter Fragebögen der Stämpfli-Enquête zeigen aber auch, dass das resolute Kategorisieren und Reduzieren sozialer Phänomene auf eindeutige Zahlenwerte wenig ausgeprägt war (Tabelle 2.1).

Mit Ausnahme der Einsetzung kantonaler Schul- und Erziehungsräte wurden die weitgehenden Reformpläne der Helvetischen Republik kaum umgesetzt. Mit dem Ende der Republik 1803 wurde die Schulpolitik rekantonalisiert und der Einfluss Geistlicher auf die Schulen nahm wieder zu. Allerdings wiesen die Tendenzen weiterhin in Richtung eines Ausbaus des Volksschulwesens. Darin waren sich Konservative wie Liberale einig, differierten aber in Umsetzungs- und Detailfragen wie dem Einfluss der Kirche oder der Lehrerbildung (Bütikofer, 2006, S. 233–234). Auf kantonaler Ebene lassen sich denn auch weitere Entwicklungen hin zu einem umfassenden Volksschulwesen im frühen 19. Jahrhundert ausmachen. Besonders in neu entstandenen Kantonen wie der Waadt sollte die Schule zur Identitätsbildung einer «kantonalen Nationalkultur» beitragen (Rothenbühler, 2010). Die Kantone sind, vor dem erst 1848 gegründeten Bundesstaat, ein wichtiger Referenzpunkt für Vorstellungen von Staat und Gesellschaft in der Schweiz im 19. Jahrhundert. Etwa wurde das Schulhaus

in ländlichen Gegenden oft als Symbol für eine unerwünschte kantonale Zentralmacht über die Gemeinden gesehen (Criblez, Hofstetter und Magnin, 1999, S. 16).

III.12	Schulkinder. Wie viele Kinder besuchen überhaupt die Schule?	
III.12.a	Im Winter. (Knaben/Mädchen)	50. Schulkinder überhaupt besuchen die Schule. im Winter. Namlich. 28. Knaben. 22. Mädchen.
III.12.b	Im Sommer. (Knaben/Mädchen)	Jm Sommer ist es sehr ungleich wegen der anzahl der Schul- Kinder. es werden Wochentlich 3. mahl Catichismus übungen gehalten, namlich. Mitwochen Samstag, und Sonntag. Zu wünschen were es das die Kinder bey diesen gesetzten Stunden, sich zahlreicher einfinden würden.

TABELLE 2.1: Beispielantwort der Stapfer-Enquête für Nr. 165, Oberembrach, Schmidt et al., 2015. Verfügbar 12. Februar 2023 unter <http://www.stapferenquete.ch/db/165>.

## 2.1.2 Konfliktlinien des schweizerischen Volksschulwesens, 1830–1870

Als eigentlicher Durchbruch eines staatlichen Schulwesens mit Schulpflicht in der Schweiz gilt der Forschung die Zeit um 1830, als liberale Bewegungen sich in zahlreichen Kantonen durchsetzen und neue Kantonsverfassungen erringen konnten. Dort kam es in der Folge zu neuen Schulgesetzen und dem Versuch, das Volksschulwesen auszubauen, zu professionalisieren und einen allgemeinen Schulbesuch zu verordnen (siehe die Beiträge in Criblez, Jenzer et al., 1999).

Die Liberalen sahen in der Volksbildung eine wichtige staatliche Aufgabe, auch um den Einfluss von Konservativen und Religiösen auf die Bildung zurückzudrängen. Die Durchsetzung der Schulpflicht blieb im 19. Jahrhundert einer der massgeblichen Konfliktlinien zwischen liberalen Schulreformern sowie Schulbehörden auf der einen Seite und Teilen der Landbevölkerung, Geistlichen und Konservativen auf der anderen Seite. Für eine Zurückhaltung gegenüber dem Schulbesuch gab es aus Sicht der Landbevölkerung wirtschaftliche wie auch kulturelle Gründe: Kinder älter als 10 Jahre wurden vor allem im Sommer als landwirtschaftliche Arbeitskräfte benötigt und der wirtschaftliche Wert einer Schulbildung erschien den Erziehungsberechtigten oft zweifelhaft. Für Bildungs- und Sozialreformer wurde mangelhafter Schulbesuch zu einer Chiffre für die Rückständigkeit gesellschaftlicher Gruppen und damit zur Legitimation von Interventionen (Scandola, 1991, S. 602).

Im 1848 neu gegründeten Bundesstaat blieb das Schulwesen in der Hand der Kantone. Dass die Schulbildung in staatlicher Hand liegen sollte und eine allgemeine Schulpflicht bestand, war zu diesem Zeitpunkt auch unter Konservativen nicht mehr bestritten. Die revidierte Bundesverfassung von 1874 konnte das bis dahin in den Kantonen erreichte festschreiben: die Pflicht zum Schulbesuch bis zur Primarstufe in staatlichen und konfessionell neutralen öffentlichen Schulen (Grunder, 2012).

Weitere Diskussionen betrafen mehr die innere Gestaltung der Schulen, den Ausbau der Schulformen und der sekundären Bildung, die Armen- und Waisenbildung sowie den Mädchenunterricht (siehe die Tabelle der von der Schweizerischen Gemeinnützigen Gesellschaft besprochenen pädagogischen Themen bei Criblez, 2013, S. 173). Um 1870 kann daher die staatliche Volksschule als in der Schweiz breit akzeptierter Normalfall gelten (Tröhler, 2011, S. 167).

## 2.2 «The avalanche of numbers»: Statistikgeschichte

Im 19. Jahrhundert wurde die Statistik zum massgeblichen Modus gesellschaftlicher Selbstbeobachtung und Selbstbeschreibung, was sich parallel zur Entwicklung des modernen Staates und dem Aufbau eines staatlichen Schulwesens vollzog. Erst das 19. Jahrhundert dachte in *nationalen* Gesellschaften, deren Wachstum, Reichtum oder Entwicklung mit Zahlen beschrieben werden konnte (Osterhammel, 2010, S. 57–62).

Vorläufer waren im deutschsprachigen Raum die frühneuzeitlichen Polizeywissenschaften. Michel Foucault macht im 17. Jahrhundert einen Umbruch in den europäischen Regierungsstilen aus, indem für den Staat die genaue Kenntnis des Landes, seiner Bevölkerung und seiner Ressourcen ins Zentrum der Regierungstätigkeit rückt und die Regierungen Informationen zur Stärke ihres Landes sammelten. Als wichtiger Faktor zur Einschätzung staatlicher Macht war dieses Wissen geheim. Es konnte also nicht öffentlich diskutiert werden und damit gesellschaftliche Selbstbeschreibung prägen (Foucault, 2004, S. 395–400).

Im 18. Jahrhundert begannen sich Private und Gelehrte für den Zustand der Staaten, der Wirtschaft oder der Bevölkerung zu interessieren. Der Begriff Statistik wurde 1749 geprägt, hatte aber damals noch keinen exklusiven Bezug zu Zahlen oder einer tabellarischen Darstellung, die wir heute mit Statistiken verbinden. Vielmehr präsentierten die statistischen Länderbeschreibungen eine Mischung geografischer, historischer und anderer Informationen meist in narrativer Form, um etwa eine Region zu beschreiben. Erst zu Beginn des 19. Jahrhunderts setzte sich allmählich die Vorstellung der Statistik als numerischer Wissenschaft durch (Hacking, 1990, S. 16–26).

Die Mathematisierung der Statistik begann aus dem Interesse an der Bevölkerungszahl. In England analysierten Pioniere der Volkszählungen Pfarreiregister und leiteten daraus Geburts- und Sterberaten ab, die fundiertere Schätzungen der Bevölkerung und ihrer Entwicklung ermöglichten. Um 1800 begannen in Grossbritannien und Frankreich staatlich unterstützte statistische Datenerhebungen zu Bevölkerungszahlen, Gesundheitsstatistiken und Zahlen zu Institutionen wie Gefängnissen. Sie entdeckten, dass sich soziale Phänomene nicht zufällig ereigneten, sondern überindividuelle Muster im Verhalten zeigten, obwohl die Entscheidungen den einzelnen Individuen frei erschienen. Beispielsweise war nicht nur die Zahl der Suizide in Gesellschaften über die Jahre hinweg bemerkenswert konstant, die Zahlen schwankten auch regelmässig im Jahresverlauf. Sogar die Arten des Suizids wiesen stabile Muster auf. Diese Regelmässigkeit erlaubte die Annahme, dass «Gesellschaft» ein stabiles Beobachtungsobjekt bildete, dessen Muster Einsichten in die Ursachen individuellen Verhaltens ermöglichte (Hacking, 1990, S. 75–76). «Soziale Probleme» wurden entdeckt,

über die Statistiken Aufschluss geben konnten, zum Beispiel anhand von Kriminalitätsraten oder Sterblichkeitsursachen (Porter, 1986, S. 27–31).

Aus dieser Erkenntnis folgte ein «enthusiastisches Zeitalter» der Statistik: Europäische Staaten, gemeinnützige Gesellschaften und Sozialreformer erhoben Zahlen zu allen Arten sozialer Phänomene, welche breit rezipiert wurden. Ian Hacking bezeichnet diese Phase als «avalanche of printed numbers», während der von 1820 bis 1840 die Anzahl gedruckter Zahlen exponentiell gestiegen sei (Hacking, 1982, S. 282). Grundlegend ist für Hacking eine Transformation des Verständnisses dieser Zahlen. Für die Zeitgenossen bilden sie so etwas wie Naturgesetze sozialen Verhaltens ab, die Ansatzpunkte für soziale Interventionen sind. Damit die Dingezählbar werden, müssen sie kategorisiert werden. Die Kategorisierung, Zählung und anschliessende Publikation dieser Zahlen hatte wiederum entscheidende Implikationen für das Selbstverständnis der so beschriebenen Gesellschaft. Die Beschreibung einer Gesellschaft als Klassengesellschaft etwa begünstigt die Identifikation der so Beschriebenen mit einer ebendieser Klasse (Hacking, 1982).

### **2.2.1 Öffentliche Statistik als Funktion einer Digitalisierung**

Für Armin Nassehi ist die Popularität der öffentlichen Sozialstatistik die «erste Entdeckung der Gesellschaft» und eine Digitalisierung der Gesellschaft. Digitalisierung meint hier, dass die Gesellschaft lernt, sich selbst «digital zu sehen». Digital sehen, das bedeutet Nassehi zufolge, dass reiche, analoge soziale Phänomene auf eine Form, hier Zahlen, reduziert werden, die einen Vergleich und eine Rekombination ermöglichen und damit latente Muster aufdecken und vorhersehbar machen. Das öffentliche Teilen und diskutieren dieser Zahlen ermöglicht ein gesellschaftlich geteiltes Verständnis der Gesellschaft, die sich als Ensemble dieser Kategorien begreift (Nassehi, 2019, S. 32–36).

Die europäischen Gesellschaften des frühen 19. Jahrhunderts sind mit einem Problem konfrontiert, für das Digitalität eine Lösung oder zumindest Erklärungen anbietet. Liess sich früher im *Ancien Régime* aus der ständischen Geburt meist zuverlässig auf Welt- und Wertvorstellungen sowie das Schicksal eines Menschen schliessen, sind die Bürger in den neuen Nationalstaaten nominell frei, ihr Leben zu gestalten. Wie die Statistiker jedoch erstaunt feststellten, treffen sie ihre Entscheidungen zu heiraten oder sich selbst umzubringen mit einiger Regelmässigkeit (Nassehi, 2019, S. 44–47). Können wir in unserem vorliegenden Korpus nachweisen, dass die Teilnehmer des Deutschschweizer Diskurses zur Volksschulbildung lernen, «digital zu sehen»?

### **2.2.2 Schweizer Statistik**

Während der Forschungsstand zur Geschichte des Schweizer Schulwesens umfassend ist, ist der Forschungsstand zur Geschichte der Statistik in der Schweiz wesentlich dünner. Ähnlich wie bei der Bildungsgeschichte war aber auch hier das Gebiet der heutigen Schweiz in die europäischen Diskurse zur Entwicklung der Statistik eingebunden. Daher könnte diese Arbeit einen Beitrag zum Statistikwesen in der Schweiz von 1820 bis 1870 leisten.

Die ältere staatswissenschaftlich-qualitative Statistik wurde im Gebiet der heutigen Schweiz während des 17. und 18. Jahrhunderts beispielsweise auch in den Kantonen Bern oder Zürich betrieben (Pfister, 1995). Auch diese ältere beschreibende Statistik interessierte sich für die Schulen, wie die Zürcher Schulumfrage von 1770 zeigt (Bloch Pfister, 2007, S. 50–55). Die schon erwähnte Stapfer-Enquête zum Zustand der Schulen war nur eine von mehreren Untersuchungen, die die Helvetische Republik ab 1799 initiiert hatte, um den Zustand des Landes und mögliche Probleme in Erfahrung zu bringen. Damit folgte sie ganz dem französischen Beispiel, wo der post-revolutionäre Staat statistische Büros eingerichtet hatte (Holenstein, 2014).

Zu einer Einrichtung eines statistischen Büros für die ganze Schweiz kam es während der Helvetik jedoch nicht. Dafür war die Helvetische Republik zu kurzlebig und die Zentralregierung zu schwach. Gängig ist deshalb das Urteil, die Schweizer Statistik habe nach der Helvetik den Anschluss an europäische Entwicklungen verloren, also gerade in der Phase von 1820 bis 1850, während der sich die Sozialstatistik etablierte (Jost und Malaguerra, 2016, S. 17–22). Dieses Urteil fällt meist mit Blick auf die sehr späte Etablierung eines bundesweiten Eidgenössischen Statistischen Büros, dem heutigen Bundesamt für Statistik. Dieses wurde erst 1860 eingerichtet, also zwölf Jahre nach der Gründung des Schweizerischen Bundesstaates 1848 (Jost und Malaguerra, 2016, S. 25).

Einzelne Kantone und Private betrieben aber durchaus statistische Erhebungen, die sich an sozialreformerischen Statistikbewegungen andernorts orientierte. Für die Liberalen waren der Aufbau eines staatlich dominierten Volksschulwesens und die öffentliche Statistik miteinander verbundene Projekte. So überschnitten sich beispielsweise die Personenkreise, die sich in beiden Diskursen engagierten. Beispiele dafür sind etwa Stefano Franscini. Nach seiner Tätigkeit als Lehrer wandte er sich der Statistik zu. 1828 publizierte er auf Italienisch eine der ersten statistischen Beschreibungen der Schweiz und war gleichzeitig in Vereinen zur Volksbildung engagiert. Als Mitglied des ersten Bundesrates ab 1848 war er als Leiter des Departements des Inneren für die Statistik zuständig und versuchte 1850, eine erste Volkszählung der Schweiz zu lancieren (Marcacci, 2022). Eine wichtige Plattform stellte die Schweizerische Gemeinnützige Gesellschaft dar. Sie bot sowohl ein Forum für Diskussionen über das Schulwesen als auch der Statistik bot. Gegen Ende des 19. Jahrhunderts hatte sie den Anspruch, durch ihre Aktivitäten eine Akteurin in der «Erfindung der Schweizer Nation» zu sein (Criblez, 2013; Rothenbühler, 2010).

### **2.2.3 Staat, Schule und Statistik**

Die enge Verbindung zwischen der Verstaatlichung der Erziehung und der Entstehung der modernen Sozialstatistik, wie sie sich seit Beginn des 19. Jahrhunderts zeigt, ist kein Zufall, sondern sie gehören zum gleichen Prozess: der Entstehung des modernen (National-)Staates. Der moderne Staat unterscheidet sich vom vorangehenden Ständestaat, dass auf seinem Gebiet nominell gleichberechtigte Bürger leben. Auch die Kantone der alten Eidgenossenschaft wiesen ständische Züge auf. Träger des vormodernen Staates waren die patrizischen Oberschichten in Städten wie Bern, Zürich oder Genf sowie in den Landkantonen. Ausserhalb der Hauptorte der Stadtkantone sowie in

den gemeinen Herrschaften lebten Untertanen mit unterschiedlichen Rechtsstatus (Maissen, 2011, S. 137–156).

Die Vorstellung, dass die auf einem Gebiet lebenden Personen durch eine gemeinsame Staatsbürgerschaft verbunden seien, setzte sich erst Ende im 19. Jahrhundert. Die Stiftung dieser «vorgestellten Gemeinschaften» bedurfte intensiver intellektueller Arbeit, etwa durch das Schreiben einer Nationalgeschichte, der Vereinheitlichung der lokalen Dialekte zur einer Hochsprache oder durch eine Kartografierung der Nation, die eine Imaginierung des Landes als zusammengehörigen Raum ermöglichte (Anderson, 1991). Die Statistik, die in ihren älteren Formen einer Landesbeschreibung viele Überschneidungen zur Geografie und zur Geschichte aufwies, gehörte mit zu denjenigen Wissenschaften, die den Nationalstaat erfanden. Sie konzipierten ihn als Container, sprich als abgeschlossenes Objekt, das durch die «Einheit von Staatsgewalt, Staatsvolk und Staatsgebiet» definiert wurde (Woolf, 1989).

Der Schule kam in diesem Prozess eine wichtige Rolle der Popularisierung zu. Nicht nur vermittelte sie den Schüler\*innen durch die neuen Fächer wie Geschichte und Landeskunde, dass sie Teil eines Staatsvolks waren, die allgemeine Schulpflicht betonte auch die rechtliche Gleichstellung der Staatsbürger (Criblez und Hofstetter, 1998; Dahn und Boser, 2015). Die Verbindung zwischen Schule und Statistik ist noch enger. So wurde die Schulbildung selbst zum Objekt der statistischen Forschung. Der Zustand der Schulen und mehr noch die Frage nach den Verhältnissen der Schüler\*innen, die Regelmässigkeit des Schulbesuchs, ihre Gesundheit, ihr Schulerfolg, werden vermessen und auf eindeutige Zahlen gebracht. Diese werden als Chiffren für dahinterstehende gesellschaftliche Zustände gelesen, welche staatliche Interventionen legitimieren, sei es zur Volksgesundheit, zur Verbesserung der Lehrerausbildung oder zur Bekämpfung von Armut und Alkoholismus in den Elternhäusern. Für die Schweiz ist dies bisher für die Zeit am Ende des 19. Jahrhunderts dokumentiert (Ruoss, 2018), für einzelne Kantone auch früher. Diese Arbeit möchte versuchen, Hinweise auf die Erfindung der Schweizer Gesellschaft anhand der Schulstatistiken bereits für einen früheren Zeitraum zu finden.

## 2.3 Zwischenfazit

Vor dem Hintergrund des historischen Forschungsstands verspricht diese Arbeit in zwei Punkten einen neuen Beitrag: Zum einen ist die Geschichte der Schweizer Statistik im Zeitraum vor Gründung des Bundesstaats 1848 noch wenig erforscht. Anhand des Diskurses um die Volksschule kann untersucht werden, welchen Einfluss Statistik auf die Diskussion gesellschaftlicher Probleme hatte. Zum anderen ist die hier angewendete Methodik einer linguistisch informierten Korpusanalyse für die Geschichtswissenschaften ungewöhnlich. Im folgenden Kapitel wird deshalb die Analysemethode theoretisch begründet.

# 3 Methodik: Historische Diskursanalyse mit korpuslinguistischen Methoden

## 3.1 Korpuslinguistik

Die Korpuslinguistik versteht sich als empirische Disziplin der Linguistik, die Sprache anhand einer grösseren Menge an beobachteten und gesammelten Sprachdaten untersucht. In Saussurschen Termini ausgedrückt fokussiert sie auf *langue*, Sprachgebrauch, statt *parole*, das Sprachsystem (Lemnitzer und Zinsmeister, 2015, S. 30–31).

Die Annahme, dass sich Sprache nicht anhand theoretischer Überlegungen, sondern primär anhand real beobachteten Sprachgebrauchs analysieren lässt, macht sie zu einer empirisch vorgehenden Disziplin (Lemnitzer und Zinsmeister, 2015, S. 21).

### 3.1.1 Was ist ein Korpus?

Um den Sprachgebrauch analysieren zu können, müssen grosse Mengen von Sprachdaten gesammelt werden. Eine solche Zusammenstellung grösserer Mengen an Text wird Korpus genannt. Korpora werden in mehreren Teildisziplinen der Linguistik sowie auch anderen Wissenschaften, die mit Textdaten arbeiten, verwendet. Linguistische Korpora unterscheiden sich in der Regel dadurch, dass sie versuchen, Sprachgebrauch (sei es allgemeinsprachlich oder die Sprache spezieller Gruppen) repräsentativ zu dokumentieren und dafür grosse Mengen möglichst unterschiedlicher authentischer Sprachdaten zu sammeln. Häufig werden Korpora annotiert, sodass den sprachlichen Rohdaten Wortarten, syntaktische Funktionen, Lemmatisierungen oder andere Anmerkungen hinzugefügt wurden. Da reine Textdaten meist mehrdeutig sind, aber ihr Kontext im Satz häufig eine eindeutige Bestimmung erlaubt, können an einen annotierten Korpus abstraktere Anfragen gestellt werden (Stefanowitsch, 2020, S. 22–23, 56, Lemnitzer und Zinsmeister, 2015, S. 58–59). Eine wichtige Vorbedingung für die Annotierung eines Korpus ist seine Zerlegung in Worteinheiten. Dies wird als Tokenisierung bezeichnet und nicht nur Wörter, sondern auch Zahlen, Satzzeichen und andre Symbole annotiert. Dafür wird häufig davon ausgegangen, dass eine von Leerzeichen umgebene Zeichenfolge eine Einheit darstellt, die als *Token* bezeichnet wird. Nicht jedes Token ist also ein Wort im herkömmlichen Verständnis (Lemnitzer und Zinsmeister, 2015, S. 61–62).

Erste Korpora wurden bereits vor dem Aufkommen elektronischer Datenverarbeitungssysteme erstellt. Aufgrund des hohen Aufwands einer manuellen Auswertung grösserer Textmengen blieben sie auf wenige Gebiete beschränkt, wie etwa die Konkordanzen religiöser Texte. Erste elektronische Korpora entstanden seit den 1960er-Jahren (Tognini-Bonelli, 2001, S. 52).

### 3.1.2 «Corpus-based» und «corpus-driven»

Aufgrund des hohen Aufwands, der für eine manuelle Erstellung und Auswertung eines Korpus nötig wäre, wurden die ersten Korpora zunächst meist als Belegsammlungen und zur Bestätigung von aus der Theorie abgeleiteten Hypothesen eingesetzt. Dieser deduktive Ansatz wird heute auch als «corpus-based» oder «korpusgestützt» bezeichnet.

Die wachsende Leistungsfähigkeit und leichtere Verfügbarkeit von Computern machte aber seit den 1990er-Jahren die maschinelle Auswertung elektronischer Korpora einfacher. In der Folge wurde ein induktives Vorgehen bei der Auswertung der Korpora möglich. Dieser «corpus-driven» oder «korpusbasierte» Ansatz versucht nicht, aus der Theorie formulierte Vermutungen mit Korpusdaten zu bestätigen, sondern umgekehrt, die Analysekategorien, Theorien und verallgemeinerbare Aussagen aus den Korpusdaten selbst abzuleiten (Tognini-Bonelli, 2001, S. 65).

Als Teilgebiet der Linguistik sind die Methoden der Korpuslinguistik zunächst entwickelt worden, um Fragen zu beantworten, die sich auf Sprache beziehen. Kritisch für das hier verfolgte Interesse ist, dass korpuslinguistische Analysen sprachliche Phänomene rein auf statistischer Grundlage untersuchen. Was sie liefern sind Zahlen und Aggregate. Fundstellen und Belege sind aus ihrem sprachlichen und para-sprachlichen Kontext entrissen und werden statistisch untersucht, nicht aufgrund inhaltlicher oder semantischer Zusammenhänge (Scharloth, 2018, S. 61–62). Um nachvollziehen zu können, wie Korpora auch Fragen zu sozialem Handeln, Kultur und Gesellschaft beantworten können, also auch die semantische Dimension von Sprache erfassen können, müssen wir uns dem Konzept des Diskurses zuwenden.

## 3.2 Diskurs- und Korpuslinguistik

### 3.2.1 Was ist ein Diskurs?

Der Begriff Diskurs wird je nach theoretischem Zugang und Erkenntnisinteresse unterschiedlich definiert. Als einflussreich hat sich Michel Foucaults Verständnis von Diskursen erwiesen. Für Foucault sind Diskurse charakterisiert als regulierte Systeme, die bestimmte Aussagen über ein Thema überhaupt möglich machen. Diskurse sind Netze von aufeinander bezogenen Aussagen und Handlungen. Damit sind sie wesentlich für die Produktion gesellschaftlich geteilter Wahrheiten und überhaupt Erkenntnismöglichkeiten. Gesellschaftliche Prozesse, regelrechte Sagbarkeitsregime, regulieren, wer innerhalb eines Diskurses etwas sagen und was gesagt werden kann. Sie sind zentral für die Produktion einer gesellschaftlich geteilten Wirklichkeit, die durch die Diskurse immer sprachlich vermittelt wird. Es gibt keine Möglichkeit, eine «Wahrheit» hinter den Diskursen zu erfassen, denn jedes Verständnis der Welt, das wir uns machen, wird bereits durch Sinnproduktionen und Diskurse geprägt (Foucault, 2021).

Foucaults Ansatz der Untersuchung von Diskursen zielte darauf zu untersuchen, welche Instanzen die Regelmässigkeit des Diskurses sicherstellten, indem sie Zugang und Ausschluss zu Diskursen regelten. Sein Ansatz ist anschlussfähig an wissenssoziologische Konzepte, die der Sprache eine

zentrale Rolle bei der «sozialen Konstruktion der Wirklichkeit» zuwiesen. Da in Gemeinschaften, die über die *face to face*-Community hinausgehen, Wirklichkeit immer sozial konstruiert und zuvor-derst sprachlich vermittelt wird (Berger et al., 2018), ist die sprachliche Wirklichkeit auch die einzige zugängliche Wirklichkeit.

Die erkenntnistheoretischen Folgen eines solchen Verständnisses sind, dass der Sprache eine zentrale Rolle als Analyseobjekt zukommt, um die Wirklichkeit zu beschreiben. Sprechakte und Sprechweisen erfolgen regelmässig. Eine Analyse der innerhalb eines Diskurses vorkommenden sprachlichen Muster könnte daher Aufschluss darüber geben, welche Regeln einen Diskurs beherrschten und was damit den Diskursteilnehmer\*innen als sag- und denkbar und damit überhaupt als «wahr» vorkommt (Bubenhofer, 2009, S. 32).

### **3.2.2 Korpuslinguistische Untersuchung von Diskursen**

Foucault entwickelte keine explizite Methode einer Diskursanalyse, seine Ideen wurden aber in vielen Disziplinen rezipiert. So haben sowohl die Linguistik als auch die Geschichtswissenschaft Methoden einer Diskursanalyse entwickelt. Da Diskurse als zusammenhängende Netze sprachlicher Äusserungen verstanden werden, also als eine grosse Menge an sprachlichen Äusserungen, bietet sich eine korpuslinguistische Analyse von Diskursen an. Wenn wir davon ausgehen, dass die in einem Korpus versammelten Sprachdaten einen diskursiven Zusammenhang aufweisen und Diskurse «wirklichkeitsstiftend» sind, dann erlaubt eine Untersuchung dieses Korpus Rückschlüsse auf die Prozesse, Möglichkeiten und Grenzen der sozialen Konstruktion der Wirklichkeit. Wir gehen also davon aus, dass die sprachliche Oberfläche, die wir in einem Diskurs vorfinden, nicht zufällig so entstanden ist, sondern dass ihre Produktion einer Regelmässigkeit unterliegt.

Signifikant häufig auftretende sprachliche Muster können deshalb als das Ergebnis rekurren-ter Sprachhandlungen der Sprecherinnen und Sprecher gedeutet werden, in die typische Ver-wendungskontexte, Handlungsziele und Interpretationsrahmen eingeschrieben sind (Scharloth, 2018, S. 63).

### **3.2.3 Konzepte der Diskursanalyse in der Geschichtswissenschaft**

Die Geschichtsschreibung ist eine der paradigmatischen Geisteswissenschaften. Ihre Verwissen-schaftlichung im 19. Jahrhundert hängt eng mit der Entwicklung der historisch-kritischen Methode der Quelleninterpretation zusammen. Die Hermeneutik prägt das Fach stark und im Fokus steht oft immer noch das Verstehen und die umfassende Rekonstruktion des Einzelfalls ohne das Ziel, dar-aus induktive Schlüsse zu ziehen (Dilthey, 1992).

Quantitative Ansätze haben in der Geschichtswissenschaft daher unterschiedliche Konjunkturen erfahren. Besonders die Sozialgeschichte, die im deutschsprachigen Raum vor allem mit der soge-nannten Bielefelder Schule verbunden wurde, betonte gegenüber der traditionellen Historiografie die Bedeutung von sozialen und wirtschaftlichen Strukturen gegenüber individuellem Handeln. Ge-sellschaftliche Makroprozesse sollten nicht zuletzt durch quantitative Analysen erklärt werden. Auch

international erlebte ab etwa 1960 eine deutlich quantitativ ausgerichtete Wirtschafts- und Sozialgeschichte (*Cliometrics*) eine starke Verbreitung. Der *cultural turn*, der in den 1980er-Jahren einsetzte, forderte die Sozialgeschichte jedoch heraus. Das Interesse der «neuen Kulturgeschichte» an Alltags- und Mikrogeschichte, historischer Anthropologie und ihre konstruktivistische Erkenntnistheorie, die nach den Sinngebungsprozessen und Deutungen historischer Akteure fragte und sich methodologisch eher an der Ethnologie als Leitwissenschaft orientierte, wurde als Gegensatz zur Sozialgeschichte gedeutet, die sich als Teil der Sozialwissenschaften verstand (Nathaus, 2012).

Seitdem nehmen quantitative Ansätze eine eher periphere Rolle in der Geschichtswissenschaft ein, wenn man von der Wirtschaftsgeschichte absieht. Die starke Tradition der Hermeneutik im Fach, die mit der historisch-kritischen Methode das verstehende Lesen und sorgfältige Rekontextualisierung von Einzelquellen als Kernmethode des Fachs verteidigte und die Dominanz der Kulturgeschichte mit einer konstruktivistischen Epistemologie seit den 1990er-Jahren liessen und lassen viele Historiker\*innen mit Skepsis auf quantitative Ansätze blicken (Hohls, 2018, S. A.1-21).

### **3.2.4 Begriffsgeschichte und Historische Semantik**

Entsprechend der Bedeutung schriftlicher Quellen für die Geschichtswissenschaft, gibt es eine lange Tradition an Ansätzen, die Schnittmengen mit linguistischen Fragestellungen haben. Im deutschsprachigen Raum ist zunächst an die Begriffsgeschichte und die historische Semantik zu denken, die anhand von Schlüsselbegriffen historischen Wandel und Bedeutungsverschiebungen von Begriffen untersuchte (Koselleck, 1972; Koselleck et al., 1972). Mit ihrem Interesse an einzelnen Wörtern in ihren konkreten Verwendungszusammenhängen und dem Versuch, den semantischen Gehalt des Begriffs in den jeweiligen Belegstellen zu identifizieren und daraus wiederum auf sozial- und kulturhistorischen Wandel, etwa Mentalitätswandel oder sozialstrukturelle Umbrüche und Kontinuitäten zu schliessen, ist die Begriffsgeschichte sehr nahe an dem hier verfolgten linguistischen Vorhaben. Angesichts der Nähe des begriffshistorischen Zugriffs zur Korpus- und Diskurslinguistik erstaunt, dass der interdisziplinäre Austausch zwischen den beiden Disziplinen nur zögerlich in Gang kam (Busse et al., 2000).

Ein wichtiger Unterschied liegt jedoch in der Aufmerksamkeit für die sprachliche Oberfläche. Die Begriffsgeschichte ist meist nicht an Wörtern interessiert, sondern an Konzepten und Institutionen, die mit den Begriffen ausgedrückt werden. Aus linguistischer Sicht ist die Auswahl und das Verständnis der Untersuchungsgegenstände der Begriffsgeschichte problematisch. Dietrich Busse hat die linguistische Kritik an Programm und Umsetzung der Begriffsgeschichte in Gestalt der «Geschichtlichen Grundbegriffe» pointiert zusammengefasst. Ihr Begriffsbeispiel bleibt unbestimmt und kann sich nicht von der impliziten Annahme lösen, dass die Begriffe doch auf dahinterliegende Dinge verweisen, die das eigentliche Forschungsobjekt bilden. Das führt zu einer theoretisch nur schwer begründbaren Unterscheidung zwischen untersuchungsrelevanten Begriffen und einem grossen Rest an blossen Wörtern und Äusserungen. Die reale begriffsgeschichtliche Praxis präsentiert sich weniger als semantische Analyse, sondern meist als herkömmliche Ideengeschichte, die sich auf normative Quellen zur Erklärung des jeweiligen Begriffs stützt (Busse, 1987, S. 50–71).

Wichtig für diese Arbeit ist Busses Hinweis, dass eine historische Semantik die Verwendungskontexte miteinbeziehen muss, in der die Begriffe historisch ihren Sinn entfaltet haben. Er verweist dazu auf die Diskurskonzeption Michel Foucaults, die sich im gleichen Zeitraum auch in den Geschichtswissenschaften etablierte.

### **3.2.5 Historische Diskursanalyse**

Das Werk Michel Foucaults wurde in den 1970ern, angesichts der sozialhistorischen Dominanz, noch zögerlich rezipiert. Seit den 1990er-Jahren ist sein Werk aber von Historiker\*innen auf sehr verschiedene Weisen als theoretische Orientierung verwendet worden. Foucaults Diskursarchäologie hat dabei viele Arbeiten angeregt, die versuchten, soziale Wirklichkeit als historische Diskurse zu rekonstruieren, indem die wechselseitige Konstitution von Diskursen und Diskursteilnehmer\*innen in den Blick genommen wurde (Landwehr, 2018).

Während es der Begriffsgeschichte darum geht, den semantischen Gehalt der Begriffe zu verstehen, vor allem auch in seinem Wandel über die Zeit hinweg, zielt die Diskursanalyse im Gefolge Foucaults weniger auf den Sinn eines Wortes, sondern auf die Position einer Aussage in einem Diskurssystem: Welche diskursiven Formationen machten es überhaupt möglich, eine Aussage so zu treffen und sie als sinnvoll verstanden zu wissen? In welchem Zusammenhang mit dem «Vorher» und «Nachher» des Diskurses steht dieser Ausdruck (Sarasin, 2012, S. 159–160)?

### **3.2.6 Digital Humanities und Interesse für quantitative Methoden**

Die seit den 2000er-Jahren stetig wachsenden Digital Humanities, computergestützten Geisteswissenschaften, hat auch die Geschichtswissenschaft erfasst. Mit zunehmender Leistungsfähigkeit der Computer und der Entwicklung benutzerfreundlicher Software wird auch in der hermeneutisch geprägten Geschichtswissenschaft die maschinelle Auswertung von Quellen wieder vermehrt diskutiert. Neben den klassischen seriellen Quellen der Sozialgeschichte werden auch Analysemethoden zur Auswertung grosser Textmengen rezipiert (z. B. «Topic Modelling», Graham et al., 2012). Gerade niederschwellige Tools wie Googles *n-gram*-Viewer wecken bei Historiker\*innen die Faszination, wie alte Fragestellungen neu bearbeitet werden können. Auch hier liegt das Interesse aber mehr in der Anwendung als in einem Verständnis der linguistischen Hintergründe (zumindest kommt die Diskussion bei Sarasin, 2012 ohne einen Verweis auf die linguistische Diskussion aus). Erste Versuche, neuere korpuslinguistische Methoden für die Begriffsgeschichte fruchtbar zu machen, wurden indes mittlerweile unternommen (Kämper, 2018; Schwandt, 2018). Dass schon die aus korpuslinguistischer Sicht einfache Methoden von Historiker\*innen als anregend innovativ empfunden werden und neue Fragen und Einsichten ermöglichen, zeigen Burckhardt u. a., die ausgewählte Begriffe in der DDR-Presse diachron verfolgen (Burckhardt et al., 2019).

### 3.3 Korpuslinguistik für kulturanalytische Fragestellungen

Die Voraussetzungen für eine korpuslinguistisch inspirierte Untersuchung historischer Diskurse sind also gegeben. Im Folgenden werden kurz die wichtigsten theoretischen Grundannahmen skizziert. Ausgegangen wird davon, dass gesellschaftliche Wirklichkeit sprachlich konstituiert ist. Jede Handlung kann nur durch sprachlich geleitete Konstruktion von Sinn ein *soziales Handeln* werden. Dadurch, dass diese Sprachhandlungen aufeinander bezogen sind, entsteht ein Diskurs, der die geteilten Konstruktionen von Sinn ermöglicht. Jede sprachliche Äusserung ist eine Handlung im Diskurs. Sie schliesst an vorhergehende Äusserungen und Sinnkonstruktionen an und beeinflusst folgende Sinnkonstruktionen und damit zukünftiges Handeln. Die Analyse der sprachlichen Äusserungen innerhalb eines Diskurses erlaubt daher, von der sprachlichen Oberfläche auf die soziale Wirklichkeit zu schliessen (Lemke und Stulpe, 2015, S. 63).

#### 3.3.1 Modellierung des Diskurses in einem Korpus

Entscheidend für die Aussagekraft einer korpuslinguistischen Untersuchung ist die Zusammenstellung des Korpus. Linguistische Korpora versuchen eine grösstmögliche Repräsentativität herzustellen, das heisst den Sprachgebrauch umfassend abzubilden. In der Praxis ist eine absolute Repräsentativität nicht erreichbar, weshalb in der Regel eine möglichst grosse Diversität an Sprachdaten und Quellen im Korpus angestrebt wird (Stefanowitsch, 2020, S. 34–35). Bei begriffsgeschichtlichen und diskursanalytischen Korpora stellt sich noch dringender die Frage nach der Repräsentativität der Korpusauswahl. Angesichts der begrenzten Anzahl an Texten, die mit einem qualitativen Ansatz untersucht werden können, stellt sich die Frage, ob die jeweils analysierten Texte genügend typisch sind, um Aussagen über die Struktur des Diskurses machen zu können (Landwehr, 2018, S. 101–103).

Allgemeinsprachliche Korpora bieten sich auch für die Untersuchung von Spezialthemen an. Wenn ein Diskurs tatsächlich wirkmächtig ist, sollten seine Sprechmuster und Sagbarkeitsregimes auch in einem allgemeinen Sprachgebrauch Niederschlag finden. Für Fragen an Spezial- oder Gegendiskurse kann es aber auch erkenntnisfördernd sein, einen möglichst spezifischen Korpus zu kompilieren. So lassen sich etwa einfacher Erkenntnisse gewinnen, die im Rauschen der anderen Diskurse untergehen würden (Hodel, 2013).

Die Frage der Repräsentativität stellt sich bei zu kulturanalytischen Zwecken zusammengestellten Korpora in verschärftem Masse. Die Auswahl der Texte des Korpus, die als zu einem Diskurs gehörig verstandenen werden, unterliegt der Fragestellung und damit eben der Annahme, dass sie in einem historisch-semantischen Zusammenhang stehen. Das ist unvermeidbar, aber es unterwirft die Texte einem Vorverständnis (Busse und Teubert, 1994, S. 15–16). Eine gänzlich «naiv»-induktive Diskursanalyse, bei der die Daten «für sich selber sprechen» ist daher theoretisch unmöglich. Diskurse, ob gegenwärtige oder historische, existieren nicht unabhängig für sich als beobachtbare Objekte, sondern sie entstehen, weil Forscherinnen annehmen, dass eine nach gewissen Kriterien bestimmbare Menge an Texten in einem diskursiven Zusammenhang stehen. Jedes Korpus, das für

eine Diskursanalyse modelliert wird, ist abhängig von der Fragestellung, dem Erkenntnisinteresse und den zur Verfügung stehenden Daten und Auswertungsmöglichkeiten (Dreesen und Stücheli-Herlach, 2019, S. 143–144).

### 3.3.2 Operationalisierung: Frequenzanalyse und Kollokationen

Als hauptsächliche Analysemethoden werden in dieser Arbeit Frequenzanalysen und Kollokationsprofile angewendet. Frequenzanalysen sind eine recht einfache Auswertungsmöglichkeit. Sie zählen, wie häufig ein Token innerhalb des Korpus auftaucht. Das Verfahren ist unter Historiker\*innen seit Googles *n-Gram Viewer* vertraut. Dieser ermöglicht eine Frequenzanalyse von Begriffen in den von Google Books gescannten Büchern ermöglicht (Sarasin, 2012). Die Häufigkeiten können absolut oder relativ per 1 Million Wörter ausgegeben werden. Kategorisiert man die Treffer nach Erscheinungsjahr des Textes ergeben sich schnell Einblicke in die Konjunkturen des Suchbegriffs: Wird er im Laufe der Zeit mehr oder weniger verwendet? Gibt es auffällige Peaks? In diesem Fall ist es meist sinnvoller, die relative Häufigkeit der Token zu betrachten, um über die Jahre hinweg ungleich verteilte Textmengen auszugleichen.

Werden bei Frequenzanalysen einzelne Token gezählt, so wird bei der Kollokationsanalyse das Umfeld eines Suchbegriffs in den Blick genommen. Bubenhofer summiert das verbreitete «empirische Verständnis» von Kollokationen als «Paare von Worteinheiten [...], die innerhalb einer bestimmten Distanz zueinander kookkurrieren und eine statistisch feststellbare Bindung zueinander aufweisen.» Das Mass für die Bindung wird meist statistisch bemessen, nämlich danach, ob gewisse Wortpaare häufiger in einem Korpus auftauchen, als aufgrund einer zufälligen Verteilung statistisch erwartbar ist. Die Teile einer Kollokation müssen nicht unbedingt in direkter Nachbarschaft stehen, sondern die Suche kann auch auf Token im Umfeld ausgedehnt werden. Theoretisch offen ist die Definition auch für Einheiten aus mehr als zwei Wörtern, n-Gramme, deren Assoziation aber statistisch schwieriger zu konzeptualisieren ist (Bubenhofer, 2017, S. 69–70).

Dem Einsatz von Kollokationen in der kulturanalytischen Korpuslinguistik liegt die Annahme der linguistischen Pragmatik zugrunde, dass der semantische Gehalt von Wörtern von ihren Verwendungskontexten bestimmt wird (Bubenhofer, 2009, S. 44–46). Sie sind die sprachlichen Muster, in denen sich soziale Phänomene ausdrücken oder diese (mit-)konstituieren. Vom überzufällig häufigen gemeinsamen Auftritt von Wörtern kann auf Bedeutzungszuschreibungen und Kontexte geschlossen werden (Scharloth, 2018, S. 63).

# 4 Explorative Annäherung

## 4.1 Das Korpus

Die Datengrundlage für die Analyse des Diskurses über die Volksschule in der Deutschschweiz ist ein Korpus aus retrodigitalisierten Drucken von den Plattformen e-rara.ch und e-periodica.ch. E-rara.ch versammelt Digitalisate von Drucken und Einzelauflagen aus Schweizer Institutionen, während auf e-periodica.ch Schweizer Zeitschriften zu finden sind.<sup>1</sup>

Im Korpus wurden die Quellen der Plattformen versammelt, die zwischen dem 1. Januar 1800 und dem 31. Dezember 1870 publiziert wurden und deren Volltexte mindestens eine der Zeichenfolgen *schulbildung*, *schulwesen*, *kantonalschul*, *volksschul* oder *primarschul* enthalten.

Die Digitalisate wurden vonseiten der Plattformen mit einer Texterkennung (OCR) versehen. Durch die unterschiedliche Qualität der Digitalisate können nicht immer alle Worte im Text zuverlässig maschinenlesbar gemacht werden. Das hat zum einen zur Folge, dass die Anzahl und Frequenz der Suchtreffer nicht genau ist. Ausserdem können falsch erkannte Worte zu Fehlern bei der Lemmatisierung oder Tokenisierung führen, wie bei den ersten explorativen Beispielen gezeigt wird. Zuletzt bedeutet eine nicht-zuverlässige Erkennung auch, dass nicht garantiert werden kann, dass ein Begriff, der nicht per Abfrage gefunden wurde, nicht dort im Korpus auftaucht.

Die heruntergeladenen rohen OCR-Daten wurden anschliessend tokenisiert, das heisst in einzelne Worteinheiten und andere Einheiten zerlegt und linguistisch annotiert. Bei der Annotation wurden den einzelnen Tokens maschinell weitere Informationen hinzugefügt. Hier wurden die Lemmata der Wörter, also die Grundform des jeweiligen Wortes, ergänzt sowie Informationen zur Wortart hinzugefügt (Lemnitzer und Zinsmeister, 2015, S. 57–87). Die Wortformen (z. B. finites Verb, Eigennamen, etc.) werden als «Part of speech» (POS) und können entsprechend gesucht werden. Daneben werden auch Interpunktionszeichen oder Token, die nur aus Zahlen bestehen, als eigenständige POS attribuiert. Beispielsweise zeigt [word="Tabelle"] [pos="VFIN"] alle Treffer, wo das Token *Tabelle* von einem finiten Vollverb gefolgt wird. Verwendet wird das Stuttgart-Tübingen Tagset (STTS), das sich im deutschen Sprachraum als Standard etabliert hat (Schiller et al., 1999). Das erlaubt die Suche nach komplexeren Mustern im Korpus. Angesichts der sehr schwankenden OCR-Qualität der im vorliegenden Korpus enthaltenen Texte dürfte die Lemmatisierung und Annotation in vielen Fällen nicht erfolgreich gewesen sein. Da bei einer Konzentration auf die annotierten Daten semantische Informationen verloren gehen können (Bubenhofer, 2009, S. 124–127), werde ich öfter nach einzelnen Wortformen und nicht Lemmata suchen. Das heisst, dass eine Abfrage wie [word="Schülerinnen"] nur Token gefunden werden, die genau diese Zeichenfolge enthalten,

---

<sup>1</sup> Das Korpus wurde von Philipp Dreesen zusammengestellt, aufbereitet und annotiert zur Verfügung gestellt.

während eine Suche wie [lemma="Schülerinnen"] sowohl das Token *Schülerin* wie auch *Schülerinnen* findet.

Insgesamt enthält das Korpus 13'306 unterschiedliche Texte mit insgesamt 74'096'351 Token. Eine Einteilung des Korpus in Subkorpora per Dekade zeigt eine gleichmässige Verteilung der Wortmenge im Zeitraum 1830 bis 1870. Auffällig ist die sehr geringe Menge an Texten und Token für den Zeitraum 1810 bis 1829. Ebenfalls auffällig ist, dass die Tokenmenge pro Text sich sehr verändert. Beispielsweise kommen 1800–1809 4.4 Millionen Token auf 1'189 Texte, also 3'735 Token pro Text, während 1830–1839 ein Text durchschnittlich 8'306 Token enthält (Tabelle 4.1).

Subkorpus	Grösse	Anzahl Token
1800–1809	1'189 Texte	4'440'324
1810–1819	84 Texte	770'109
1820–1829	304 Texte	3'976'613
1830–1839	1'323 Texte	10'988'279
1840–1849	1'564 Texte	28'525'988
1850–1859	4'429 Texte	17'088'456
1860–1869	4'151 Texte	16'937'411
1870	262 Texte	2'013'970

TABELLE 4.1: Grösse der Subkorpora nach Dekaden

Den Ansätzen der korpusgestützten Linguistik folgend wird das Korpus induktiv untersucht, das heisst Hypothesen und Kategorien für die weitere Analyse werden aus dem Korpus gewonnen. Zu beachten ist aber, dass auch solche induktive Analyse nicht ohne Vorannahmen auskommt. Diese betreffen einerseits erkenntnistheoretische Grundannahmen wie auch die Zusammenstellung des Korpus, als auch Vermutungen zu Aussagemöglichkeiten des Korpus und die verwendeten Methoden.

Das Korpus wird mit der Software CWP/CQP bzw. dessen webbasiertem Suchoberfläche CQPWeb analysiert. CQPWeb ermöglicht gängige statistische Untersuchungen von Korpora wie Kollokations- und Häufigkeitsanalysen und, sofern das Korpus entsprechend annotiert wurde, eine Untersuchung nach Lemmata und Wortarten (Hardie, 2021).<sup>2</sup>

## 4.2 Entwicklung von Begriffen im Zusammenhang mit Statistik

### 4.2.1 Diachronische Keywords

Ein gängiges Verfahren, um einen ersten Überblick über ein grosses Korpus und typische Wörter zu bekommen, ist die Keyword-Analyse. Hier wird das Korpus mit einem Referenzkorpus verglichen, um signifikante Unterschiede festzustellen. Wörter, die im Untersuchungskorpus deutlich frequenter sind als im Referenzkorpus, können als typisch für das Korpus gelten (Bubenhofer, 2017, S. 74). Im

<sup>2</sup> Ich danke Klaus Rothenhäuser für die technische Unterstützung bei der Verwendung von CQPWeb.

vorliegenden Fall steht jedoch kein Referenzkorpus zur Verfügung, das einen allgemeinen Sprachgebrauch des 19. Jahrhunderts abbilden würde und mit dem Untersuchungskorpus verglichen werden könnte. Stattdessen wurde das Korpus in zwei Subkorpora unterteilt, die beide je die Hälfte des Untersuchungszeitraumes umfassen. Das erste Korpus enthält Texte aus dem Zeitraum 1800 bis 1834, das andere Texte mit Veröffentlichungsdatum in den Jahren von 1835 bis 1870. Ein Vergleich der beiden Subkorpora könnte erste Einblicke in signifikante Unterschiede zwischen den beiden Korpushälften geben (für andere Ansätze im Umgang mit einem historischen Korpus, siehe Bubenhofer und Scharloth, 2014).

Tabelle 4.2<sup>3</sup> zeigt die 50 Begriffe des Subkorpus mit den Texten von 1835 bis 1870, die im Vergleich zu den Jahren 1800 bis 1834 von CQPWeb nach dem «Log Ratio»-Wert als Keywords berechnet wurden.<sup>4</sup> Log Ratio ist ein geeigneter Masswert, um zu berechnen, wie häufig ein Token in einem Korpus gegenüber einem anderen Korpus auftritt (Hardie, 2014). Mass für die Bewertung eines Keywords ist nicht die absolute Häufigkeit eines Tokens im Subkorpus, sondern sein statistisch signifikantes, überzufällig häufiges Vorkommen in dem Korpus im Vergleich zu einem Referenzkorpus (Gabrielatos, 2018). Das bedeutet, je höher hier der Log Ratio-Wert, desto stärker ist ein Begriff im Subkorpus überrepräsentiert. Aus strenger statistischer und korpuslinguistischer Sicht genügt die Analyse nicht den Massstäben für Signifikanz, aber ermöglicht dennoch eine erste Annäherung, um Hinweise für mögliche weitere Untersuchungsschritte zu gewinnen.

Position	Token	Häufigkeit	Häufigkeit	Log ratio
		pro Mio.	pro Mio.	
		Wörter	Wörter	
		1835–1870	1800–1834	
1	wät	22.17	0.22	6.63
2	Pfr	113.41	1.19	6.57
3	Soeben	25.57	0.3	6.42
4	gum	24.76	0.3	6.37
5	unb	621.21	8.36	6.22
6	bie	443.25	6.34	6.13
7	Jnterlaken	25.85	0.37	6.11
8	Ba«	24.91	0.37	6.06
9	Kt.	45.26	0.75	5.92
10	Bem	137.68	2.54	5.76
11	»or	11.93	0.22	5.74
12	Sekundar-	11.86	0.22	5.73
13	Staatsbeitrag	15.34	0.3	5.68
14	bth	14.61	0.3	5.61
15	CtS	44.79	0.97	5.53
16	Bundesversammlung	13.71	0.3	5.52
17	Behaart	16.64	0.37	5.48
18	Referat	13.16	0.3	5.46

Fortsetzung auf nächster Seite

3 Die vollständigen Ergebnisse der Abfragen dieser Arbeit stehen unter [https://github.com/mac-fvb/ZHAW\\_thesis-applied-linguistics/tree/master/Data](https://github.com/mac-fvb/ZHAW_thesis-applied-linguistics/tree/master/Data) zur Verfügung.

4 Token, die nur aus Zahlen bestehen, wurden aus den von CQPWeb erzeugten Ergebnissen herausgefiltert.

Tabelle 4.2 – Fortsetzung von vorheriger Seite

Position	Token	Häufigkeit	Häufigkeit	Log ratio
		pro Mio.	pro Mio.	
		Wörter	Wörter	
		1835–1870	1800–1834	
19	fid	12.9	0.3	5.43
20	amtliches	18.76	0.45	5.39
21	turnen	24.6	0.6	5.36
22	Std	18.04	0.45	5.33
23	aud	20.74	0.52	5.31
24	mln	8.17	0.22	5.19
25	Volksschule	98.02	2.69	5.19
26	Anmeldungen	8.12	0.22	5.18
27	3eit	8.04	0.22	5.17
28	bte	40.12	1.12	5.16
29	»on	162.72	4.63	5.14
30	resp.	10.48	0.3	5.13
31	lehrerstand	7.81	0.22	5.12
32	Durchschnittlich	31.1	0.9	5.12
33	lehrerbildung	7.73	0.22	5.11
34	Vorlagen	10.3	0.3	5.11
35	Msgr	7.71	0.22	5.11
36	Seminardirektor	20.36	0.6	5.09
37	Balb	7.53	0.22	5.07
38	nad	22.42	0.67	5.06
39	Gefleckt	7.46	0.22	5.06
40	Sekundarschulen	29.42	0.9	5.04
41	Kantonstheil	7.23	0.22	5.01
42	Gymn	7.22	0.22	5.01
43	Primarlehrer	9.6	0.3	5.01
44	beren	7.13	0.22	4.99
45	fût	23.56	0.75	4.98
46	nationalen	7.05	0.22	4.98
47	Narbe	7.03	0.22	4.97
48	laffen	11.48	0.37	4.94
49	Jgfr	6.89	0.22	4.94
50	Eisenbahnen	11.33	0.37	4.92

TABELLE 4.2: Keywords im Surbkörper 1835–1870 im Vergleich zu 1800–1834

Die Tabelle 4.2 zeigt die Schwierigkeit in der korpuslinguistischen Arbeit mit retrodigitalisierten und OCR-erkannten Texten: Zeichen werden nicht richtig erkannt und zusammengehörige Wörter werden nach dem Schriftsatz getrennt, womit sie im Korpus als eigenständige Token erscheinen. Dies hat zur Folge, dass auch die automatisierte Erkennung der Wortformen weniger zuverlässig ist. Die Tabelle listet denn auch einige Fragmente auf.

Unter diesem Vorbehalt zeigt die Übersicht, welche Begriffe in der diachronisch zweiten Hälfte des Gesamtkörpers überzufällig häufig auftauchen. Im Zusammenhang mit statistischen Diskursen ist einzig das Wort «Durchschnitt» einschlägig. In Bezug auf Diskurse um die Schule lassen sich Anzeichen erkennen, dass Themen der Sekundarschulen und der Lehrerbildung ab 1835 häufiger diskutiert wurden. Eine Hypothesenbildung für weitere Anfragen an den Korpus in Bezug auf die

Fragestellung, die Etablierung einer statistischen Selbstbeschreibung der Schweizer Gesellschaft im Diskurs um die Volksschule, findet aber wenig Anknüpfungspunkte.

#### 4.2.2 Wörterbuchansatz

Buchner et al. nutzen einen Wörterbuchansatz um Trends der quantitativen Forschung in der deutschsprachigen Geschichtswissenschaft des 20. Jahrhunderts zu erforschen. Zur Erstellung nutzten sie ein Korpus von deutschsprachigen Statistiklehrbüchern mit Erscheinungsdatum zwischen 1946 und 2013, aus deren Glossaren und Indizes sie 1'081 Lemmata bzw. 2'928 flektierte Wörter mit eindeutigem Statistikbezug identifizierten. Durch dieses Verfahren soll das Wörterbuch möglichst objektiv einen charakteristischen Sprachgebrauch abbilden, wenn es um Statistik geht. Die verwendeten Begriffe erlauben zudem einen Schluss auf das methodische Niveau der im Text besprochenen statistischen Arbeiten (Buchner et al., 2020, S. 596–599).

Aus forschungspraktischen Gründen stütze ich mich auf dieses Wörterbuch,<sup>5</sup> auch wenn ein aus Quellen des 20. Jahrhunderts erstelltes Buch nicht zuverlässig den Sprachgebrauch und die statistischen Praktiken des 19. Jahrhunderts abbilden kann. Für einen ersten Überblick über das vorliegende Korpus und die Frage danach, wie weit statistische Begriffe im Diskurs verankert sind, ist es allerdings brauchbar.

Position	Token	Anzahl
1	Tabelle	2197
2	Durchschnitt	1785
3	Erhebung	1357
4	Tabellen	1322
5	Schätzung	1191
6	Statistik	1129
7	Prozent	961
8	Quotienten	698
9	Logarithmen	575
10	Durchschnitte	567
11	Quotient	499
12	Zählung	474
13	durchschnitten	436
14	Messungen	374
15	Messung	342
16	Exponenten	323
17	Multiplikator	309
18	Schätzungen	264
19	Exponent	238
20	zählungen	200
21	Erhebungen	194
22	Logarithmus	168
23	Ordinate	168
24	Quadratwurzel	162

Fortsetzung auf nächster Seite

<sup>5</sup> Das Wörterbuch wurde freundlicherweise von den Autoren zur Verfügung gestellt. Es ist einsehbar unter [https://github.com/mac-fvb/ZHAW\\_thesis-applied-linguistics/tree/master/Data/Buchner\\_et\\_al\\_Woerterbuch.csv](https://github.com/mac-fvb/ZHAW_thesis-applied-linguistics/tree/master/Data/Buchner_et_al_Woerterbuch.csv).

Tabelle 4.3 – Fortsetzung von vorheriger Seite

Position	Token	Anzahl
25	Prozente	135
26	Index	119
27	Integral	119
28	Prozenten	99
29	Durchschnitts	76
30	graphen	76
31	Multiplikators	68
32	integrale	51
33	Ordinaten	51
34	Summanden	50
35	Fruchtbarkeitsziffer	44

TABELLE 4.3: Treffer für Einträge aus dem Wörterbuch «Quantifizierung» von Buchner et al., 2020, Frequenz &gt; 40

Aus den 2'928 Einträgen im Wörterbuch konnten 104 im Korpus nachgewiesen werden (s. Tabelle 4.3). Bei näherer Betrachtung der häufigen Treffer im Kontext in der sogenannten KWIC-Ansicht (Keywords in Context) zeigen sich erneut die Schwächen der OCR-erkannten Texte, wenn etwa das kleingeschriebene «s» als «f» erkannt wird. Die Treffer zu *durchschnitten* (Position 13 in Tabelle 4.3) entstammen geografischen Beschreibungen oder einem geometrischen Lehrbuch (vgl. Tabelle 4.4).

Position	Kontext vor	Suchbegriff	Kontext nach
1	grünenden können ; und durch Hülfe derselben fünfmal das Flußbett der Lanquart	durchschnitten	, ste iu ihrem Lauf abge- sperrt , « m sie in
2	nordöstlichen Seite des Tiefentobel , ist dasselbe von unregelmäßigen Gängen und Trümmern	durchschnitten	, die im ganzen genom- men gleiche Richtungs- und Fallungslinie mit diesem
3	An derjenigen Stelle des Tiefentobel , wo dieses Erzführende Lager von jenem	durchschnitten	, vom Kegel abge- wikkelt und in einer Ebene ausgebreitet
4	einmal ward dunkle Nacht , die von den Feuerfunken der Eisenbahn bisweilen	durchschnitten	wurde ; ein unwillkührlichcs Still- schweigen bewältigte die ganze Menge , gleich als
5	diese Männer vereinigt . Keine Aufregung war sichtbar , keine wilden Cheers	durchschnitten	die Luft , kein rhetorisches Gepränge unerquicklich wie Wüsten- sand , sondern
6	beiden gleichlaufenden Geraden all und El ) von der Linie E ?	durchschnitten	; so fragt es sich : wie die übereinstimmenden W. , die
7	II. Es seien umgekehrt die Geraden und EO 46 von EE	durchschnitten	, und es sei IN - II , oder u X oder
8	X. 4 ) s. Werden zwei gleichlaufende Gerade von einer dritten Linie	durchschnitten	: so sind die über- einstimmenden W. einander gleich , die innern
9	Zweig da wo das neue einjährige und das alte Holz sich scheidet	durchschnitten	und so eingesetzt . Es bildet sich hart um den Schnitt ein
10	nach oben und unten getrennt wird , mit einem scharfen Messer quer	durchschnitten	und auf eine Länge von zirka 1/ mit 4—5 Augen abgekürzt .

Fortsetzung auf nächster Seite

Tabelle 4.4 – Fortsetzung von vorheriger Seite

Position	Kontext vor	Suchbegriff	Kontext nach
11	der Pyramide , des Ke- gels und Cylin Vers ; dieselben	durchschnitten	und deren Oberfläche auf eine Ebene entwickelt als Netz gezeichnet . Durchdringung
12	seitige ; 3. ein Cylinder , wo mög- lich parallel mit der Grundfläche	durchschnitten	; 4. mehrere Pyramiden , darunter ein Tetraeder und etwa eine parallel
13	der Grundfläche durchschnitte- ne ; 5. ein Kegel , parallel mit der Grundfläche	durchschnitten	; K. eine Kugel , wo möglich zwei- mal , nach einem größten
14	wo möglich zweimal , nach einem größten und einem kleineren Krei- se	durchschnitten	. Wünschenswerth find ferner : Einige Drahtnetze , welche nur die Kanten
15	— auch fände sich der Canton durch einen Distrikt des Cantvns Zürich	durchschnitten	, lind dadurch im Zusammenhang unterbrochen . — Diese geographi- sche Schilderung wird

TABELLE 4.4: KWIC-Ansicht für [word="durchschnitten"%c]

Bei den Treffern für *Durchschnitt* (Position 2) zeigt sich hingegen ein gemischtes Bild: Sie tauchen sowohl in geografischen Kontexten als auch solchen mit Rechnungen und Zahlenwerten auf. Für die Treffer *Zählung* (Position 12) zeigt die KWIC-Ansicht, dass viele Treffer ebenfalls aus falsch erkannten Wörtern stammen, die im Satzspiegel der Quellen umbrochen wurden wie beispielsweise *Er-zählung* (Positionen 7, 12, 14 und 15 in Tabelle 4.5).<sup>6</sup> Darunter sind aber ebenfalls Themen der zeitgenössischen Statistikdiskussion wie *Volkszählung* (Position 11) oder mehrere Zählungen in Bezug auf statistische Erhebungen (Positionen 1–6, 8, 11, 13). Dasselbe trifft auf *schätzung* zu, dass sowohl Treffer aus dem Bereich der Wertung enthält (*Geringschätzung*, *Wert[h]schätzung*, *Unterschätzung*, *Überschätzung*) wie auch einschlägig statistische Wörter wie *Gebäudeschätzung* und *Größe[n]schätzung* enthält.

Position	Kontext vor	Suchbegriff	Kontext nach
1	mit den amtlichen , im Amtsblatte enthaltenen Angaben nicht überein . Die	Zählung	des Viehstandes vom Jahre 1846 ist insoweit unrichtig , weil in jenem
2	vom Jahre 1846 ist insoweit unrich- tig , weil in jenem Jahre keine	Zählung	stattfand . Der Verfasser scheint die im gleichen Jahre in den Verhand- lungen
3	verzeichnet . Heute dürfte Florenz etwa 186,666 Einwohner zählen ( eine offizielle	Zählung	hat seit der Ausdehnung der städ- tischen Grenzen und dem aus 36,666 Köpfen
4	Cantons das immerwährende Woh- nungsrecht besaßen und noch besitzen . Die letzte	Zählung	dcr Landflächen geschah im Jahr 1818 und betrug 2569 Köpfe . Zum
5	das Departement dcs Innern im Laufe des JahrS 1832 eine zweite genaue	Zählung	veranstalten , allein dicscS Bcgchr- en blieb bisher ohne Erfolg , weil ,

Fortsetzung auf nächster Seite

6 Der gleiche Befund gilt auch für *zählungen* (Position 20).

Tabelle 4.5 – Fortsetzung von vorheriger Seite

Position	Kontext vor	Suchbegriff	Kontext nach
6	Bevölk- Eine runq n , d. Wirth , schaue » .	Zählung	schijft aus von l<>z> . Kopfe . Amtsbezirk Aarberg 20 12702 636
7	Auge- denken , als daß wir , durch hier ganz überflüßige Her-	Zählung	dieselben , von Ihrer — nicht der Rüge ftü- herer Mißgriffe ,
8	88 ! c S Activbürger waren in Schams und Rhein- wald , nach	Zählung	vom Jahr 1799. Schams : Zillis ioo .
9	Novelle<- von L. A. Ohorn . Das Gespenst . Et-	Zählung	Andeer ns . Do-1 von Julius Uliczny . — Allerlei , Re bus und Illustrationen. .
10	trugen sie in der verfassungsmäßig anberaumten Zeit gen Liestal . Bei der	Zählung	der Stimmen zeigte sich , daß die Einsprache des Volkes gegen das
11	zur Statistik des Kantons , welches den Schluß der Ergebnisse der Volks-	Zählung	von 1860 und der vergleichenden Zusammenstellungen derselben und der Ergebnisse früherer Volkszählungen
12	jede Woche des lah- res eine kurze Betrachtung mit einer entsprechenden Er-	Zählung	. „Kalender für Zeit und Ewigkeit , von Alb ,
13	Andere muß- ten nach der ersten Probezeit entlassen werden . Eine Auf-	Zählung	der Taubstummen vom Jahre 1839 wies im Kau- ton 571 solcher Unglücklichen
14	77 S. Der Grund und Boden , worin der Inhalt dieser Er-	Zählung	.wurzelt , sind einerseits die Kruz- züge , von denen IV . 23
15	Gebetver- trauen habe sich der orientalischen Phantasie eingekleidet , wie die Er-	Zählung	von Josua lautet , wollten wir nicht unbedingt so nennen . Aber

TABELLE 4.5: KWIC-Ansicht für [word="Zählung"%c]

#### 4.2.2.1 *tabelle, durchschnitt und prozent*

Aus der Überprüfung der weiteren Treffer von Tabelle 4.3 ergibt sich, dass allein die Token *Tabelle*, *Durchschnitt*, *Statistik*, *Prozent*, *Fruchtbarkeitsziffer* und mit Abstrichen *Zählung[en]* eindeutig oder vorwiegend in Bezug auf die Diskussion von Zahlen und Statistik verwendet werden und hinreichend frequent (d. h. mehr als fünf Treffer) auftreten. Diese fünf Token wurden jeweils im Korpus gesucht, um weiteren Aufschluss über ihre Verwendungskontexte zu erhalten. Die 44 Treffer zu *Fruchtbarkeitsziffer* entstammen jedoch alle demselben Text, weshalb sie nicht berücksichtigt werden.

Im Hinblick auf die Fragestellung interessant sind die Treffer zu [word=".\*tabell.\*%c"]. Unter den häufigeren Treffern finden sich die Token *Schultabelle* sowie *Absenz[en]tabelle* und *Versäumnistabelle[n]*. Letztere verweisen auf die Durchsetzung der allgemeinen Schulpflicht, für das die Tabelle als Kontrollinstrument verwendet wurde. Dieser Gedanken wird in Kapitel 5.4 weiterverfolgt.

Die signifikante Häufung des Tokens *Durchschnitt* wurde schon in Tabelle 4.2 als ein möglicher Indikator für ein Aufkommen statistischer Denkmuster in der zweiten Hälfte des Korpus ausgemacht.

Auch in der Wörterbuch-Analyse wurde das Token identifiziert. Da die Kontextansicht darauf hindeutet, dass die Form *durchschnitten* häufig in geografischem Sinn gebraucht wird, wurde die Abfrage mit [word="durchschnitt|durchschnittl.\*%c"] eingegrenzt. So werden nur Treffer gefunden, die das Token *Durchschnitt* enthalten oder ein Token, das mit der Zeichenfolge *durchschnitl* beginnt, jeweils sowohl gross- als auch kleingeschrieben. Interessant hier ist die diachrone Entwicklung der relativen Häufigkeit der entsprechenden Token im Gesamtkorpus (Abbildung 4.1). Sie bestätigt die Kernaussagen des Forschungsstandes zur Geschichte der Statistik, nämlich dass sie ab den 1830er-Jahren an Verbreitung gewinnt und, das zeigt das Korpus, dies auch in der Schweiz passiert. Die grossen Ausschläge der relativen Häufigkeit für die Jahre vor 1815 lassen sich auf einzelne Texte zurückführen. Ab etwa 1820 wird der Gebrauch stetiger und verteilt sich auf mehrere unterschiedliche Texte.

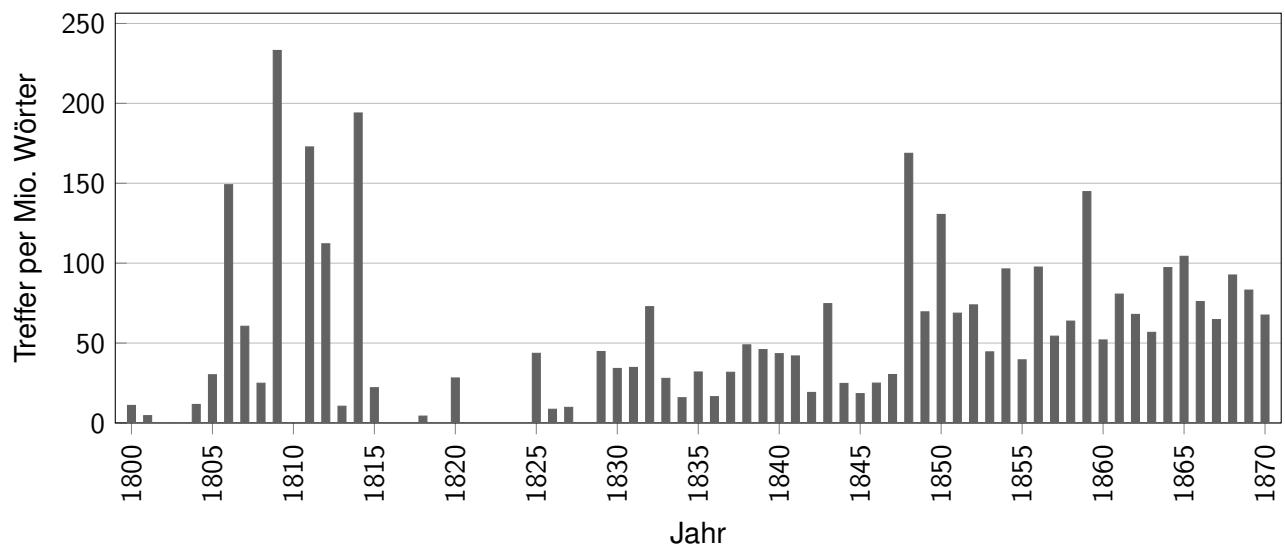


ABBILDUNG 4.1: Relative Häufigkeit für [word="durchschnitt|durchschnitl.\*%c"]

Für das Token *prozent* ergibt sich eine stärkere Etablierung als adjektivische Zusammensetzung mit einer Zahl wie *fünf-prozentiger*. Diese Verwendungen fallen in den Kontext von Finanzen und Steuern und selten von Statistik, wie die KWIC-Ansicht ergibt. Sie werden deshalb in dieser Arbeit nicht weiter verfolgt.

#### 4.2.2.2 *Statistik*

Zuletzt bleibt aus den Wörterbuchbefunden noch *Statistik*. Die Abfrage [word = ".\*statist.\*%c"] nach allen Token, die die Zeichenfolge *statist* enthalten und damit sowohl die verschiedenen Formen der Substantive, Adjektive und Verben wie auch Komposita — zusammengesetzte Wörter — umfasst, liefert zahlreiche Treffer. Die 50 häufigsten Treffer ergeben einen ersten Einblick in die Verwendungsweisen von *Statistik* (Tabelle 4.6).

Position	Token	Anzahl
1	Statistik	1'129
2	statistischen	613
3	statistische	577
4	Statistisches	110
5	statistisch	108
6	statistischer	76
7	Statistiker	75
8	statist	70
9	statistique	35
10	Schulstatistik	25
11	Forststatistik	20
12	statisti	19
13	Bevölkerungsstatistik	15
14	Obstbaustatistik	15
15	Kirchenstatistik	13
16	Statistikern	13
17	Criminalstatistik	12
18	Statistiken	12
19	statisti-	10
20	Alpenstatistik	8
21	statisti-	8
22	geographisch-statistisches	6
23	Militärstatistik	6
24	bevölkerungsstatistischen	5
25	Kriminalstatistik	5
26	Mortalitätsstatistik	5
27	Statisten	5
28	statistischem	5
29	Sterblichkeitsstatistik	5
30	Armenstatistik	4
31	geographisch-statistische	4

TABELLE 4.6: Ergebnisse für [word=".\*statist.\*%c"] (Frequenz &gt; 3).

Betrachtet man die sprachliche Oberfläche, wird deutlich, dass «Statistik» im Korpus ein morphologisch produktives Wort ist, das die Bildung einiger Komposita angeregt hat. Diese lassen sich gruppieren nach:

- geografisch-agrarischen Statistiken (*Forst-, Alpen-, Obstbau-, geographisch-*),
- eine Reihe von Sozialstatistiken (*Schul-, Criminal- bzw. Kriminal-, Armen-, Bevölkerung- und Mortalität- bzw. Sterblichkeit-*),
- sowie staatskundlichen Statistiken (*Militär-, Kirchen-*).

Vorderhand bestätigt der Blick auf die sprachliche Oberfläche der Korpusdaten die Expansion der Statistik im 19. Jahrhundert auch in der Schweiz. Staat, Wirtschaft, Landschaft und Gesellschaft werden Themen der Statistik und ein Wort wie *Alpenstatistik* wird vorstellbar. Das häufige Auftreten des Token *Schulstatistik* dürfte sich aus der Kompilationsweise des Korpus begründen. Dass in einem mit Fokus auf Schulbildung zusammengestellten Korpus allerdings ein so bunter Strauss an «Bindestrichstatistiken» zu sehen ist, ist nur durch die Etablierung von Statistik als «Denkstil» zur Behandlung gesellschaftlicher Probleme erklärbar, wie auch durch die Nähe des Bildungsthemas

zu weiteren Sozialstatistiken. Der Kontrollüberschuss, den Sozialstatistiken erzeugen, wird hier bereits ebenfalls angedeutet.

Eine diachrone Ansicht dieser «Bindestrichstatistiken» bestätigt weiterhin den Forschungsstand, indem sie zeigt, dass die morphologische Produktivität von *Statistik* in unserem Korpus erst ab zirka 1830 einsetzt (Abbildung 4.2). Bis 1870 werden die Erwähnungen solcher Statistiken zunehmend häufiger, dies nicht in absoluter Zahl, sondern bezogen auf die Anzahl der Token im Gesamtkorpus. Hier sind die vier Peaks in den Jahren 1825, 1848, 1854 sowie 1865 auffällig. Ein Blick in die Treffer zu den vier Jahren zeigt, dass die sehr hohen Trefferzahlen sich jeweils auf einen Text zurückführen lassen. Zuerst sorgt 1825 die *Geographisch-statistische Beschreibung des Cantons Bern* von Johann Rudolf Wyss für 28 der 30 Treffer des Jahres. 1848 ist es die in diesem Jahr auf Deutsch erschienene zweite Auflage von Stefano Franscini's *Neue Statistik der Schweiz*, die für 309 der 314 Treffer verantwortlich ist. 1854 ist es das Werk *Bevölkerungswissenschaftliche Studien aus Belgien*, das ebenfalls für zirka 300 der insgesamt 394 Erwähnungen sorgt. Für den letzten Peak 1865 sorgt der bayerisch-pfälzische Statistiker Georg Friedrich Kolb mit der vierten Auflage seines *Handbuchs der vergleichenden Statistik der Völkerzustands- und Staatenkunde*. Diese vier Texte können für diese Arbeit als statistische Ausreisser betrachtet werden, die die Aussage einer stetig zunehmenden Verwendung von «Statistik» nicht infrage stellen. Die Peaks können aber auch als Hinweis gesehen werden, um der Bedeutung dieser Werke für die Etablierung von Begriffen im Zusammenhang mit Statistik nachzugehen. Das ist allerdings nicht Gegenstand dieser Arbeit.

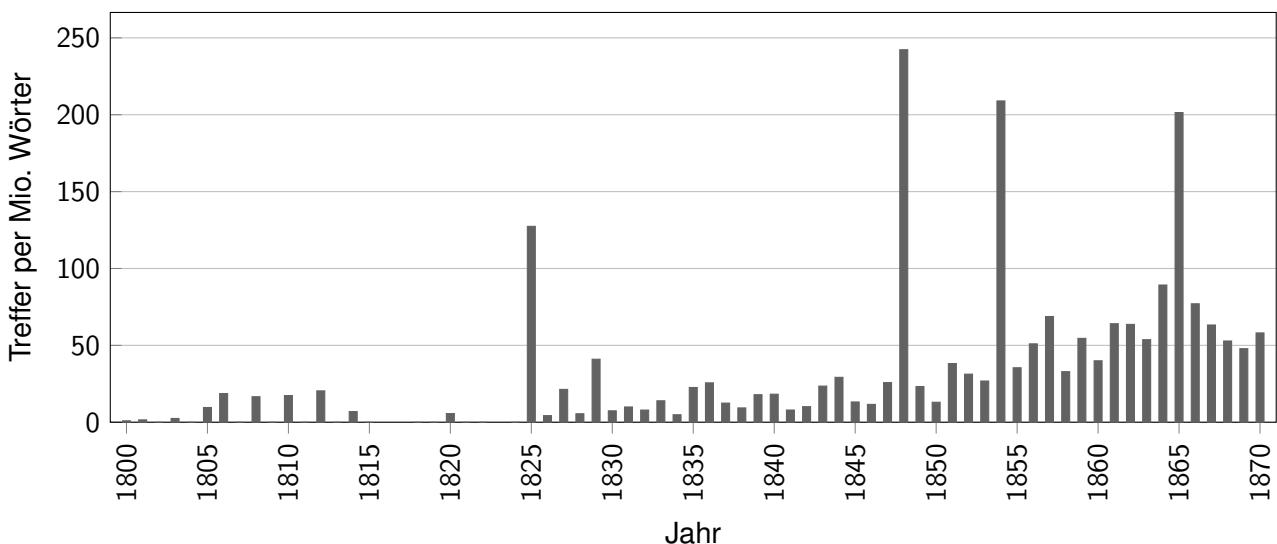


ABBILDUNG 4.2: Relative Häufigkeit für [word=".\*statist.\*%c"]

### 4.3 Zwischenfazit

Mit Vorbehalt der unsicheren Datenlage, lässt sich aus dieser ersten Exploration bereits schliessen, dass im Zusammenhang mit Statistik übliche Wörter durchaus ihren Niederschlag in unserem Korpus, und damit im schweizerischen Schuldiskurs des 19. Jahrhunderts, gefunden haben. Die oben untersuchten Begriffe *Tabelle*, *Durchschnitt*, *Prozent* und *Statistik* stehen alle für grundsätzliche Konzepte statistischen Denkens und seiner textlichen Repräsentation. Statistische Spezialtermini wie etwa Normalverteilung oder Median wurden nicht gefunden. Dies schliesst allerdings nicht aus, dass es sie im Korpus gibt und sie aufgrund der fehlerhaften OCR-Erkennung nicht gefunden werden konnten. Dennoch kann darauf geschlossen werden, dass im vorliegenden Korpus eher die Ergebnisse und Inhalte von Statistiken referiert werden, mit der Angabe von Durchschnitten als einzige analytische Prozedur.

Aus der Exploration ergibt sich auch, dass eine rein induktive Perspektive auf den hier vorliegenden Korpus zu kurz greifen würden, nicht zuletzt da die Daten mit erheblichen Unreinheiten belastet sind. Im Hinblick auf die Fragestellung nach der Digitalisierung der Schweizer Gesellschaft, verstanden als Beschreibung der Gesellschaft durch diskrete Zahlen und Kategorien, muss die Verwendung der Begriffe in ihren Kontexten betrachtet werden. Wie werden Statistiken und Tabellen verwendet? Worauf verweisen sie und welche Effekte zeitigen sie diskursiv? Dazu ist es auch hilfreich, deduktiv vorzugehen und einige Themenfelder der schweizerischen Schulgeschichte abzuschreiben.

# 5 Hypothesengestützte Annäherung

## 5.1 Verwendungszusammenhänge statistischer Termini

In Bezug auf die Fragestellung haben sich aus der induktiven, *corpus-driven* Perspektive zunächst keine klaren Erkenntnisse herausgestellt. Ziel der Arbeit ist es ja, anhand des deutschschweizerischen Schuldiskurses im 19. Jahrhundert aufzuzeigen, dass Statistik eine Funktion der «Digitalisierung» der Gesellschaft ist. Sie soll die zunehmende gesellschaftliche Komplexität der modernen, post-ständischen Gesellschaft handhabbar und beschreibbar machen. Diese Komplexität ist in der Schule besonders drängend, da sie in dieser Gesellschaft erstmals den theoretischen Anspruch erhebt, dass alle Staatsangehörigen ein gewisses Ausmass an standardisierter Volksschulbildung geniessen sollen.

Die Korpusexploration hat zumindest den Nachweis erbracht, dass statistische Termini im Untersuchungskorpus signifikant vorkommen und dies besonders in der zweiten Hälfte, das heisst ab etwa den 1830er-Jahren. Davon ausgehend sollen einige Hinweise weiterverfolgt und mit Annahmen kombiniert werden, um die Fragestellung beantworten zu können. Im Sinne des Forschungsdesigns ist dies der zweite Strang einer hypothesengestützen Korpusanalyse.

## 5.2 Was Statistiken zählen

### 5.2.1 Schülerinnen und Schüler

Die grundlegende Operation der Statistik ist das Zählen, sei es die Zahl der Staatsangehörigen, Schüler\*innen oder Diebstähle. Zählen ist aber nicht voraussetzungslos. Es braucht eine vorherige Definition dessen, was gezählt werden soll. Wie Hacking betont, braucht man zum Zählen Kategorien. Gezählt werden kann nur, was vorher als addierbar bestimmt wurde. Dafür müssen unter Umständen disparate Realitäten auf einen gemeinsamen Aspekt reduziert werden (Hacking, 1982, S. 280). Tatsächlich sind die Kategorien historischer Statistiken hochgradig variabel und ihre Entwicklung gibt Aufschluss über die Statistikgeschichte, aber auch gesellschaftliche Konstruktionen dessen, was untersucht wird. Einmal geschaffene Kategorien konnten dabei wieder auf den Untersuchungsgegenstand zurückwirken. Vanderstraeten zeigt dies am Beispiel der Einteilung in berufstätige und nicht-berufstätige Personen. Dass Individuen und ihr Beschäftigungsstatus gezählt werden, empfinden wir als normal. Es ist aber die Folge statistischer Kategorisierungen, die institutional verankert sind und auf individueller Ebene als gesellschaftliche Wirklichkeit akzeptiert werden (Vanderstraeten, 2006).

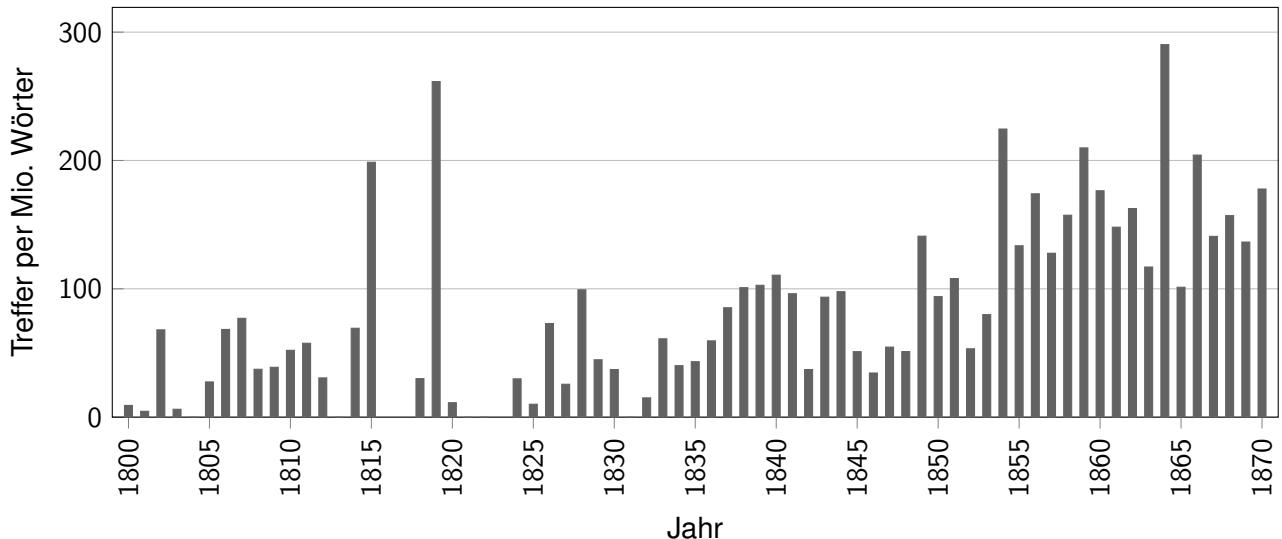


ABBILDUNG 5.1: Relative Häufigkeit für ([pos="CARD"] [lemma=".\*schüler- | .\*schülerin|Knabe|Mädchen|Tochter|Kind|Schüler|Schülerin|Bube "]) | ([lemma=".\*schüler"] .\*schülerin|Knabe|Mädchen|Tochter|- Kind|Schüler|Schülerin|Bube") [pos= "CARD"] )

Entsprechend bietet sich auch für diese Arbeit an, zu fragen, was im Korpus gezählt wird, wenn es um Schüler\*innen geht. Dafür wurden alle Vorkommnisse von Zahlen gesucht, die von Begriffen angeführt oder gefolgt werden (Abbildung 5.1).<sup>1</sup> Die Peaks in den Jahren 1815 und 1865 erklären sich durch die geringen Mengen an Text, die für diese Jahre im Korpus vertreten sind. Daher sorgen hier schon wenige Treffer (21 bzw. 11) für eine vergleichsweise hohe relative Häufigkeit. Beide röhren jeweils von einem einzigen Text her, der für die Fragestellung hier nicht relevant ist. Während die Form mit vorangestellter Zahl eher in Fließtexten auftaucht und auf separate oder textexterne Tabellen verweist, findet sich zweitere Form mit nachgestellter Zahl wiederholt in Auflistungen.

In den Ergebnissen zeigt sich, nach welchen Kategorien Schüler\*innen und Schulen gezählt wurden. Die Schulstatistik des Kantons Bern für das Schuljahr 1838/39 beispielsweise, also zu Beginn des «avalanche of numbers», zählte die Volksschüler\*innen nach Geschlecht sowie die unterschiedlichen Schularten (Abbildung 5.2).

Die Statistik macht so sichtbar, wie viele Primarschulen im Kanton bestanden und wie viele männliche und weibliche Schüler\*innen diese besuchten. Darüber hinaus bestand offensichtlich ein Interesse an öffentlichen Gemeindeschulen und wie viele seit 1831, dem Erlassjahr einer neuen liberalen Kantonsverfassung, gegründet wurden. Sie verschaffte also den Diskursteilnehmer\*innen einen Überblick über die Expansion der öffentlichen Primarschule, einem zentralen Anliegen der liberalen Bewegung.

1 Die Idee für diese Abfrage stammt von Philipp Dreesen.

II. Schulstatistisches aus dem Schuljahr 1838/39.

**A. Zahl der Volksschulanstalten im Jahr 1838.** a) Der Kanton Bern hatte in 255 Kirchengemeinden 75 Knaben-, 79 Mädchen- und 907 gemischte Schulen, zusammen 1061 öffentliche Primarschulen; davon bestanden 872 vor 1831 und wurden 189 seit 1831 errichtet. In diesen Schulen befanden sich 42,110 Knaben und 40,726 Mädchen, zusammen 82,836 Kinder. b) Mädchen-Arbeitschulen waren von den Gemeinden errichtet 329, von Privatpersonen 43, zusammen 372 mit 11,262 Kindern. c) Kleinkinderschulen: von Gemeinden errichtet 8, von Privatpersonen 14, zusammen 22 mit 564 Kindern. d) Privatprimarschulen bestanden 40 mit 681 Knaben, 696 Mädchen, zusammen mit 1377 Kindern. e) Zusammenzug. Es gab so mit 1398 Gemeindeanstalten, 97 Privatanstalten, zusammen 1495.

Die Maimühlebauschule ist nicht im Kanton Bern enthalten.

ABBILDUNG 5.2: Schulstatistik des Kantons Bern 1838/39 aus «Kanton Bern», 1840, S. 356.

Demgegenüber zeigt ein Beleg aus dem Kanton Thurgau gegen Ende des Untersuchungszeitraums eine deutlich komplexere Übersicht (Abbildung 5.3). Erhoben wird eine Reihe soziodemografischer Merkmale, anhand derer die Schüler\*innen unterschieden werden:

- Geschlecht
- Konfession
- geografische Herkunft (aus dem Schulort, bis 1 Stunde entfernt, mehr als 1 Stunde entfernt)
- Berufsstand der Eltern
- Schulbesuch

So entsteht ein Bild nicht nur über die thurgauischen Sekundarschüler\*innen, sondern auch über ihre soziale Herkunft und welche Kategorisierungen zur gesellschaftlichen Selbstbeschreibung herangezogen werden: Die Unterscheidung zwischen männlichen und weiblichen Schüler\*innen war bereits 1840 relevant und ist es immer noch. Daneben werden die Schüler\*innen kategorisiert nach Religion (reformiert, katholisch, jüdisch) und Wohnort (Stadt, Umland, Land). Besonders interessant ist die Unterteilung der Schüler\*innen nach dem elterlichen Beruf («Landwirthe», «Handwerker», «Handwerker mit Landwirtschaft», «Handelsleute», «Fabrikanten», «Beamte», «Lehrer», «Arme»). Hier zeigt sich, wie die Vielfalt der Berufe auf übersichtliche Kategorien eingegrenzt wird und der Beruf — und nur ein Beruf pro Person — zu einem Identitätsmerkmal wird. Dass es nicht unüblich war, mehrere Tätigkeiten auszuüben (vgl. Vanderstraeten, 2006, S. 199), schimmert noch in der Angabe «Handwerkern mit Landwirtschaft» durch, aber die Kategorienbildung ist schon weit fortgeschritten. Ebenso muss auch der erfasst werden, dessen Eltern keinen festen Beruf haben, hier in der Kategorie «Arm».

In ähnlicher Aussführlichkeit verbreiten sich andere Tabellen über die Aufnahme neuer Schüler (Total im Frühjahr 1864 311), über den Klassenbestand (1864/65 Total 686, nämlich 514 Knaben und 172 Mädchen, 559 Reformierte, 124 Katholiken und 3 Hebräer; 294 aus den Sekundarschulorten, 337 aus einer Entfernung bis auf 1 Stunde, 55 über 1 Stunde vom Schulort entfernt; 272 Kinder von Landwirthen, 79 von Handwerkern, 131 von Handwerkern mit Land-

ABBILDUNG 5.3: Sekundarschulstatistik Thurgau 1864/65 aus «Schulnachrichten. Thurgau», 1866, S. 24.

### 5.3 Was Statistiken machen

Im Rahmen des induktiven Vorgehens wurden die Lemmata *Tabelle*, *Durchschnitt*, *Prozent* und *Statistik* identifiziert, deren zunehmende Frequenz in unserem Korpus als Indiz für eine Zunahme statistischer Denk- und Redeweisen im Verlauf des 19. Jahrhunderts gelesen werden kann. Die blosse Zunahme dieser Ausdrücke ist aber im Zusammenhang mit der Fragestellung noch wenig aussagekräftig. Zwar haben die Komposita im Zusammenhang der Grundformen bereits gewisse Aufschlüsse über Verwendungsformen von Tabellen und Statistiken geliefert, aber für eine umfassendere Analyse bietet es sich an, die Kollokationsprofile zu analysieren. Kollokationen sind statistisch signifikante Wortpaare oder Wörter, die in Nachbarschaft zu einander auftreten (siehe 3.3.2).

Interessant für eine weitergehende Analyse ist etwa Treffer 25 aus der KWIC-Ansicht zu [word=-"Statistik"%c]: «Wie noch in vielen andern Sachen, herrscht gegenwärtig in den Schulen unseres Kantons ein wahres Wirrwarr. Die Statistik der verschiedenen Lehrmittel beweist dies» («Korrespondenzen», 1858, S. 14). Aus sprachlicher Sicht interessant ist hier, dass einer Statistik ein handelnder Charakter zugesprochen wird, im obigen Beispiel «beweist» sie etwas. Aus wissenschaftssoziologischer Perspektive lässt sich das als ein Akt der Konstruktion einer Tatsache lesen. Wie Heintz schreibt, werden hier «Zahlen [...] nicht als selektive Beschreibungen einer zugrunde liegenden Wirklichkeit angesehen, sondern mit dieser selbst gleichgesetzt» (Heintz, 2007, S. 74). Wie äussert sich dies auf sprachlicher Ebene? Dazu wurden zu den Termini .\*tabell.\* und .\*statist.\* ein Kollokationsprofil von Token erstellt, die im Abstand von bis zu fünf Token rechts oder links der Suchbegriffe stehen. Die Kollokationen wurden anschliessend nach finiten Vollverben gefiltert. Angesichts der sehr problematischen Datenqualität sind die annotierten Daten aber nur mit Vorbehalt zu lesen.

Die Treffer wurden nach dem «Log-likelihood»-Wert sortiert. Dieser ist ein Mass für die statistische Signifikanz einer Verbindung: Je höher der Log-likelihood-Wert für ein Token-Paar ist, desto eher können wir davon ausgehen, dass beide Token nicht zufällig miteinander auftauchen (Bubenhofer, 2009, S. 139–147). Wir erhalten eine Tabelle (5.1) mit 27 Treffern.

<b>Position</b>	<b>Token</b>	<b>Anzahl Kollokationen</b>	<b>Log-likelihood</b>
1	enthält	21	136.72
2	nähern	5	71
3	ergibt	13	49.05
4	entnehmen	10	47.514
5	Fügen	6	44.588
6	zahlt	8	29.178
7	weist	7	23.625
8	zeigt	14	19.798
9	ergiebt	5	15.57
10	läßt	7	14.953
11	geben	21	13.747
12	folgen	12	11.741
13	u	40	11.277
14	ließ	5	11.251
15	od	6	9.616
16	besitzen	6	8.255
17	liefert	5	7.873
18	weisen	5	7.565
19	gaben	5	3.353
20	geht	9	3.148
21	gibt	9	2.341
22	zeigen	5	2.171
23	giebt	6	1.564
24	lassen	10	-0.029
25	steht	5	-0.544
26	macht	5	-1.254
27	d	6	-8.463

TABELLE 5.1: Kollokationen zu [word=".\*tabell.\*|.\*statist.\*"] gefiltert  
nach finiten Vollverben ([pos="VFIN"])

Das Kollokationsprofil ergibt, dass vor allem Verben des Zeigens und Beweisens statistisch überzufällig häufig mit Tabellen und Statistiken korrelieren. Unter den 23 Verben mit einem positiven Log-likelihood-Wert sind allein fünfzehn, die sich diesem Wortfeld zuordnen lassen oder auf Substantive wie *Aufschluss* oder *Hinweis* verweisen (*enthält*, *ergibt*, *entnehmen*, *weist*, *zeigt*, *ergiebt*, *läßt*, *geben*, *folgen*, *liefert*, *weisen*, *gaben*, *geht*, *gibt*, *zeigen*, *giebt*). *Geht* an Position 20 bezieht sich häufig auf «hervor».

Position	Kontext vor	Suchbegriff	Kontext nach
2	Schenken wir darum schließlich auch diesem Falle noch unsere Aufmerksamkeit ! Unsere	Schulstatistik	zeigt allerdings , daß verhältnismäßig mehr Lehrer ihren Stand wechseln , als
4	zu genügen . Und in der That , ein Blick ans die	Zensurtabellen	zeigt , daß die Schulen um die Stadt herum fast alle nur
7	im hiesigen Cantone unmittelbar vor der Einführung des nencn Schulgesetzes zeigt in	statistischer	Hinsicht die auf Beilage Nr. VI. befindliche , vom Erziehungsdepartemente damals veranstaltete
8	mehrere Präsidenten dieser Behörden aus . An mehrern Orten , mie die	Vericht-erstattungstabelle	zeigt , steht es mit Rücksicht auf den Schulbesuch der Schulpflegen noch
9	Real-Abtheilung zusammen 264 , und gegenwärtig 384 Schüler ; die Zusammenstellung der	Jahrestabellen	seit 1856 zeigt , einige Schwankungen abgerechnet , eine stetige Zunahme der
10	Gemeinden wenig Er » freuliches berichtet werden . In Starkenbach zeigt die	Absenzen-tabelle	auf einen Schüler durchschnittlich 9Ve , in Wintersberg 12 und in Ennetbühl

TABELLE 5.2: KWIC-Ansicht für **zeigt** im Zusammenhang mit [word=". \*tabell.\*|.\*statist.\*"]

Die Treffer der KWIC-Ansicht für das Verb *zeigen*, die im Zusammenhang mit Schule stehen (Tabelle 5.2), behandeln Unterschiedliches, etwa die Erwerbsbiografien des Lehrpersonals (Pos. 2), einen Vergleich der Schulqualität anhand der Noten (Pos. 4), die quantitative Entwicklung der Schülerzahlen (Pos. 10) oder die Entwicklung der Schulabsenzen (Pos. 11). Gemeinsam ist den Treffern, dass Statistik etwas sichtbar macht, was ohne sie nicht sichtbar wäre und nun mit einem einfachen finiten Verb als Beweis in den Diskurs eingeführt werden kann.

Entgegen der ursprünglichen Erwartung finden sich in den Treffern kaum Belege dafür, dass Kategorien der gesellschaftlichen Selbstbeschreibung kombiniert werden. Am häufigsten zu finden sind Aufstellungen der Zahl der Schulen, der Schülerinnen und Schüler und besonders der Absenzen. Die Angaben zur Zahl der Schulen sind recht summarisch gehalten und nennen in der Regel nur Orte. Das Beispiel in Abbildung 5.4 kommt einer anfänglich erwarteten Konstruktion der schweizerischen Gesellschaft durch Statistiken noch am nächsten. Hier wird die Anzahl der Schulen pro Einwohner für einige Kantone verglichen. Ein Versuch, daraus beispielsweise zu argumentieren, dass katholische Kantone sich in Bezug auf die Schulbildung von reformierten Kantonen unterschieden, wird hingegen nicht unternommen.

**Verschiedenes.**

Der „Appenz. Ztg.“ entnehmen wir ein statistisches Fragment betreffend das Volksschulwesen, nämlich eine vergleichende Uebersicht der Zahl der Volksschulen (Primarschulen) aus elf verschiedenen Kantonen im Verhältnis zur Bevölkerung:

- 1) Zürich hat bei einer Bevölkerung von 268,000 Seelen 514 Schulen, also trifft es 1 Schule auf 524 Einwohner;
- 2) Luzern hat eine Einwohnerzahl von 132,000 Seelen und 446 Schulen, also kommt 1 Schule auf 296 Einwohner;
- 3) Schwyz hat 45,000 Einwohner und 100 Schulen, also kommt auf 1 Schule 450 Seelen;
- 4) Glarus hat 33,660 Einwohner und 57 Schulen, also 1 Schule auf 585 Seelen;
- 5) Zug hat 19,600 Einwohner und 44 Schulen, also 1 Schule auf 445 Seelen;
- 6) Freiburg hat 105,000 Einwohner und 308 Schulen, also 1 Schule auf 341 Seelen;
- 7) Solothurn zählt 70,000 Einwohner und 176 Schulen, also 1 Schule auf 397 Seelen;
- 8) St. Gallen zählt 181,000 Einwohner und 390 Schulen, also 1 Schule auf 464 Seelen;
- 9) Graubünden zählt 91,000 Seelen und 451 Schulen, also auf 1 Schule 202 Einwohner;
- 10) Aargau zählt 194,000 Seelen und 504 Schulen, also 1 Schule auf 385 Einwohner;
- 11) Appenzell A. Rh. hat bei einer Bevölkerung von 48,600 Seelen 84 Schulen (mit Inbegriff der Waisen- und Mittelschulen), also kommt 1 Schule auf 580 Einwohner.

ABBILDUNG 5.4: Auszug aus «Verschiedenes», 1864.

## 5.4 Wie Statistiken Kontrolle ausüben

Was die Statistiken und Tabellen aus Kapitel 5.3 häufig «zeigen» oder worüber sie Aufschluss geben, sind die Absenzen der Schüler\*innen. Die Tabelle und die Statistik als Kontrollinstrument staatlicher Macht ist ein klassisches Thema der Statistikgeschichte. Die Erfassung von deviantem Verhalten in Form von Suizid- oder Kriminalstatistiken machte einen grossen Teil der Flut an Statistiken ab 1830 aus (Porter, 1986, S. 30). Auch im vorliegenden Korpus spielt die Erfassung von Normabweichungen eine grosse Rolle. So sollten mit statistischen Erhebungen Schüler\*innen und Lehrer\*innen überwacht werden, um Probleme zu identifizieren und behandeln zu können (Ruoss, 2018, S. 217).

Bis 1858 hatten alle Kantone eine allgemeine Schulpflicht für den Primarschulbesuch eingeführt. Deren Durchsetzung war das ganze 19. Jahrhundert hindurch indes umstritten. Die Arbeitskraft der Kinder war vielerorts wirtschaftlich notwendig, gerade im Sommer in der Landwirtschaft. Daneben wurde die Schule auch als Ort (kantonal-)staatlicher Normierungsversuche wahrgenommen,

sodass Eltern ihre Kinder nicht auf die Schulen schicken wollten. Dagegen versuchten die Schulverwaltungen die Schulpflicht durchzusetzen (Criblez, Hofstetter und Magnin, 1999, S. 26–27). Wie bereits in Kapitel 4.2.2.1 angedeutet, spielten Statistiken und Tabellen dabei eine wichtige Rolle. Einige Treffer im Kontext des Suchbegriffs [word=".tabell.\*"] betreffen das Thema der Schulabsenzen. Diese sind in Tabelle 5.3 dargestellt.

Position	Token	Anzahl
18	Schultabellen	12
22	Versäumnißtabellen	11
23	Absenztabellen	8
38	Schultabelle	5

TABELLE 5.3: Auswahl aus Treffern für [word=".tabell.\*%c"]

Wenn wir die Suche auf Begriffe im Zusammenhang mit *Versäumniß* und *Absenz* ausdehnen, finden sich noch weitere Wörter, die die Funktion von *-tabelle* einnehmen (Tabelle 5.4).

Position	Token	Anzahl
7	Versäumnißtabellen	11
10	Absenzenverzeichnisse	8
11	Absenztabellen	8
15	Absenzenlisten	4
16	Absenzenverzeichniß	4
21	Absenzenlistcn	3
22	Absenzentabellen	3
23	Versäumnißrodel	3
24	Versäumnißtabelle	3
25	Absenzenliste	2
37	Absenzcnrapporte-	1
38	Absenzcntabellcn	1
46	Absenzen-Verzeichnisse	1
47	Absenzenberichten	1
53	Absenzenrgtster	1
55	Absenzenrodel	1
56	Absenzenrodels	1
59	Absenzentabelle	1

TABELLE 5.4: Auswahl aus Treffern für [word="Versäumniß.\*|Absenz.\*"]

Wir können daher die Begriffe *-listen*, *-verzeichnis*, *-rodel*, *-rapport* und *-berichte* in die Suche aufnehmen. Außerdem tauchen auch vom OCR falsch erkannte Token auf wie *Absenzcn*- oder *listcn*. Nehmen wir diese Begriffe in eine Suchanfrage auf, erhalten wir insgesamt 41 Belegstellen, die sich in Texten ab 1835 finden lassen. Aufgrund der insgesamt niedrigen Frequenz der Token sind die Balken in Abbildung 5.5 mit Vorbehalt zu lesen. Dennoch lässt sich festhalten, dass die Praxis der

Dokumentation von Schulabsenzen in Tabellen, Listen oder ähnlichen Verzeichnissen einen Niederschlag auf der Textoberfläche findet.

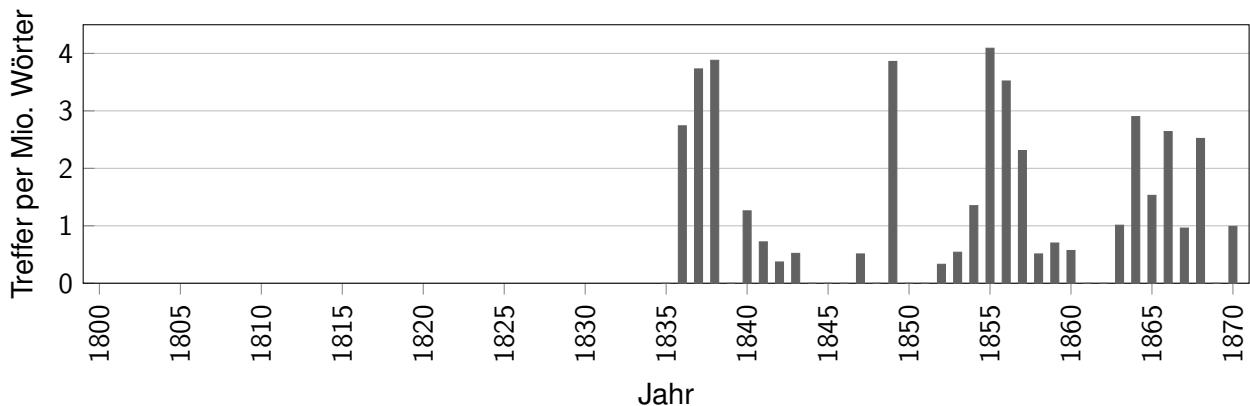


ABBILDUNG 5.5: Relative Häufigkeiten für [word="Absenztabellen|Absenzenlisten|Absenzenverzeichniß|Absenzenverzeichnisse|Absenzenlisten|Absenzentabellen|Versäumnißrodel|Versäumnißtabelle|Absenzenliste|Absenzenrapporte|Absenzen-Verzeichnisse|Absenzenberichten|Schullisten|Schultabellen|Versäumnißtabellen"]

Mit den Absenzenlisten sollte die Einhaltung der in vielen Gegenden noch ungewohnten gesetzlichen Schulpflicht durchgesetzt und kontrolliert werden. Dem Lehrpersonal wurde aufgetragen, die Listen genau zu führen und regelmässig zum Bericht einzusenden, wie etwa das Beispiel aus Appenzell Ausserrhoden zeigt: «Allein an einer richtigen und genauen Tabellenführung und konsequenter Bestrafung der Schulversäumnisse liegt viel mehr, als es den Anschein hat. Es klingt paradox und ist doch buchstäbliche Wahrheit, daß die Schultabellen einen wesentlichen Einfluß auf die Hebung der Schulen in Außerrhoden ausgeübt haben, vorab auf einen regelmässigen und fleißigen Schulbesuch; ja sie haben ganz allmälig den obligatorischen Schulzwang eingeführt» («Schulverordnungen», 1868, S. 68). So wurde die Zahl der Absenzen zu einem Gradmesser für die Reichweite der Volksbildung und es liessen sich säumige Eltern, Lehrer und Schulgemeinden identifizieren (vgl. Abbildung 5.6). Der Fortschritt der Bildungsexpansion wird also auf eine eindeutige, diskrete Zahl reduziert, welche Vergleiche ermöglicht, Muster aufzeigt und auf Steuerungsbedarf hinweist.

<p>Aus den Verfäumnißtabellen aller Schulen im ganzen Lande, vom Mai bis Weinmonat 1837, ergiebt sich die bereits <sup>13)</sup> erwähnte Anzahl nach dem Reglement entschuldigter und nicht entschuldigter Schulversäumnisse. Diese Summen unter die sämmtlichen Alltagsschüler des ganzen Landes vertheilt, ergiebt sich auf jeden derselben im Durchschnitte eine Zahl von <math>12^{2487}/5534</math> entschuldigten, von <math>6^{621}/5534</math> nicht entschuldigten, in Allem also von <math>18^{1554}/5534</math> Versäumnissen.</p> <p>Nach den einzelnen Gemeinden stellt sich die Durchschnittszahl der unentschuldigten Versäumnisse heraus, wie folgt:</p>	
Urnäsch	$9^{334}/359$
Hertsau	$6^{21}/249$
Schwellbrunn	$9^{30}/287$
Hundwil	$6^{133}/145$
Stein	$3^{98}/171$
Schönengrund	$10^{40}/67$
Waldstatt	$10^{2}/55$
Teufen	$5^{7}/613$
Bühl	$7^{1}/56$
Speicher	$2^{65}/326$
Trogen	$1^{203}/340$
Nehetobel	$7^{3}/132$
Wald	$4^{26}/41$
Grub	$5^{49}/22$
Heiden	$5^{88}/103$
Wolfhalden	$6^{7}/299$
Luzenberg	$4^{84}/121$
Walzenhausen	$3^{45}/260$
Neute	$3^{6}/148$
Gais	$9^{55}/363$

ABBILDUNG 5.6: Schulabsenzen in Appenzell 1837 aus «Nachlese», 1838, S. 29.

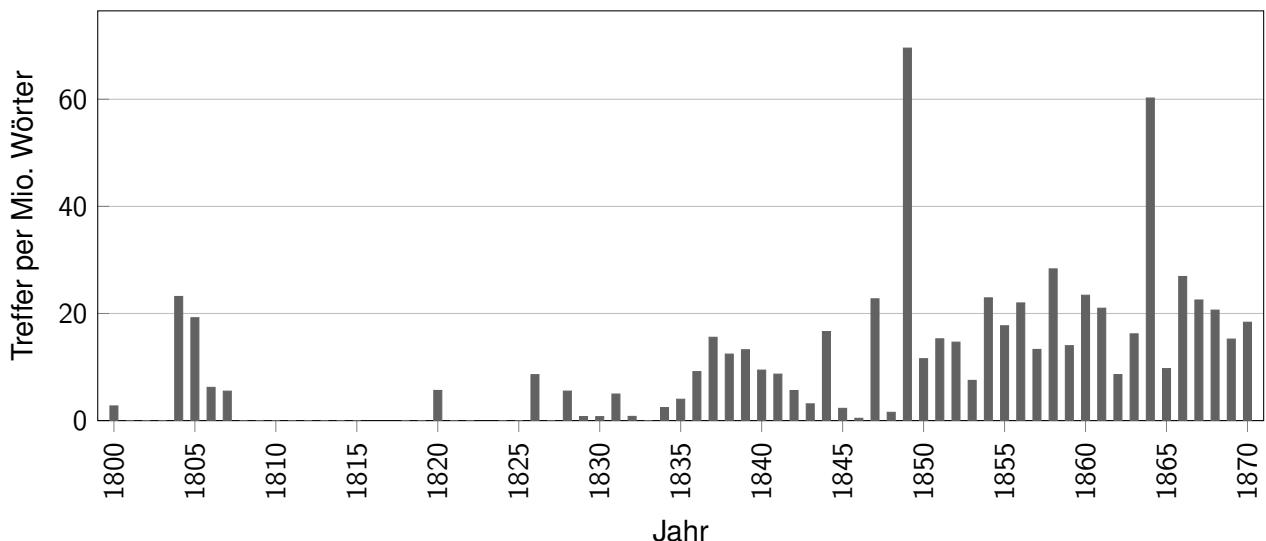


ABBILDUNG 5.7: Relative Häufigkeiten für [word="Absenz.\*|Versäumniß"&gt;%cd]

Angesichts der geringen relativen Häufigkeit der Treffer zu den Absenztabellen wurde zusätzlich nach den Token, die mit *Absenz* oder *Versäumniß* anfangen, gesucht. Diese treten ab 1834 auf (Abbildung 5.7).<sup>2</sup> Der erste Treffer ist ein 1834 Antrag des Schulkapitels Bülach bei der Zürcher Schulsynode: «Der hohe Regierungsrath möchte durch kräftige Maaßregeln (sic!) [...] den zahlreichen Absenzen Abhilfe verschaffen zu suchen» («General-Bericht der Schulkapitel an die Schulsynode», 1834, S. 30, zur Zürcher Schulsynode siehe Tröhler, 2007). Ab diesem Zeitpunkt sind die beiden Zeichenfolgen *Absenz* und *Veräumniß*, mit gewissen Konjunkturen, fast durchgängig im Korpus vertreten. Der Ausschlag für das Jahr 1849 in Abbildung 5.7 erklärt sich mit einer geringeren

2 Alle Treffer vor 1834 behandeln nicht Schulabsenzen.

Menge an Texten im Korpus mit diesem Erscheinungsjahr, sodass die relative Häufigkeit deutlich ansteigt.

— 184 —								
			Winter.	Sommer.				
Bannwyl	2. Klasse	2,5	7		Gehlken	2	5,7	
Bleienbach	3. "	4	5		Ruppoldsried	2	8	
Thunfetten	3. "	3,6	5		Bätterkinden	1. Klasse	5	10
Bübbberg	2. "	2,5	3,6	"	2. "	4	6	
Noggwyl	2. "	4	11	"	3. "	2	4	
"	4. "	3	6	"	3. "	4,6	6	
Langenthal	3. b "	5	5	"	4. "	3,6	5	
"	4. a "	3	6	"	5. "	5	6	
"	4. b "	3,5	5		Wyler	4,5	7	
Schoren		5	6		Zielebach	3,7	6	
Vogwyl	1. "	4	8					
"	2. "	4	7,5					
"	3. "	3	6,5					
"	4. "	1	3					
Madiswyl	1. "	5	10					
"	4. "	4	9					
Wobach	2. "	3	7					
Wattenbach	1. "	4	8,7					
"	2. "	5	7					
Rohrbach	4. "	4,5	10					
Kleindietwyl	2. "	2,4	2,4					
C. Amt Burgdorf.								
Burgdorf	1. Klasse	3	4					
"	2. "	5	5					
"	3. a "	5	5					
"	3. b "	5	4					
"	4. a "	1,8	3					
"	4. b "	2	2,6					
"	5. a "	1,8	1					
"	5. b "	3	1,3					
Buhwyl		5	6					
Kaltacher	2. "	5	6					
Hindelbank	1. "	5	9					
"	2. "	1,7	2,5					
Wötschwyl		3,3	6					
Kirchberg	2. "	3	10,5					
"	3. "	3	5,5					
Niederösch	2. "	3	6					
Kernenried		5	12					
Lysach	2. "	2	3,5					
Koppigen	2. "	4	8					
"	3. "	5	7					
Höchstetten	2. "	2,5	8					
Oberburg	2. "	5	4,5					
D. Amt Fraubrunnen.								
Münchenbuchsee	1. Klasse	3,5	10,5					
"	2. "	1,4	10					
Moosseedorf	1. "	3	6					
"	2. "	2	2,6					
Diemerswyl		4,5	11					
Zegenkorf	3. "	5	5					
Buzwyl	3. "	3,5	5					
Grafenried	2. "	5	5,5					
Fraubrunnen	1. "	5	7					
"	2. "	2,7	5					
Vimpach	1. "	3,5	8					
"	2. "	1,5	2,5					
Büren zum Hof	1. "	2	7,5					
"	2. "	1	6					
Schalunen		2	11					
Bangerten		3,5	6					
Scheuren		2	7,5					
Eghofen								
Ruppoldsried								
Bätterkinden								
"								
Ukenstorf								
"								
Wyler								
Zielebach								
III. Schulen mit 6–10 unentschuldigten Absenzen per Kind im Winter.								
A. Amt Wangen.								
Herzogenbuchsee	3. Klasse	9	8					
Öberönz	1. "	10	10					
"	2. "	6	7					
Röthenbach	1. "	7	9					
Graben		9	9					
Neuhaus	2. "	8	8					
Thörigen	1. "	10	11					
"	2. "	6	6					
Öschwand	1. "	7	8					
Seeberg	1. "	9	12					
"	2. "	6	6					
Walliswyl, Wangen	1. "	10	11					
Altiswyl	1. "	7,5	11					
"	2. "	6	8					
Numisberg	1. "	7	11					
Farnern		7	7					
Dürrmühle	2. "	7	12					
Dürrmühle	3. a "	6	10					
Dorf Niederbipp	3. b "	6,5	12					
Schwarzhäusern	2. "	5,5	9					
Urtenbach	3. "	5,5	6					
"	2. "	10	10					
B. Amt Narwangen.								
Narwangen (Knaben)	1. Klasse	8	10					
Bannwyl	1. "	8	8					
Bleienbach	2. "	6	8					
Noggwyl	1. "	7,5	10,5					
"	3. "	7	6,5					
Wynau	2. "	8	6,5					
"	3. "	7	6					
Langenthal	2. "	10	8					
"	3. a "	8	6					
Unterstechholz		8	9					
Rütiholen	2. "	6	6					
Madiswyl	2. "	6,6	10					
"	3. "	5,5	9					
Rohrbach	3. "	8,6	5,6					
Auswyl	2. "	8	9					
Kleindietwyl	1. "	7	8					
Leimiswyl	2. "	9	7					
Melchnau	4. "	7	6					
Buhwyl	2. "	8,5	6					
C. Amt Burgdorf.								
Heimiswyl	1. Klasse	9,7	11					
"	2. "	9,5	7					
Bariswyl		7,7	7,5					
Krauchthal	1. "	9	15					
"	2. "	8	9					

ABBILDUNG 5.8: Schulabsenzen im Oberaargau 1864 aus «Die bernische Schule auf der Anklagebank (V.)», 1864, S. 184.

Im Jahr 1864 sorgt ein mehrteiliger Aufsatz in der «Neuen Berner Schul-Zeitung», «Die bernische Volksschule auf der Anklagebank», für einen weiteren deutlichen Peak. Es ist eine bernische Entgegnung auf einen kritischen Artikel in der «Schweizerischen Lehrerzeitung» zum mangelnden Zustand der Schulen im Kanton Bern, der besonders an den vielen Absenzen festgemacht wurde. Herzstück der Replik ist eine umfassende Tabelle der Schulabsenzen im Oberaargau, aufgeteilt nach Ämtern. Angegeben werden die Absenzen der Schüler nach Ort, Klassenstufe und die Zahl der Absenzen. Als signifikantes Muster werden die Unterschiede der Absenzen im Sommer und Winter ausgemacht (Abbildung 5.8).

Wie Abbildung 5.6 tragen solche Statistiken zum «Nation Building» des Kantons Bern bei, wenn sie die Gemeinde- und Bezirkszugehörigkeiten im Diskurs reifizieren. Allerdings werden keine Bezüge zwischen den Ämtern oder Orten und ihrer Sozialstruktur hergestellt. Höchstens im Subtext lässt sich die Botschaft identifizieren, dass Bern im Vergleich mit dem Kanton Zürich nicht rückständig ist. Auffällig im Vergleich zu Abbildung 5.6 ist die Darstellungsform, die sich deutlich in Richtung von statistischen Tabellen bewegt. Noch deutlicher wird diese im Beispiel des Kantons Zürich. Sie zeigt auch die fortschreitende Kategorisierung der Absenzen, die nach verantwortbaren Absenzen, etwa durch Krankheit, und strafbare Absenzen eingeteilt werden. Es werden Durchschnitte pro Schüler berechnet und nach Bezirken aufgestellt, was einen Vergleich der Bezirke erlaubt. Warum aber die Absenzenzahlen im Schuljahr 1857/58 in Zürich 15.02, in Winterthur aber nur 9.26 betragen, wird im Begleittext nicht erklärt (Abbildung 5.9).

Übersicht der Schulversäumnisse.

Im Bezirk	Es kommen Absenzen auf den Schüler							
	der Alltagschule.			der Repetitor- und Singschule.			Gesamtdurch-	
	verant-	straß-	Total.	verant-	straß-	Total.	1856	1857
	wortete	bare.		wortete	bare.		auf 1857.	auf 1858.
Zürich . .	20,23	1,20	21,43	3,02	1,20	4,22	11,47	15,02
Affoltern .	10,07	1,05	11,12	1,10	1,04	2,14	9,18	8,03
Gösgen .	21,22	1,10	22,32	2,11	1,10	3,21	13,28	15,12
Meilen .	23,04	1,08	24,12	2,12	1,04	3,16	12,10	15,22
Hinwil .	16,21	1,19	17,40	2,22	1,12	3,34	11,40	11,29
Uster . .	18,00	1,02	19,02	2,06	1,04	3,10	8,18	11,20
Pfäffikon	21,02	0,00	21,02	3,01	1,02	4,03	11,02	14,02
Winterthur	13,05	0,00	13,05	2,09	1,03	3,12	8,55	9,26
Andelfingen	10,11	1,00	11,11	1,07	1,02	2,09	7,18	8,02
Bülach . .	12,07	2,13	14,20	2,06	2,04	4,10	7,00	10,06
Regensberg	15,07	0,00	15,07	2,12	1,06	3,18	8,24	11,27

ABBILDUNG 5.9: Schulabsenzen im Kanton Zürich 1856–1858 aus «Jahresbericht der h. Direktion des Erziehungswesens über den Zustand des zürcherischen Erziehungswesens im Schuljahre 1857–58», 1858, S. 24.

Unter den Treffern zu den Absenzen finden sich nun einige, die die Absenzenzahlen mit sozio-demografischen Mustern in Verbindung bringen, so etwa ein Bericht des Waadtländer Staatsrates

zu den Schulen im Jahr 1856. Für den Bezirk Orbe werden die sozio-ökonomischen Unterschiede zwischen den Orten im Jura und denen der Ebene als Gründe für Unterschiede der Schulabsenzen angeführt:

in den Bergen ist er [der Schulbesuch, M. K.] viel besser als in der Ebene; hier halten die ländlichen Arbeiten die Kinder oft von der Schule ab. Ferner ist der Geschmack am Unterricht in der Ebene weniger entwickelt als in den Bergen; das erklärt zum Theil die häufigen Absenzen in dem Landesteile, welcher am Fuße des Jura liegt («Mittheilungen über den Zustand und die Entwicklung des schweizerischen Schulwesens. Waadt», 1858, S. 287).

Für den Bezirk Yverdon heisst es an gleicher Stelle:

Was den Schulbesuch betrifft, so ist ein merkbarer Unterschied zwischen der Stadt und dem Lande. In der ersteren sind die Absenzen und Disciplinarfälle zahlreicher als in den Landgemeinden durch Schuld der Eltern, welche ihre Kinder unter den kleinlichsten Verwändten zurückhalten und sie durch ihr Beispiel und ihre Reden zur Mißachtung der Lehrer anleiten. In den Dörfern im Gegentheil, und besonders in denjenigen, wo ein gewisser Wohlstand herrscht, bemerkst man am Fleiße und an der Gelehrigkeit der Kinder leicht, daß die Eltern den Werth des Unterrichts und einer guten religiösen Erziehung schätzen. Davon sind auszunehmen die armen Gemeinden, wo die Gewohnheiten der Trunkenheit, des Nichtsthuns, des Bettelns das tägliche Leben vieler Familien geworden sind und auf eine traurige Weise auch die Schuljugend anstecken («Mittheilungen über den Zustand und die Entwicklung des schweizerischen Schulwesens. Waadt», 1858, S. 287).

Durch die Schulstatistik und besonders die Absenzen wird ein sozio-demografisches Bild des Bezirks gezeichnet. Interessanterweise ist es die Stadt, in der die Rückständigkeit beklagt wird. Dagegen wird die Landbevölkerung als bildungsfeiig dargestellt. Scharf von letzterer ausgenommen werden jedoch die armen Gemeinden. Hier werden Bildungsfeiiss und ökonomische Lage verknüpft. Für den Waadtländer Staatsrat scheint es klar zu sein, dass arme Familien auch bildungsferne Familien sind.

Andererseits zeigen die Belege auch die Schwierigkeiten im Umgang mit der Statistik als Überwachungstechnologie. Zahlreich werden nachlässig geführte Listen beklagt oder Lehrer verweisen auf die Kompliziertheit und die Mühen des Ausfüllens der Formulare. Der «avalanche of numbers» erzeugte nicht nur Ordnung erzeugt, sondern auch ein zu viel an Daten. Zudem war auch den Zeitgenossen bewusst, dass hinter den scheinbar eindeutigen Zahlen oft zweifelhafte Erhebungspraktiken standen. Dies ist ebenfalls ein Befund der kritischen Statistikgeschichte (nur ein Beispiel: Brückweh, 2015, S. 88). Entsprechend finden sich auch im Korpus Belege, die den Wert einer auf Zahlen reduzierten Messung der Schulqualität anzweifeln:

So wie sich aber diese Kontrolle fast ausschließlich nur auf die Absenztabellen beschränkte, so eben auch der Jahresbericht, und man gab sich so ziemlich dem Glauben hin, als seien die Absenztabellen der sicherste Maßstab zur Beurtheilung des Zustandes der Schulen («Schulvisitationen», 1855, S. 272).

## 5.5 Kontrollüberschuss von Statistiken

Das Interessante an der Entbergung latenter Muster anhand statistischer Daten ist, dass Daten stets mehr verraten können als den Zweck, für den sie erhoben werden. Dafür sorgt die ihre «radikale Rekombinationsfähigkeit» (Nassehi, 2019, S. 128). Auf Zahlen reduzierte soziale Phänomene sind frei kombinierbar. Sie können ausserhalb ihres Verwendungskontextes eingesetzt und mit anderen Zahlen in Bezug gesetzt werden: Daten werden verknüpft und neue Einsichten in gesellschaftliche Muster sind im Diskurs formulierbar. Auch dieses Phänomen findet sich im vorliegenden Korpus im Fall von Statistiken. So behandelte ein anonymer Autor in der «Neuen Berner Schulzeitung» 1865 die Frage «Ist die physische Entartung der jetzigen Generation eine Thatsache?» und kombinierte in seiner Problemanalyse medizinische und Schulstatistiken:

Ueber den nachtheiligen Einfluß unzweckmäßiger Einrichtungen geben uns den sichersten Aufschluß die Journale der Aerzte, die Todtenlisten und die Zahl der stets auf betrübende Weise zunehmenden Fälle von Rückgratskrümmungen, Brustkrankheiten, Kopfschmerzen und Augenleiden unserer Schuljugend und unsere daher rührenden Absenzenverzeichnisse (J., 1865, S. 70).

Auf sprachlicher Ebene äussert sich dies in diesem Zitat in einer Aufzählung verschiedener Arten von Listen, deren Kombination «Aufschluß gibt», und zwar über die baulichen Zustände der Schulhäuser. Das ist etwas völlig anderes als das, wofür die Absenzenverzeichnisse, Medizinalstatistiken und Mortalitätslisten ursprünglich erstellt worden waren.

Den Statistiken ist also ein «Kontrollüberschuss» (Dirk Baecker, zit. n. Nassehi, 2019, S. 43–44). Durch ihre Rekombinationsfähigkeit enthalten Daten stets ein mehr Informationen, dass sich für Rückschlüsse auf Regelungsbedürftiges nutzen lässt. Ein Beispiel aus dem Korpus ist dafür Treffer 10 in Tabelle 5.2. Darin kommentiert der Autor den Amtsbericht des evangelischen Erziehungsrats des Kantons St. Gallen für das Jahr 1855. Die in der Statistik enthaltenen Zahlen werden als Fakten zitiert, denen eine Beweisfunktion zukommt. Auf dieser Grundlage schliesst der Kommentator, die auf tiefgehende Missstände im Umkreis der Schule:

«In Starkenbach zeigt die Absenzentabelle auf einen Schüler durchschnittlich 9½, in Wintersberg 12 und in Ennetbühl 17 unentschuldigte Abwesenheiten.» Wo solche Krebsschäden sich zeigen, da suche man die Ursache nicht an einem Orte, und erinnere herhaft nicht nur die Eltern, sondern auch Lehrer und Schulräthe an ihre Pflichten. Nur radikales Einschreiten kann da helfen, wo Nachlässigkeit und Gleichgültigkeit das Schulleben anzugreifen drohen (Sch., 1856).

Hier wird von den diskursiv eingeführten Zahlen auf gesellschaftliche Zustände geschlossen. Die Statistik erweist sich nicht nur als Kontrollmittel der Schülerinnen und Schüler sowie ihrer Eltern, sondern auch von Amtspersonen wie den Schulräten. Die Tabellen entdecken nicht nur ein Muster im Absenzverhalten der Schülerinnen, sondern vielmehr Nachlässigkeiten im Disziplinierungsapparat. Sie sind ein Mittel der «Biopolitik». Es wird nicht gemessen, um zu wissen, sondern um andere zu einem erwünschten Verhalten zu disziplinieren. Eine solche datengestützte Form der Disziplinierung ist nur möglich in einer Gesellschaft, die digital sehen gelernt hat, ergo digitalisiert ist (Nassehi, 2019, S. 309–310).

## 5.6 Zwischenfazit

Statistiken im Schweizer Schuldiskurs des 19. Jahrhunderts zählten Schüler\*innen vor allem nach Schulort und Geschlecht. Festgestellt wurde auch eine zunehmend komplexere Kategorisierung, die weitere Kategorien erhab. Allerdings konnte im Korpus weder eine klare Tendenz dahingehend festgestellt werden, dass die Erhebung dieser zusätzlichen Kategorien weiter verbreitet gewesen wäre. Auch Belege zu einer expliziten Kombination der Schüler\*innenzahlen wurden kaum gefunden. Was die Statistiken im Diskurs «zeigen», wenn auf sie sprachlich verwiesen wurde, sind meist eben die Zahlen zur Anzahl nach Ort aufgeschlüsselt.

Anhand der Schulabsenzen konnten dagegen Belege gefunden werden, die die Hypothesen dieser Arbeit stützen. Es konnte gezeigt werden, dass die Statistiken und Tabellen eine Überwachungs- und Kontrollfunktion innehatten. Diese erstreckte sich nicht nur auf die Schüler\*innen und ihre Eltern, sondern auch auf Lehrer und die Schulgemeinden.

# 6 Fazit

## 6.1 Die Digitalisierung der Schweizer Gesellschaft anhand des Schuldiskurses

Ziel der Arbeit war es, den Diskurs um die Volksschule in der Schweiz im 19. Jahrhundert mit korpuslinguistischen Methoden zu analysieren. Dabei sollte herausgearbeitet werden, wie Statistik auch in der Schweiz dazu beitrug, Vorstellungen von Gesellschaft möglich zu machen und wie sich dies diskursiv im vorliegenden Korpus niederschlug. Die Annahme war, dass wir deshalb bereits im 19. Jahrhundert von einer Digitalisierung der Schweizer Gesellschaft sprechen können, die sich nur durch digitale Techniken beschreiben liess. Diese digitalen Techniken reduzierten die komplexe Gesellschaft der Individuen auf diskrete Zahlen und Kategorien und erlaubten es so, latente Verhaltensmuster sichtbar zu machen.

### 6.1.1 Statistikgeschichte

Zunächst wurde in Kapitel 1.3 gefragt, ob es beobachtbare Veränderungen auf der sprachlichen Oberfläche des Diskurses gibt, die auf eine zunehmende Bedeutung statistischer Redeweisen gibt. Die explorative Analyse in Kapitel 4 konnte nachweisen, dass sich im vorliegenden Korpus ab etwa 1830 zunehmend statistische Termini finden. Tabellen, Statistiken, Durchschnitte und Prozentangaben tauchen auf der sprachlichen Oberfläche des Diskurses auf. Am Wort *Statistik* konnte gezeigt werden, dass es Kompositbildungen anregt. Im Diskurs tauchen Statistiken aller Art auf, etwa zu Schulen, zur Kriminalität oder zum Obstbau. Dass es möglich ist, solche Komposita zu bilden, illustriert, wie die statistische Selbstbeobachtung dazu führt, unterschiedliche Sozialbereiche und Themenfelder abzustecken.

### 6.1.2 Statistik im Schuldiskurs

In Kapitel 5 konnten zuerst in Bezug auf Frage 3 gezeigt werden, in welche Gruppen Schüler\*innen kategorisiert wurden und wie diese Kategorien als Repräsentation von Gesellschaft gelesen werden können. Es waren dies zunächst Geschlecht und regionale Kriterien wie die Aufteilung in Schulbezirke. Gegen Ende des Untersuchungszeitraums konnte an einem Beispiel eine zunehmende Komplexität belegt werden, die auch Religion und soziodemografischen Status des Vaters mit einbezog.

Allerdings konnte kaum, wie von Frage 4 gefragt, nachgewiesen werden, dass diese Kategorien gesellschaftlicher Selbstbeschreibung häufig zueinander in Bezug gesetzt wurden. Schüler\*innenzahlen wurden in Tabellen und Listen, meist nach Bezirk oder Ort, rapportiert, aber nur wenige Belege gefunden, wie diese Kategorien diskursiv verbunden wurden, um Muster sozialen Verhaltens

zu belegen. Argumente wie beispielsweise «Die Statistik zeigt, dass katholische Kantone weniger Schulen besitzen und ärmer und weniger gebildet sind» wurden explizit nur selten gemacht.

Anhand von Statistiken und Tabellen zu den Schulabsenzen konnte deutlicher gezeigt werden, dass Techniken des «digitalen Sehens» im Schweizer Diskurs um die Schule etabliert waren. Hier konnten zentrale Annahmen bestätigt werden, etwa dass Statistiken dafür eingesetzt wurden, Muster des Verhaltens sichtbar zu machen und zu adressieren. Außerdem wurde der den digitalen Techniken inhärente Kontrollüberschuss aufgezeigt. Durch die Rekombination der Daten wurde beispielsweise sichtbar gemacht, in welchen Orten die Schulabsenzen höher waren und dies mit der Sozialstruktur der betroffenen Gemeinden in Verbindung gebracht.

### **6.1.3 Doch keine Digitalisierung im 19. Jahrhundert?**

Die These einer Digitalisierung der Schweizer Gesellschaft zwischen 1830 und 1870 im Spiegel der Schulstatistik kann daher in Ansätzen bestätigt werden. Sie überzeugt vor allem für die allgemeine Entwicklung des Schulwesens und die Durchsetzung der allgemeinen Schulpflicht mittels Absenzenlisten. Mittels der Schul- und Schülerzahlen wurde der Fortschritt des Schulausbaus verfolgt und anschaulich gemacht. Kantonale Darstellungen zeigten, wie es um den regionalen Ausbau der Schulen stand, aber es wurden nur selten konkrete Zusammenhänge zwischen Sozialstruktur und Schulausbau formuliert.

Dies könnte darauf hinweisen, dass das Statistikwesen der Schweiz doch einen Nachzügler im europäischen Vergleich darstellt. Diese These wird von der bisherigen Forschung so vertreten und begründet mit der schwachen Position des Eidgenössischen Bureaus für Statistik gegenüber den Kantonen. Ruoss beispielsweise untersucht differenzierte Schulstatistiken ebenfalls erst ab 1890 (Ruoss, 2018). Um diese These genauer überprüfen zu können, müsste das Korpus auf Quellen nach 1870 ausgeweitet werden.

Ein weiterer Grund, weshalb die Hypothese nicht überzeugend bewiesen werden konnte, könnten die Korpusquellen sein. Genügen die im Korpus vertretenen Texte doch nicht für eine entsprechende Untersuchung oder müsste es erweitert werden? Die angewandten Suchtstrategien könnten ebenfalls im Lichte dieser Untersuchung optimiert werden. Es stehen weitere Ansätze der maschinellen Textanalyse zur Verfügung, die hier nicht angewendet wurden. Das gewählte Vorgehen war ausserdem einigermassen eklektisch.

Ein wichtiges Problem des Korpus stellt die problematische Datenqualität dar. Wie an mehreren Stellen gezeigt, sind die OCR-erkannten Texte mit vielen Fehlern behaftet, besonders die von Frakturschriften. Zahlreiche Wörter wurden nicht oder falsch erkannt und werden von den Abfragen nicht erfasst, was vor allem induktive Zugänge erschwert. Hier könnte eine auf maschinellem Lernen basierte Texterkennung mit einem für deutschsprachige Fraktur optimierten Modell die Datenqualität deutlich verbessern («Druckwerke», 2022; «German Fraktur 19th–20th century», 2022).

Als Folge der mangelnden Datenqualität ist die Lemmatisierung und Annotierung des Korpus ungenau. Allerdings kann man davon ausgehen, dass sich die OCR-Fehler als statistischer Noise gleichmäßig verteilen und die Aussagen nicht zu stark verzerrten. Generell fragt sich, ob das Problem für retrodigitalisierte Quellen in den Griff zu bekommen ist oder die Korpusgrösse so stark ausgeweitet werden muss, damit die OCR-Fehler noch mehr als Noise untergehen. Letztlich schränken diese Probleme aber die Aussagekraft dieser Untersuchung erheblich ein.

## 6.2 Das Potential der korpuslinguistischen Methode für die Sozialgeschichte

Kann eine korpuslinguistische Analyse historischer Diskurse das Versprechen einlösen, das die Begriffsgeschichte der 1970er-Jahren der Sozialgeschichte versuchte anzubieten? Programmatisch war der Anspruch der Begriffsgeschichte, mehr zu liefern als eine Ideengeschichte, die den «Hohenkamm» kanonischer Texte abschritt. Busse zufolge ist sie an diesem Vorhaben jedoch gescheitert (Busse, 1987, S. 50–71, als Entgegnung darauf Dipper, 2000, S. 306). Tatsächlich bietet die Korpuslinguistik die Chance, das begriffsgeschichtliche Arbeiten zu erweitern, indem eine grössere Spannbreite an Texten in die Analyse einbezogen wird. So können Semantisierungsprozesse der untersuchten Begriffe breiter untersucht werden und, sofern die Quellen dafür verfügbar sind, auch Texte aus den «Niederungen» einbezogen werden. Nicht zu unterschätzen ist die Funktion einer maschinellen Suche im Korpus, die Belege aus Texten zutage fördern kann, in denen sie nicht erwartet werden. Das erhöht die Repräsentativität korpusgestützt arbeitender Diskursanalysen.

CQPWeb und andere Tools stellen aber nicht von selber neue Zusammenhänge her. Entscheidend bleibt die Fragestellung, die Modellierung des Diskurses und die Auswahl der Texte. Die resultierenden Ergebnisse von CQPWeb und anderen Tools sind wiederum interpretationsbedürftig (Schwandt, 2018, S. 33). Sie müssen anhand der Korpusdaten evaluiert werden und lösen so neue Fragen an das Korpus aus, in einem fortlaufenden Feedback loop (Bubenhofer, 2009, S. 320–322).

## 6.3 Limitationen

Ist die hier verfolgte Methodik einer qualitativen Diskursanalyse überlegen? Nein, vielmehr müssen sich beide Methoden ergänzen. Der Mehrwert des Verfahrens dieser Arbeit besteht darin, dass Befunde quantitativ abgestützt und einzelne Aussagen nicht willkürlich als diskursbeeinflussend beschrieben werden können. Es gibt so eine höhere Überprüfbarkeit. Andererseits braucht es neben der quantitativen Analyse zwingend auch den qualitativen Blick in die Ergebnisse und einen hermeneutischen Verstehensprozess, um die Belege deuten zu können. Das gilt sowohl für die absoluten Frequenzen als auch die Kollokationslisten. Was plausibel und relevant ist, unterliegt immer der Fragestellung und somit auch einem Interpretationsprozess.

Die Fragestellung bestimmt ausserdem massgeblich die Zusammenstellung des Korpus. Weiter hängt die Zusammenstellung des Korpus von der Verfügbarkeit der Quellen ab: Nur ein kleiner Teil

der historischen Quellen ist digitalisiert und Ansätze wie die *microstoria* oder «Geschichte von unten» haben zurecht darauf hingewiesen, Überlieferungszusammenhänge ebenfalls in die Quellenkritik aufzunehmen. Was überliefert wird, vor allem in gedruckter Form in Archiven und Bibliotheken, sind Dokumente der Elite. Auch diese lassen sich auf die Spuren nichthegemonialer Diskurse lesen (Guha, 1987). Dies geht aber kaum mit den hier vorgestellten Methode oder anderen Formen des «distant reading». Diese Fragen verschärfen sich, denn die technologischen Voraussetzungen für die Digitalisierung von Texten sind weltweit ungleich verteilt und bevorzugen das, was in (westlichen) Bibliotheken und Archiven überliefert wurde und als bewahrens- und digitalisierungswürdig angesehen wird (Putnam, 2016).

Die in dieser Arbeit angewendeten quantitativen Verfahren sind nicht voraussetzungslos und «naiv». Sie sind statistische Versuche, Signifikanz zu modellieren, aber auch sie gehen von Voraussetzungen aus und treffen Annahmen. Es gibt gar keine Möglichkeit, einen Korpus ohne Vorannahmen zusammenzustellen und diesen komplett objektiv zu untersuchen. Ähnlich wie die Statistiken, die Thema dieser Arbeit sind, sprechen auch Sprachdaten niemals für sich selbst. Sie werden genauso unter gewissen Gesichtspunkten hergestellt und müssen in den Diskurs eingebunden werden.

Quellenkritik ist auch bei digitalen Korpora wichtig. Ramisch zeigt, dass sogar ein redaktionell aufwändig erstelltes Korpus wie die Protokolle des Deutschen Bundestags erhebliche Probleme aufweist (Ramisch, 2022). Die Datenqualität ist in dem hier verwendeten Korpus noch weniger zufriedenstellend. Viele Historiker dürften aber mit einem Blick auf die zahlreichen nicht oder falsch erkannten Wörter oder die Vermischung von Paratext und Text in der KWIC-Ansicht die Aussagekraft in Zweifel ziehen und die Vorzüge eines traditionellen hermeneutischen Verfahrens der Diskursanalyse betonen.

## 6.4 Desiderate

Diese Arbeit konnte die Hypothese einer Digitalisierung der Schweizer Gesellschaft im Spiegel der Schulstatistiken nicht vollständig positiv beantworten. Zu überlegen ist daher, das Korpus um die Jahre 1870 bis 1900 zu erweitern. Ebenfalls fruchtbar könnte es sein, die vermuteten Zusammenhänge zwischen statistischer Beschreibung und Gesellschaftskonzeption auf Ebene einzelner Kantone zu untersuchen. Kantonale Identitäten spielten in der «Erfindung der Schweizer Nation» eine wichtige Rolle und staatliche Zentralisierungsprozesse wurden oft eher mit dem Kanton als dem Bundesstaat identifiziert. Das betraf auch die Schulen.

In methodischer Hinsicht bieten die derzeit populären Ansätze des *Natural Language Processing* viele Möglichkeiten, um das vorliegende Korpus zu bearbeiten. Sie setzen oft auf maschinelles Lernen und eine datengetriebene Auswertung. Zu denken ist etwa an Methoden wie *topic modeling*, eine *ngram*-Analyse oder *Word embeddings*. Diesen Ansätzen ist gemeinsam, dass sie ein beinahe alinguistisches Verständnis von Text haben. Dieser wird als Datensatz aufgefasst, in dem sich

statistische Muster finden lassen. Bei diesen Ansätzen stellt sich die Frage, welche Rolle eine hermeneutische Analyse von Sprache noch spielen kann. Allerdings lässt sich mit guten Gründen daran festhalten, dass die Antworten von ChatGPT & Co. nicht von selbst verstanden werden können, sondern nur Ansatz für weitere Fragen sind (Bubenthaler, 2018).

# Bibliografie

- Anderson, B. (1991). *Imagined communities: reflections on the origin and spread of nationalism* (Überarb. und erw. Aufl.). Verso.
- Berger, P. L., Luckmann, T., & Plessner, H. (2018). *Die gesellschaftliche Konstruktion der Wirklichkeit: eine Theorie der Wissenssoziologie* (27. Aufl.). Fischer.
- Bloch Pfister, A. (2007). *Priester der Volksbildung: der Professionalisierungsprozess der Zürcher Volksschullehrkräfte zwischen 1770 und 1914*. Chronos.
- Bosse, H. (2012). *Bildungsrevolution 1770–1830* (N. Ghanbari, Hrsg.). Winter.
- Brückweh, K. (2015). *Menschen zählen: Wissensproduktion durch britische Volkszählungen und Umfragen vom 19. Jahrhundert bis ins digitale Zeitalter*. De Gruyter Oldenbourg.
- Bubenhofer, N. (2009). *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. De Gruyter.
- Bubenhofer, N. (2017). Kollokationen, n-Gramme, Mehrworteinheiten. In K. S. Roth, M. Wengeler & A. Ziem (Hrsg.), *Handbuch Sprache in Politik und Gesellschaft* (S. 69–93). De Gruyter.
- Bubenhofer, N. (2018). Wenn „Linguistik“ in „Korpuslinguistik“ bedeutungslos wird. Vier Thesen zur Zukunft der Korpuslinguistik. *Osnabrücker Beiträge zur Sprachtheorie*, 92, 17–30.
- Bubenhofer, N., & Scharloth, J. (2014). Korpuspragmatische Methoden für kulturanalytische Fragestellungen. In N. Benitt, C. Koch, K. Müller, S. Saage & L. Schüler (Hrsg.), *Kommunikation, Korpus, Kultur: Ansätze und Konzepte einer kulturwissenschaftlichen Linguistik* (S. 47–66). Wissenschaftlicher Verlag Trier.
- Buchner, M., Jopp, T. A., Spoerer, M., & Wehrheim, L. (2020). Zur Konjunktur des Zählens – oder wie man Quantifizierung quantifiziert. Eine empirische Analyse der Anwendung quantitativer Methoden in der deutschen Geschichtswissenschaft. *Historische Zeitschrift*, 310(3), 580–621. <https://doi.org/10.1515/hzhz-2020-0019>
- Burckhardt, D., Geyken, A., Saupe, A., & Werneke, T. (2019). Distant Reading in der Zeitgeschichte. Möglichkeiten und Grenzen einer computergestützten Historischen Semantik am Beispiel der DDR-Presse. *Zeithistorische Forschungen*, 16(1), 177–196. <https://doi.org/10.14765/ZZF.DOK-1345>
- Busse, D. (1987). *Historische Semantik. Analyse eines Programms*. Klett-Cotta.

- Busse, D., Stukenbrock, A., & Scharloth, J. (2000). Historische Diskurssemantik. Ein linguistischer Beitrag zur Analyse gesellschaftlichen Wissens. *Sprache und Literatur in Wissenschaft und Unterricht*, 31(86), 39–53.
- Busse, D., & Teubert, W. (1994). Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik* (S. 10–28). Westdeutscher Verlag.
- Bütikofer, A. (2006). *Staat und Wissen: Ursprünge des modernen schweizerischen Bildungssystems im Diskurs der Helvetischen Republik*. Haupt.
- Criblez, L. (2013). Die Schweizerische gemeinnützige Gesellschaft – ein pädagogisches Netzwerk in der Schweiz des 19. Jahrhunderts. In H.-U. Grunder, A. Hoffmann-Ocon & P. Metz (Hrsg.), *Netzwerke in bildungshistorischer Perspektive* (S. 167–178). Klinkhart.
- Criblez, L., & Hofstetter, R. (1998). Erziehung zur Nation: nationale Gesinnungsbildung in der Schule des 19. Jahrhunderts. In U. Altermatt, C. Bosshart-Pfluger & A. Tanner (Hrsg.), *Die Konstruktion einer Nation: Nation und Nationalisierung in der Schweiz, 18.–20. Jahrhundert* (S. 167–187). Chronos.
- Criblez, L., Hofstetter, R., & Magnin, C. (1999). Einleitung. In L. Criblez, C. Jenzer, R. Hofstetter & C. Magnin (Hrsg.), *Eine Schule für die Demokratie. Zur Entwicklung der Volksschule in der Schweiz im 19. Jahrhundert* (S. 19–35). Peter Lang.
- Criblez, L., Jenzer, C., Hofstetter, R., & Magnin, C. (Hrsg.). (1999). *Eine Schule für die Demokratie: zur Entwicklung der Volksschule in der Schweiz im 19. Jahrhundert*. Peter Lang.
- Dahn, N., & Boser, L. (2015). Learning to See the Nation-State. History, Geography and Public Schooling in Late 19th-Century Switzerland. *Bildungsgeschichte / International Journal for the Historiography of Education*, 5(1), 41–56.
- Die bernische Schule auf der Anklagebank (V.) (1864). *Neue Berner Schul-Zeitung*, 7(46), 183–185.
- Die neuesten Schulverordnungen von Appenzell-Außen- und Innerrhoden. (1868). *Appenzellische Jahrbücher*, 9(6), 56–73.
- Dilthey, W. (1992). *Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften* (B. Groethuysen, Hrsg.; 8., unveränd. Aufl.). Teubner.
- Dipper, C. (2000). Die „Geschichtlichen Grundbegriffe“: Von der Begriffsgeschichte zur Theorie der historischen Zeiten. *Historische Zeitschrift*, 270(1), 281–308. <https://doi.org/10.1524/hzhz.2000.270.jg.281>

- Dreesen, P., & Stücheli-Herlach, P. (2019). Diskurslinguistik in Anwendung. Ein transdisziplinäres Forschungsdesign für korpuszentrierte Analysen zu öffentlicher Kommunikation. *Zeitschrift für Diskursforschung*, 7(2), 123–162. <https://doi.org/10.3262/ZFD1902123>
- Druckwerke [OCR-BW]. (2022). Verfügbar 4. Juni 2023 unter <https://ocr-bw.bib.uni-mannheim.de/anwendung/druckwerke/#Hinweise-zur-Texterkennunge/>
- Foucault, M. (2004). *Geschichte der Gouvernementalität*. Suhrkamp.
- Foucault, M. (2021). *Die Ordnung des Diskurses* (16. Auflage, erweiterte Ausgabe). Fischer-Taschenbuch.
- Gabrielatos, C. (2018). Keynes analysis. Nature, metrics and techniques. In C. Taylor & A. Marchi (Hrsg.), *Corpus approaches to discourse: a critical review* (S. 225–258). Routledge.
- General-Bericht der Schulkapitel an die Schulsynode. (1834). *Bericht über die Verhandlungen der Zürcherischen Schulsynode*, 1, 22–30.
- German Fraktur 19th–20th century* [READ-COOP]. (2022). Verfügbar 4. Juni 2023 unter <https://readcoop.eu/de/modelle/german-fraktur-19th-20th-century/>
- Graf, R. (2022). Pestalozzi, Johann Heinrich. In *Historisches Lexikon der Schweiz*. Verfügbar 19. April 2023 unter <https://hls-dhs-dss.ch/de/articles/009054/2022-02-15>
- Graham, S., Weingart, S., & Milligan, I. (2012, 2. September). *Getting Started with Topic Modeling and MALLET* [Programming Historian]. Verfügbar 5. Februar 2023 unter <https://programminghistorian.org/en/lessons/topic-modeling-and-mallet>
- Grunder, H.-U. (2012). Primarschule. In *Historisches Lexikon der Schweiz*. Verfügbar 26. April 2023 unter <https://hls-dhs-dss.ch/de/articles/010402/2012-06-14/>
- Guha, R. (1987). Chandra's Death. In R. Guha (Hrsg.), *Subaltern Studies V. Writings on South Asian History and Society* (S. 135–165). Oxford University Press.
- Hacking, I. (1982). Biopower and the Avalanche of Printed Numbers. *Humanities in Society*, 5, 279–295.
- Hacking, I. (1990). *The Taming of Chance*. Cambridge University Press.
- Hardie, A. (2014, 28. April). *Log Ratio – an informal introduction* [ESRC Centre for Corpus Approaches to Social Science (CASS)]. Verfügbar 23. Mai 2023 unter <https://cass.lancs.ac.uk/log-ratio-an-informal-introduction/>
- Hardie, A. (2021). *CQPweb* (Version 3.2.43). <https://cwb.sourceforge.io/cqpweb.php>
- Heintz, B. (2007). Zahlen, Wissen, Objektivität: Wissenschaftssoziologische Perspektiven. In A. Mennicken & H. Vollmer (Hrsg.), *Zahlenwerk. Kalkulation, Organisation und Gesellschaft* (S. 65–85). VS Verlag für Sozialwissenschaften.

- Hodel, T. (2013). Das kleine Digitale: Ein Plädoyer für Kleinkorpora und gegen Grossprojekte wie Googles Ngram-Viewer. *Nach Feierabend. Zürcher Jahrbuch für Wissenschaftsgeschichte*, 9, 103–119. <https://doi.org/10.5167/UZH-82205>
- Hohls, R. (2018). Digital Humanities und digitale Geschichtswissenschaften. In L. Busse, W. Enderle, R. Hohls, T. Meyer, J. Prellwitz & A. Schuhmann (Hrsg.), *Clio-Guide. Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften* (2. erw. und aktualisierte Aufl., A.1–1–A.1–34). Humboldt-Universität.
- Holenstein, A. (2014). Reform und Rationalität. Die Enquêteen in der Wissens- und Verwaltungsgeschichte der Helvetischen Republik. In D. Tröhler (Hrsg.), *Volksschule um 1800. Studien im Umfeld der Helvetischen Stapfer-Enquête 1799* (S. 13–32). J. Klinkhardt.
- J., J. J. (1865). Die erste obligatorische Frage. *Neue Berner Schul-Zeitung*, 8(19), 73–74.
- Jahresbericht der h. Direktion des Erziehungswesens über den Zustand des zürcherischen Erziehungswesens im Schuljahr 1857–58. (1858). *Bericht über die Verhandlungen der Zürcherischen Schulsynode*, 25, 22–58.
- Jost, H. U., & Malaguerra, C. (2016). *Von Zahlen, Politik und Macht: Geschichte der schweizerischen Statistik*. Chronos.
- Kämper, H. (2018). Diskurslinguistik und Zeitgeschichte. In I. H. Warnke (Hrsg.), *Handbuch Diskurs* (S. 53–74). De Gruyter. <https://doi.org/10.1515/9783110296075-003>
- Kanton Bern. (1840). *Allgemeine schweizerische Schulblätter*, 6(7), 349–358.
- Korrespondenzen. (1858). *Neue Berner Schul-Zeitung*, 1(4), 14–16.
- Koselleck, R. (1972). Begriffsgeschichte und Sozialgeschichte. In P. C. Ludz (Hrsg.), *Soziologie und Sozialgeschichte* (S. 116–131). VS Verlag für Sozialwissenschaften.
- Koselleck, R., Brunner, O., & Conze, W. (Hrsg.). (1972). *Geschichtliche Grundbegriffe: historisches Lexikon zur politisch-sozialen Sprache in Deutschland* (Bde. 8). Klett-Cotta.
- Landwehr, A. (2018). *Historische Diskursanalyse* (2., aktualisierte Aufl.). Campus.
- Lemke, M., & Stulpe, A. (2015). Text und soziale Wirklichkeit: Theoretische Grundlagen und empirische Anwendung von Text-Mining-Verfahren in sozialwissenschaftlicher Perspektive. *Zeitschrift für germanistische Linguistik*, 43(1), 52–83. <https://doi.org/10.1515/zgl-2015-0003>
- Lemnitzer, L., & Zinsmeister, H. (2015). *Korpuslinguistik: eine Einführung* (3., überarbeitete und erweiterte Aufl.). Narr Francke Attempto.
- Maissen, T. (2011). *Geschichte der Schweiz* (3. korrigierte Aufl.). Hier + Jetzt.
- Marcacci, M. (2022). Franscini, Stefano. In *Historisches Lexikon der Schweiz*. Verfügbar 12. Februar 2023 unter <https://hls-dhs-dss.ch/de/articles/003508/2022-04-21/>

- Mittheilungen über den Zustand und die Entwicklung des schweizerischen Schulwesens. Waadt. (1858). *Pädagogische Monatsschrift für die Schweiz*, 3(6), 285–292.
- Nachlese von 1837. (1838). *Appenzellisches Monatsblatt*, 14(2), 27–32.
- Nassehi, A. (2019). *Muster: Theorie der digitalen Gesellschaft*. C.H. Beck.
- Nathaus, K. (2012). Sozialgeschichte und Historische Sozialwissenschaft. *Docupedia-Zeitgeschichte*. <https://doi.org/10.14765/ZZF.DOK.2.268.V1>
- OpenAI. (2023). *ChatGPT* (Version March 23). <https://chat.openai.com/>
- Osterhammel, J. (2010). *Die Verwandlung der Welt: eine Geschichte des 19. Jahrhunderts*. C.H. Beck.
- Pfister, C. (1995). «Uss gwüssen Ursachen»: Hintergründe und Methoden statistischer Erhebungen im Kanton Bern 1528–1928. *Schweizerische Zeitschrift für Geschichte*, 45(1), 29–50. <https://doi.org/10.5169/SEALS-81119>
- Porter, T. M. (1986). *The Rise of Statistical Thinking, 1820–1900*. Princeton University Press.
- Putnam, L. (2016). The transnational and the text-searchable: Digitized sources and the shadows they cast. *The American Historical Review*, 121(2), 377–402. <https://doi.org/10.1093/ahr/121.2.377>
- Ramisch, P. (2022). Open Discourse - eine Quellenkritik. Verfügbar 25. Mai 2023 unter [https://paulramisch.de/opendiscourse/beta/1\\_einleitung.html](https://paulramisch.de/opendiscourse/beta/1_einleitung.html)
- Rothen, M., & Ruloff, M. (2014). Die vergessenen Schulumfragen der Helvetischen Republik. In D. Tröhler (Hrsg.), *Volksschule um 1800. Studien im Umfeld der Helvetischen Stapfer-Enquête 1799* (S. 33–54). J. Klinkhardt.
- Rothenbühler, V. (2010). Eine neue Schule – ein neuer Staat? Diskussionen um die Volksschule, 1820–1850. In S. G. Gesellschaft (Hrsg.), *Freiwillig verpflichtet. Gemeinnütziges Denken und Handeln in der Schweiz seit 1800*. NZZ Libro.
- Ruoss, T. (2018). *Zahlen, Zählen und Erzählen in der Bildungspolitik: lokale Statistik, politische Praxis und die Entwicklung städtischer Schulen zwischen 1890 und 1930*. Chronos.
- Sarasin, P. (2012). Sozialgeschichte vs. Foucault im Google Books Ngram Viewer. Ein alter Streitfall in einem neuen Tool. In P. Maeder, B. Lüthi & T. Mergel (Hrsg.), *Wozu noch Sozialgeschichte? Eine Disziplin im Umbruch* (S. 151–174). V&R unipress.
- Scandola, P. (1991). Von der Standesschule zur Staatsschule. Die Entwicklung des Schulwesens in der Schweizerischen Eidgenossenschaft 1750–1830 am Beispiel der Kantone Bern und Zürich. In N. L. Dodde & W. Schmale (Hrsg.), *Revolution des Wissens? Europa und seine Schulen im Zeitalter der Aufklärung (1750–1825) ein Handbuch zur europäischen Schulgeschichte* (S. 581–626). D. Winkler.

- Sch., J. J. (1856). Mittheilungen über den Zustand und die Entwicklung des schweizerischen Schulwesens. St. Gallen. *Pädagogische Monatsschrift für die Schweiz*, 1(12), 364–368.
- Scharloth, J. (2018). Korpuslinguistik für sozial- und kulturanalytische Fragestellungen. In M. Kupietz & T. Schmidt (Hrsg.), *Korpuslinguistik* (S. 61–80). De Gruyter.
- Schiller, A., Teufel, S., & Stöckert, C. (1999). Guidelines für das Tagging deutscher Textcorpora mit STTS. Verfügbar 2. Juni 2023 unter <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>
- Schmidt, H. R., Messerli, A., Osterwalder, F., & Tröhler, D. (Hrsg.). (2015). Die Stapfer-Enquête. Edition der helvetischen Schulumfrage von 1799. Verfügbar 19. April 2023 unter <https://www.stapferenquete.ch/>
- Schulnachrichten. Thurgau. (1866). *Schweizerische Lehrerzeitung*, 11(3), 23–24.
- Schulvisitationen. (1855). *Appenzellische Jahrbücher*, 2(8), 268–281.
- Schwandt, S. (2018). Digitale Methoden für die Historische Semantik: Auf den Spuren von Begriffen in digitalen Korpora. *Geschichte und Gesellschaft*, 44(1), 107–134. <https://doi.org/10.13109/gege.2018.44.1.107>
- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Benjamins.
- Tröhler, D. (2007). Die Zürcher Schulsynode. Ein demokratisches Kuckucksei in der liberalen Ära Zürichs im 19. Jahrhundert. In C. Crotti, P. Gonon & W. Herzog (Hrsg.), *Pädagogik und Politik. Historische und aktuelle Perspektiven*. (S. 69–88). Haupt.
- Tröhler, D. (2011). Classical Republicanism, Local Democracy, and Education. The Emergence of the Public School of the Republic of Zurich, 1770–1870. In D. Tröhler, T. S. Popkewitz & D. F. Labaree (Hrsg.), *Schooling and the Making of Citizens in the Long Nineteenth Century. Comparative Visions* (S. 153–176). Routledge.
- Tröhler, D. (2014). *Volksschule um 1800. Studien im Umfeld der Helvetischen Stapfer-Enquête 1799*. J. Klinkhardt.
- Vanderstraeten, R. (2006). Soziale Beobachtungsraster: Eine wissenssoziologische Analyse statistischer Klassifikationsschemata. *Zeitschrift für Soziologie*, 35(3), 193–211. <https://doi.org/10.1515/zfsoz-2006-0302>
- Verschiedenes. (1864). *Neue Berner Schul-Zeitung*, 7(36), 144.
- Woolf, S. (1989). Statistics and the Modern State. *Comparative Studies in Science and History*, 31(3), 588–604.

# Abbildungsverzeichnis

4.1	Relative Häufigkeit für [word="durchschnitt durchschnittl.*"%c] . . . . .	32
4.2	Relative Häufigkeit für [word=".*statist.*"%c] . . . . .	34
5.1	Relative Häufigkeit für ([pos="CARD"] [lemma=".schüler .*schülerin -Knabe Mädchen Tochter Kind Schüler Schülerin Bube"]) ([lemma=".schüler .*schülerin Knabe Mädchen Tochter Kind Schüler -Schülerin Bube"] [pos= "CARD"] ) . . . . .	37
5.2	Schulstatistik des Kantons Bern 1838/39 aus «Kanton Bern», 1840, S. 356 . . . . .	38
5.3	Sekundarschulstatistik Thurgau 1864/65 aus «Schulnachrichten. Thurgau», 1866, S. 24. . . . .	39
5.4	Auszug aus «Verschiedenes», 1864. . . . .	42
5.5	Relative Häufigkeiten für [word="Absenztabellen Absenzenlisten Absenzenverzeichniß Absenzenverzeichnisse Absenzenlistcn Absenzen-tabellen Versäumnißrodel Versäumnißtabelle Absenzenliste Absenzcnrapporte Absenzen-Verzeichnisse Absenzenberichten Schul-listen Schultabellen Versäumnißtabellen"] . . . . .	44
5.6	Schulabsenzen in Appenzell 1837 aus «Nachlese», 1838, S. 29. . . . .	45
5.7	Relative Häufigkeiten für [word="Absenz.* Versäumniß"%cd] . . . . .	45
5.8	Schulabsenzen im Oberaargau 1864 aus «Die bernische Schule auf der Anklagebank (V.)», 1864, S. 184. . . . .	46
5.9	Schulabsenzen im Kanton Zürich 1856–1858 aus «Jahresbericht der h. Direktion des Erziehungswesens über den Zustand des zürcherischen Erziehungswesens im Schuljahre 1857–58», 1858, S. 24. . . . .	47

# Tabellenverzeichnis

2.1	Beispielantwort der Stapfer-Enquête für Nr. 165, Oberembrach, Schmidt et al., 2015. Verfügbar 12. Februar 2023 unter <a href="http://www.stapferenquete.ch/db/165">http://www.stapferenquete.ch/db/165</a> .	12
4.1	Grösse der Subkorpora nach Dekaden . . . . .	25
4.2	Keywords im Surbkorpus 1835–1870 im Vergleich zu 1800–1834 . . . . .	27
4.3	Treffer für Einträge aus dem Wörterbuch «Quantifizierung» von Buchner et al., 2020, Frequenz > 40 . . . . .	29
4.4	KWIC-Ansicht für [word="durchschnitten"%c] . . . . .	30
4.5	KWIC-Ansicht für [word="Zählung"%c] . . . . .	31
4.6	Ergebnisse für [word=".*statist.*"%c] (Frequenz > 3). . . . .	33
5.1	Kollokationen zu [word=".*tabell.* .*statist.*"] gefiltert nach finiten Voll- verben ([pos="VVFIN"]) . . . . .	40
5.2	KWIC-Ansicht für <i>zeigt</i> im Zusammenhang mit [word=".*tabell.* .*sta- tist.*"] . . . . .	41
5.3	Auswahl aus Treffern für [word=".*tabell.*"%c] . . . . .	43
5.4	Auswahl aus Treffern für [word="Versäumniß.* Absenz.*"] . . . . .	43