



ZHistLex-Papiere

2020

Einige neue explorative Visualisierungsformen in der digitalen historischen Lexikologie und Lexikographie

Armin Hoenen

Vollständige Zitation:

Armin Hoenen. 2020. *Einige neue explorative Visualisierungsformen in der digitalen historischen Lexikologie und Lexikographie*. ZHistLex-Papiere. https://zhistlex.de/papiere/hoenen_2020_visualisierungsformen_ZHistLex.pdf

ZHistLex wurde gefördert durch das



Bundesministerium
für Bildung
und Forschung

Einige neue explorative Visualisierungsformen in der digitalen historischen Lexikologie und Lexikographie

Ziel dieses Papiers ist es, neue Visualisierungsformen für die historische Lexikologie und Lexikographie im digitalen Raum zu erkennen und auch einige eigene zu konzipieren. Zunächst wird der Forschungsstand analysierend wiedergegeben, woraufhin Aufgaben der Visualisierung im lexikographischen Zusammenhang strukturiert dargestellt werden. Schließlich werden eigene Ansätze zur Umsetzung dieser Aufgaben, also eigene Visualisierungsvorschläge vorgestellt.

1 Aufgaben und Ziele von Visualisierungsformen in der digitalen, historischen Lexikologie und Lexikographie

Visualisierungen können im digitalen Medium (im Gegensatz zum Printmedium) kostengünstig und dynamisch (Parametersetzung, Interaktionsmöglichkeiten) gestaltet werden. Es lässt sich ein gegenwärtiger Trend für digitale Wörterbücher und Wörterbuchsysteme konstatieren, Visualisierungen standardmäßig miteinzubinden. Dabei stellt sich einerseits die Frage nach den Aufgaben und Möglichkeiten des Einsatzes solcher Visualisierungen, sowie die nach geeigneten Mitteln und Visualisierungsformen, diese abzubilden.

1.1 Forschungsstand

Wortgeschichte und Wortgeschichten sind im Fokus einiger Felder der Computerlinguistik mit der Visualisierung in Kontakt geraten. Das vielleicht bekannteste Beispiel stellen sog. Wortvektorräume und deren Visualisierung dar. Seit Verfügbarkeit von neuronalen Netzwerken zur Erzeugung sog. Wortvektorräume (vgl. [Mikolov et al. \(2013\)](#); [Pennington et al. \(2014\)](#)) wurde mehrfach die semantische Repräsentativität selbiger hervorgehoben [Mikolov et al. \(2013\)](#). Ein n-dimensionaler Raum repräsentiert dabei die virtuelle Umgebung in der jedes Wort eines Korpus mit Koordinaten verortet wird. Die Abstände und Anordnungen der Worte im Raum korrelieren mit der Semantik. Saussure's paradigmatische sowie syntagmatische Ähnlichkeiten fließen beide in die Berechnung der Vektorpositionierung mit ein. Eine Reduktion auf zwei Dimensionen und ein darauffolgendes Plotten des so reduzierten Vektorraums oder von Teilen davon hat sich als beliebte Visualisierung zumindest in computerlinguistischen Kreisen etabliert, vermag jedoch zunächst nur jeweils einen Vektorraum, der auf einem Korpus basiert, abzubilden. Die Dimension Zeit ist zunächst nicht eindeutig in diesen Plots erkennbar.

[Hamilton et al. \(2016\)](#) untersuchen nun die "Drift" einzelner Lexeme in Vektorräumen über die Zeit hinweg indem sie für Berechnung und Visualisierung an 2 durch geschickte Interpolation konstant gehaltenen Achsen ausrichten, siehe [Abb. 1](#). Mit diesem Verfahren gelingt eine Visualisierung als zweidimensionaler Weg ([Zhang et al., 2015](#), S.307/311) nutzen sogar 3 Dimensionen) durch den so fixierten Wortvektorraum.

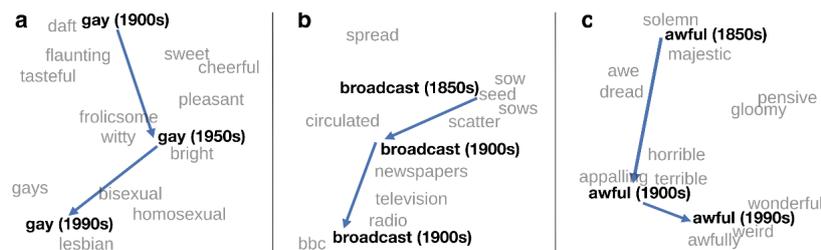


Abbildung 1: Hier zeigen [Hamilton et al. \(2016\)](#) Wege der semantischen Entwicklung dreier Lexeme durch deren enge Nachbarn in zeitlich abgestuften Vektorräumen.

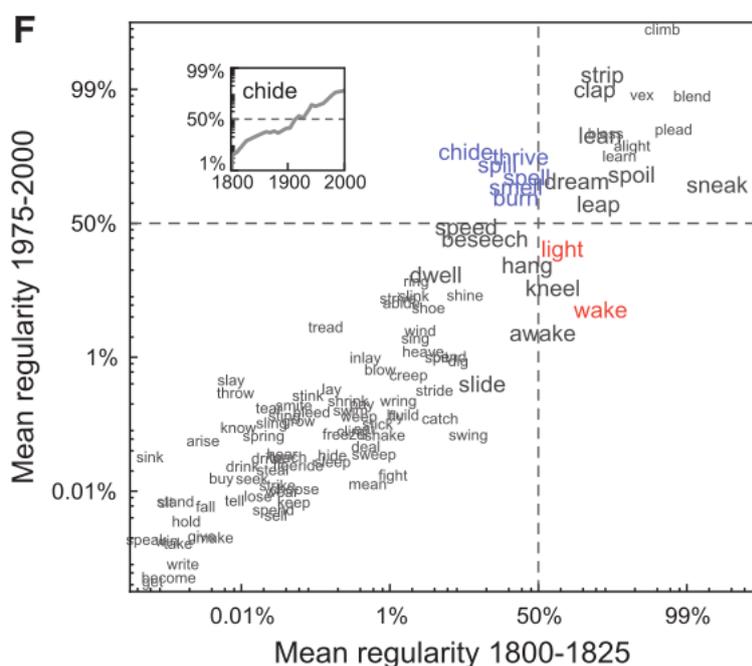


Abbildung 2: 2-d Plot mit Zeiträumen auf den Achsen.

Voraussetzung für das Erzeugen einer solchen Visualisierung ist die Implementation der vorgeschlagenen Methode. Allerdings sind zudem größere Datenmengen als Input zumindest wünschenswert und Evaluationswerte lassen ggf. je nach Frequenzen eine nicht zu marginalisierende Fehlerrate plausibel erscheinen.

1.2 Zweidimensionale Darstellungen

[Michel et al. \(2011, Figure 2 F\)](#) nutzen zwei Zeiträume als Achsen und einen vergleichbaren Regularitätswert, der angibt wie sehr ein englisches Verb regulär flektiert wird, siehe Abb. 2.

Die Achsenwerte sollten lexikologisch relevanten Größen entsprechen. Solche Visualisierungen zielen auf spezifische linguistische Phänomene ab, die teils sprachspezifisch sind. Die visuelle Nutzung zweier Achsen mit verschiedenen Zeiträumen lässt sich jedoch unabhängig auf viele solche Vergleiche übertragen.

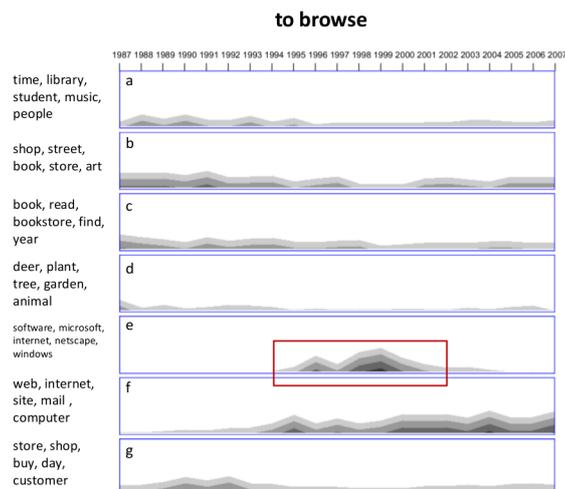


Abbildung 5: Belegfrequenz-basierte Darstellung für Bedeutungen des Neologismus *browse*.

Die Ableitungswege eines Lexems lassen sich ebenfalls sprachgeschichtlich mit Rückbezug auf Lautwandelphänomene darstellen. Eine vergleichende baumartige Darstellung skizziert [Hoenen \(2018a\)](#), siehe Abb. 4, wobei noch zusätzliche visuelle Parameter wie Farbe zur Markierung der Sprache/Sprachstufe genutzt werden könnten.

1.4 Zeitleisten und Topic Models

Ein weiterer Block an Visualisierungen für die Wortentwicklung kann unter dem Stichwort Zeitleisten zusammengefasst werden, wobei hier (nur) eine Achse, meist die X-Achse den zeitlichen Verlauf abbildet. Generell ist die Zeit als Achse aufeinanderfolgender Elemente besonders gut visualisierbar. Ein Beispiel einer klassischen und erweiterten klassischen Timeline wird in 2.1.1 genauer beschrieben. [Rohrdantz et al. \(2011\)](#) präsentieren u.a. eine nach Topics eines Topic Models induzierte mehrbändige Entwicklungskurve, siehe Abb. 5. Topics sind Wortgruppen, die nach statistischen Gesichtspunkten besonders konsistent in einem Corpus zusammen oder in ähnlichen Umgebungen auftreten. Topics, in denen der Term ‘browsen’ vorkommt, werden von [Rohrdantz et al. \(2011\)](#) mit dessen Bedeutungen gleichgesetzt. Die gemittelte relative Frequenz der Topicwörter in zeitlichen Abschnitten des Corpus zeigt in Abb. 5 das Aufkommen der technischen Bedeutung des “Browsens im Internet”.

Diese Visualisierung kann ohne manuelle Annotationen erzeugt werden, allerdings benötigt ein Topic Model eine initiale Anzahl an topics (hier: senses) für ein zu visualisierendes Lexem. Anzahlen könnten z.B. aus vorhandenen Wörterbuchartikeln ableitbar sein oder iterativ optimiert werden.

1.5 Weitere Visualisierungen

(Kollokationsbasierte) Wortwolken, Kreisdiagramme, Balkendiagramme, Kurven usw. sind allgemein bekannte Visualisierungen, die tw. bereits umgesetzt werden, z.B. im Digitalen Wörterbuch der Deutschen Sprache (DWDS). Auch kartenbasierte Visualisierungen bei denen Belegpunkte auf den Belegort projiziert werden sind verbreitet.

Im Folgenden soll auf die Aufgaben einer Visualisierung im Kontext der digitalen Lexikographie eingegangen werden.

1.6 Aufgaben

Visualisierungen können verschiedene Aufgaben erfüllen und die konkrete Aufgabe einer Visualisierung bestimmt ihr Design maßgeblich. Als übergeordnetes Ziel einer Visualisierung mag man festhalten wollen, dass durch sie ein positiver Mehrwert für den Benutzer entstehen soll. Welche umsetzbaren Aufgaben sind konkret für Visualisierungen im lexikologischen und lexikographischen Rahmen denkbar? Die folgende Aufzählung ist selbstverständlich nicht erschöpfend, manche ihrer Einträge überschneiden sich thematisch oder beziehen sich auf dieselben Beispiele:

1. **Informationsübertragung:** Visualisierungen können dazu dienen, Informationen schneller verstehbar zu machen. Text als Format ist an die sequenzielle Aneinanderreihung von Morphemen gebunden und kann nur durch anaphorische Rückreferenzen u. dgl. Querbezüge ausdrücken, vgl. auch [Ong \(2013\)](#); [Richter \(2014\)](#). Welche relevanten Informationen aber können visualisiert werden?
 - **Belegdaten** und **Lexemfrequenzen** treten als quantitative Daten in lexikologischen Artikeln auf und sind aus den Primärquellen extrahierbar. Der Mehrwert einer frequenzbasierten Visualisierung kann u.a. darin liegen, dass die lückenlose Beleglage sichtbar wird (während der Text sich auf das Wesentliche beschränken kann).
 - Metadaten, die während der lexikologischen Arbeit entstanden sind, können Quelle weiterer Visualisierungen sein, und z.B. Fragen beantworten wie *Welche Primärquellen wurden wo genutzt? Welche Jahre waren bei der Erstellung eines Lexikons besonders produktiv? Wie hoch ist der Anteil der Onlineverweise in welchen Artikelbereichen?* usw.
2. **Datenexploration:** Noch nicht erschlossene oder nicht manuell erschließbare Bereiche/Strukturen im wahrsten Sinne des Wortes 'sichtbar' zu machen, kann Aufgabe einer Visualisierung sein (man denke auch an Distant Reading ([Moretti, 2013](#))).
 - Kollokationen eines Wortes bieten eine Grundlage für die ggf. auch zeitlich kontrastierende Darstellung der Umgebungen eines Lexems.
 - Wortvektorräume können als Repräsentation semantischer Räume verstanden werden und entsprechend als Visualisierung Einblicke in die Gesamtstruktur eines solchen Raumes erlauben.

- Ontologien verbinden hierarchisch Entitäten, die auch auf Lexeme abbilden, wobei Polisemie, wie auch im obigen Fall zu berücksichtigen ist. WordNets erlauben die hierarchische Organisation auch von Sprachstufen zu codifizieren und entsprechend zu visualisieren. Die Abdeckung von Wordnets ist dabei ebenfalls zu berücksichtigen.
 - Topics als Gruppen konsistent über ein Dokumentkorpus distributionell ähnlich auftauchender Begriffsagglomerationen (wobei jeder Begriff mehreren Topics angehören kann) können in ihrem Verlauf über den Belegzeitraum hinweg visualisiert werden (Heyer et al., 2018) oder sogar in Verbindung mit dem Bedeutungsspektrum eines Einzellexems (Rohrdantz et al., 2011).
 - Belegorte können auf einer Karte dargestellt werden. Besonders in der Dialektologie erfreuen sich kartenbasierte Visualisierungen großer Beliebtheit und Verbreitung.
3. **Bestätigung lexikologischer Erkenntnisse:** In der Visualisierungsforschung werden Visualisierungen oft funktional als konfirmativ (i.Gs. zu explorativ u.a., siehe auch Rehbein (2017)) bezeichnet, und zwar genau dann, wenn sie eine bestimmte Erkenntnis anhand geeigneter visueller Mittel zu exemplifizieren im Stande sind. So wäre es denkbar, die Existenz eines Phänomens 'Pejorativisierung' visuell zu demonstrieren. Hierzu zählen aber auch Strukturen und Zusammenhänge des Wortschatzes, wie die Wortfamilienstruktur eines Lexems, die z.B. durch Graphen visualisiert werden kann.
 4. **Multimodale Darbietung des Materials:** Obgleich Visualisierungen vor allem Mehrwerte erzeugen sollen, kann ihre Redundanz mit dem Textteil zu einer besseren Merkbarkeit der dargebotenen Information führen.
 5. **Lexikonüberblick:** Viele statistische Gegebenheiten können einen Überblick über die Zusammensetzung, Vernetzung und Entstehung eines Lexikons erlauben. Insofern besteht eine Aufgabe der Visualisierung darin, Transparenz über den Prozess der Lexikongenese und einen Überblick über dessen Inhalt zu ermöglichen. Man denke z.B. an eine Visualisierung der Anzahl Artikel pro Buchstabe.

Mit Kollokationen, Wortvektorräumen, Ontologien und Topics sind hier viele in der klassischen präcomputationalen Lexikologie unbekannte oder ungenutzte Visualisierungsquellen genannt. Da die Computerlinguistik eine stark quantitativ durchdrungene Wissenschaft ist verwundert die gute Visualisierbarkeit von aus dieser Disziplin (oder der Korpuslinguistik) erwachsenen Datenkonstrukten kaum. Ein Problem ergibt sich daraus nur dann, wenn das Grundmaterial (Lexikon) z.B. in einer Zeit einstanden ist oder wesentliche Teile aus dieser Zeit enthält, welches die Extraktion oder Generierung solcher Daten nicht erlaubt. Für Kollokationen, Wortvektorräume und Topicmodels sind die Belegkorpora als Datenquelle nötig, Ontologien oder WordNets sollten als Ressource vorliegen, bevor eine Visualisierung aufsetzen kann.

Eine Visualisierung ist, es sei denn künstlerisch, an eine numerische Komponente gebunden. Die Aufgabe der Visualisierung für Lexika im eigentlichen Sinne kann aber das Vorhandensein des Belegkorpus oder eine Verknüpfung damit nicht automatisch voraussetzen. Wenn dieses vorhanden ist, dann lassen sich eine Vielzahl von Aufgaben aus der modernen Korpuslinguistik ableiten. Ist dies nicht der Fall, bleiben weniger Alternativen.

Die eher klassisch genutzten quantitativen Datenpunkte sind neben Belegen und Beleghäufigkeiten, Datumsangaben, Eigennamen und Ereignisse, Belegkontexte, Beziehungen zu anderen Teilen des Wortschatzes oder anderen Sprachen, sowie die Kontextualisierung der

Belegquelle (Themengebiet, Funktion, intendierte Leserschaft etc.) uvm. Dieses Material ist ggf. sehr heterogen und einzelwerkspezifisch und eröffnet die Möglichkeit individueller Visualisierungen je nach Lexikon/Lexem.

Eine wichtige Stellung kommt den Belegen und Belegfrequenzen bei der Visualisierung zu, da sie an der Schnittstelle klassisch lexikologischer und lexikographischer Arbeit mit der Korpuslinguistik anzusiedeln sind. Für Hapax legomena und Dislegomena ist eine solche Visualisierung jedoch kaum sinnvoll, obwohl sie u.U. große relative Anteile der Types eines Korpus ausmachen.

Eine weitere interessante Frage ist die der Automatisierbarkeit. Dabei ist mit automatisch zu visualisierenden Datenlagen zu rechnen (z.B. Frequenzen, wenn ein Korpus vorliegt), mit solchen, welche menschliche Intervention erfordern (z.B. manuelle Annotation bestimmter Gegebenheiten) sowie mit solchen, die nicht ohne signifikante menschliche Beteiligung entstehen können (z.B. visuelle Metaphorik).

2 Abbildung auf Formen der Visualisierung

Für das oben skizzierte Portfolio an Aufgaben der Visualisierung lassen sich nun bestimmte konkrete Visualisierungsformen definieren. Allerdings ist die Anzahl möglicher Visualisierungen sehr groß. Visualisierung ist ein Prozess, der u.a den Schritt der Auswahl der zu visualisierenden Daten aus einer größeren Datenmenge impliziert und zum anderen müssen die Visualisierungsparameter visuell/interaktiv abgebildet werden, so z.B. Frequenz auf Balkenhöhe, Zeitpunkte auf horizontale Position, relative Frequenzen auf prozentuale Anteile einer Fläche usw. Der Reduktionsschritt sowie der Abbildungsschritt implizieren bereits kombinatorisch eine Fülle an Möglichkeiten und müssen selbst noch miteinander verkettet werden. Schlechte Abbildungen können zu schlechten Visualisierungen führen mit denen keines der o.g. Ziele zu erreichen ist. Eine besondere Gefahr ist hierbei das visuelle Überladen, siehe auch [Tufte \(1983\)](#); [Shneiderman \(1996\)](#).

2.1 Lexikologie

Für die Lexikologie steht das Einzellexem bzw. stehen Bedeutungspositionen und deren Verflechtungen im Vordergrund. [Gould \(1993\)](#) konstatiert besonders für die Geistes- und Sozialwissenschaften stark nicht visuelle Modi und besonders Printlexika bzw. einzelne Artikel darin enthalten selten Visualisierungen.

2.1.1 Text-dominante Visualisierungen

[Richter \(2014, p.180\)](#) unterscheidet Visualisierungen danach ob sie 'language-like' oder 'picture-like' sind. Da die Schrift selber ein primär visuelles Korrelat der Sprache ist, ist eine Grenze zwischen Schrift und Visualisierung zu ziehen nicht so trivial, wie es erscheinen mag. Die oben angesprochene Sequenzialität zeichnet zwar Sprache aus, kann aber im Zusammen-

hang eines Bildes durchaus als gewollt sinntragendes nicht primär sprachliches Designelement aufgefasst werden.

Der Leseprozess ist neuronal eine erlernbare Aktivität, welche sich als solche von anderen visuellen Erfassungsprozessen abgrenzen lässt. Interpunktion und visuelle Textanordnung (Hoenen, 2018b) exemplifiziert wie schwer es sein kann, Schrift und Visualisierung qualitativ zu trennen. Vor allem aber sind es höhere strukturelle Anordnungen von Text, die mit nicht viel mehr visuellen Mitteln als Interpunktion und Textanordnung arbeiten, die ohne in der Visualisierungsforschung als solche behandelt zu werden, bereits zumindest visuell getriebene Modi der Darstellung sind, so z.B. Tabellen. Sofern eine Aufzählung als Fließtext gilt oder im Falle, dass ganze Sätze in Tabellenfeldern stehen, ist eine Tabelle kaum noch strikt von Poemen oder anderen graphisch arrangierten Texten zu unterscheiden. Sie kann beispielsweise, muss aber nicht, streng sequenziell gelesen werden. Vielmehr können Sprünge von bestimmten in nicht benachbarte andere Zellen stattfinden. Da auch im normalen Leseprozess Sprünge ständig stattfinden, führen diese nicht dazu, dass der neuronale Prozess *Schriftlesen* (worunter auch das Lesen von Aufzählungen fällt) beim Betrachten einer Tabelle unterbrochen werden muss.¹

An dieser Stelle sollen daher Tabellen und weitere Visualisierungen, deren primärer Zugang das Lesen ist als Visualisierungen in die Betrachtung integriert werden. Solche werden und wurden in der Lexikologie eingesetzt, vgl. (Splett, 2002), während eher bildliche Visualisierungen wie Belegkurven im Vergleich mit dem digitalen Medium eher selten anzutreffen sind.²

Verschiedene graphische Elemente, die für eine reine schriftliche Repräsentation von Sprache nicht nötig wären und teilweise erst spät entstanden, werden oder wurden also visualisierend eingesetzt: Textfarbe und Unterlegung, Unterstreichung (Durchstreichung), Font, Spatien, Gross - Kleinbuchstaben, Schriftgröße, Abkürzungstechniken, Interpunktionszeichen, Schrift-richtung, Hoch- und Tiefstellung, Zeilenumbrüche und Textanordnung auf einer Seite, Symbole und graphische Formen (Dreieck, Kreis etc.). Diese Elemente unterliegen oft genauer Planung, werden aber i.d.R. nicht als Visualisierung erfasst. Ein historisches Beispiel sind bedeutungstragende Anordnungen von zwischenspaltigen Worten im Notker, siehe Abb. 7, wobei die Tabelle in all ihren Formen der prävalenteste Typ text-dominanter Visualisierungen bleibt, vgl. Abb. 6.³

Erkennt man alle oder doch einige der o.g. Elemente als Visualisierungen an, hat die klassische Lexikologie mit den verschiedenen Strukturen der Wörterbuchartikel und intentionalen Anordnungen (z.B. in zwei Spalten pro Seite) eine starke visualisierende Komponente text-dominanter Visualisierung herausgebildet, die es gilt im Digitalen umzusetzen und mit Interaktionsmöglichkeiten anzureichern. So kann eine rein alphabetische Anordnung je nach Nutzerpräferenz z.B. auf Knopfdruck in eine nach Wortklassen überführt werden.

¹Im Digitalen trägt zur Beliebtheit von Wortwolken vielleicht gerade dieser Aspekt bei.

²Die Einteilung ist vergleichbar mit der von Richter (2014) 'language-like' genannten Visualisierung, aber nicht struktur, sondern prozessbasiert und teilt in Beantwortung der Frage: *Ist das Lesen (Sprachlesen) der primäre Zugang zur Visualisierung* ein.

³Online unter <http://zhistlex.de/materialien/zemanTabelle.html>

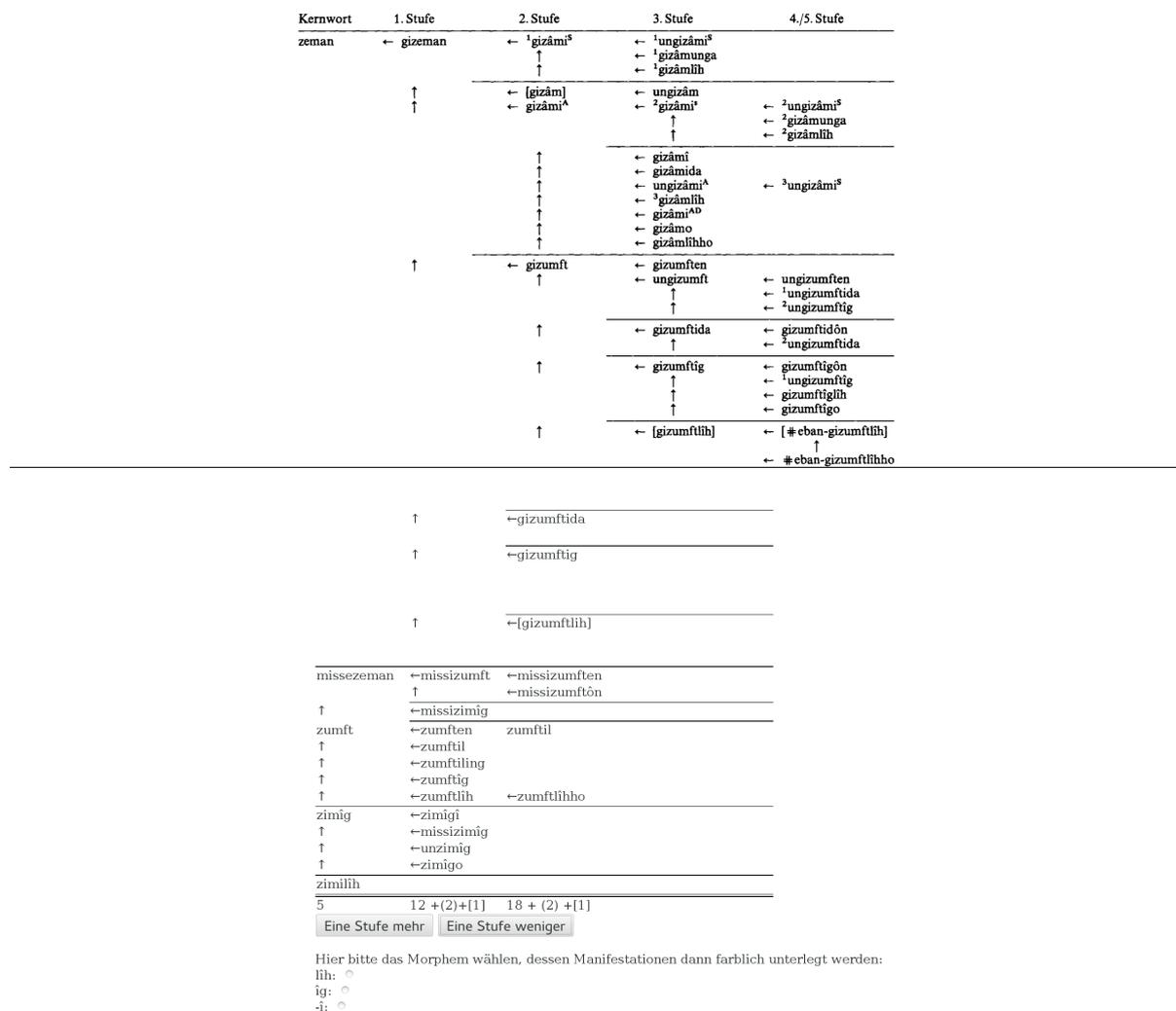


Abbildung 6: Tabelle von J. Splett aus dem Handbuch Lexikologie, welche Wortfamilienstrukturen darstellt und digitales Korrelat, das u.a. dynamische Aus- und Einblendfunktionen implementiert.

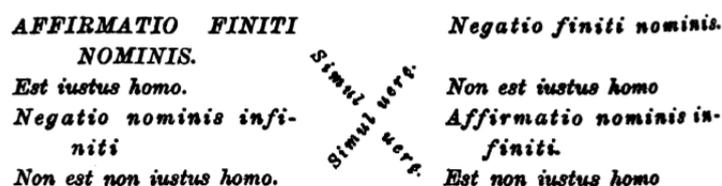


Abbildung 7: Historisches Beispiel mit Textanordnung, die linienhaft Verbindungen herstellt (Piper et al., 1883, S.540). Handelt es sich hier um eine Visualisierung?

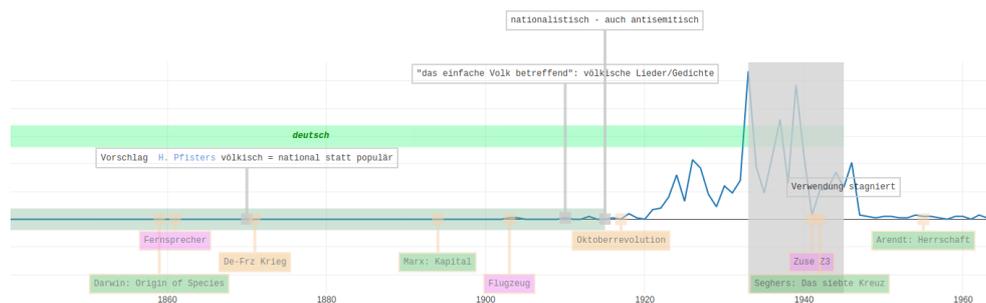


Abbildung 8: Timeline zur narrativen Wortgeschichte von ‘völkisch’, frei nach der Analyse des Lexems durch V. Harm. Die Darstellung ist im Original zoombar und Links zu Belegen in den Wörterbüchern klickbar.

2.1.2 Vorgeschlagene Visualisierungen

Im Rahmen des ZHistLex Projektes sind Visualisierungen entstanden, für die z.T. digitale Prototypen ausprogrammiert und publiziert wurden, siehe (Hoenen, 2019). Belegverteilungskurven wurden bereits eingangs genannt und werden in digitalen Wörterbuchsystemen eingesetzt. Da die X-Achse einer Verteilungskurve die Dimension Zeit abbildet, da des Weiteren aus der Belegkurve heraus allein die relevante Entwicklung nicht ersichtlich ist, bietet sich eine Ergänzung der Belegkurve um historische Ereignisse und Perioden an. Die X-Achse fungiert sozusagen als Timeline. Besonders gut lässt sich diese Visualisierung mit dem Konzept der narrativen Wortgeschichten verbinden, bei dem einzelne Stationen und Eigennamen, die meist mit konkreten Daten in Verbindung stehen, wichtige Elemente sind. In Abb. 8 ist das Ergebnis der technischen Umsetzung zu sehen.⁴ Über der Belegkurve-Timeline zeigen Querbalken Synonyme an (falls vorhanden) - paradigmatisch. Unten werden Ereignisse aufgezeigt, die in 3 Bereiche gegliedert sind und die separat ein- und ausgeblendet werden können: *Wissenschaft*, *Gesellschaft* und *Katastrophen und Kriege*. Diese Ereignisse beruhen auf Listen großer Erfindungen, Wetterereignissen und Publikationen. 563 Masterevents wurden aus entsprechenden Listen der Wikipedia extrahiert und manuelle nach Relevanz für Deutschland/Europa gefiltert. Die Events verteilen sich auf einen Zeitraum vom Jahr 98 bis zum Jahr 2003 und stellen nur eine beispielhafte Basis allgemeiner kontextualisierender Ereignisse dar. Technisch wurde die Timeline+ mit dem Javascript Framework plotly⁵ umgesetzt.

Eine Auszeichnung der narrativen Wortartikel z.B. in einem XML Format würde eine automatische Generierung ermöglichen. Dabei sollten im Artikel genannte Ereignisse, Zeitpunkte und Eigennamen, sowie Metainformationen wie Themengebiete zu denen das Lexem gehört (Politik, Wissenschaft, etc.) annotiert sein.

Belegkurven und Timeline-Belegkurven für mehrere Lexeme können vergleichend kombiniert werden, dabei können aber Skalierungsprobleme entstehen. Schliesslich könnten Karten des semantischen Raums angefertigt werden.

Als weitere Darstellung könnte eine sog. ‘stacked area’ erzeugt werden. Diese ähnelt entfernt der *Histomap* McNally’s, einem sehr bekannten und weit verbreiteten Vorläufer, wurde aber tatsächlich unabhängig von ihr entwickelt, siehe Abb. 9. Die X-Achse steht für die Zeit, die

⁴Online unter <https://zhistlex.de/timeline/>

⁵<https://plot.ly/>

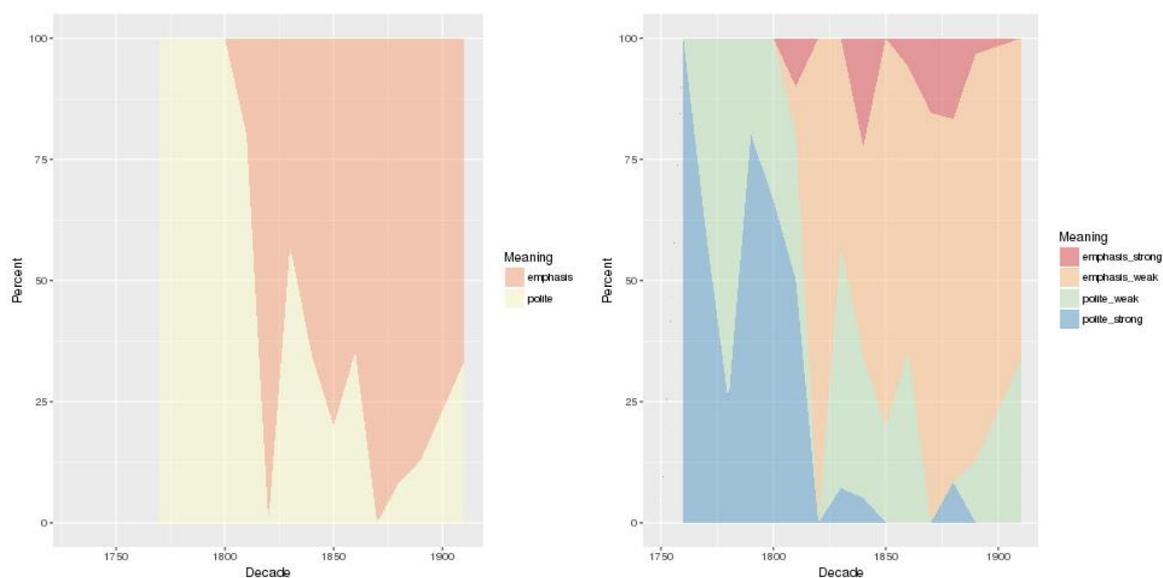


Abbildung 9: Stacked area plot für 'gefälligst', frei nach der Analyse des Lexems durch T. Gloning.

Y-Achse für die relative Häufigkeit (einer Bedeutung). Einzelne Bedeutungen/Lesarten werden als Farbflächen markiert. Die absolute Frequenz des Lexems (pro Jahr) wird nicht visualisiert, zu jedem beliebigen Zeitpunkt der X-Achse (Zeit) summieren sich alle Werte zu 100% Belegmaterial und die Größe der jeweiligen Gebiete zeigt den prozentualen Anteil der Belegstellen des Lexems an, an denen es eine bestimmte Bedeutung trägt. Im abgebildeten Beispiel wurden alle ca. 200 DTA Belege des Lexems 'gefälligst' dahingehend annotiert, ob sie höflich bittend (gefägend, ältere Lesart) oder in ihrer Nachdruckslesart (den früheren Gebrauch ironisierend, auffordernd) verwendet wurden. Die linke Graphik unterscheidet nur diese 2 Lesarten, während die rechte 4 Kategorien (stark höflich, schwach höflich, stark nachdrücklich, schwach nachdrücklich) erhebt. Es werden 2 Farbpolaritäten (kalt und warm) genutzt. In den Graphiken zeichnet sich deutlich ein Dominant-Werden der Nachdruckslesart ab, wobei der Verlauf nicht strikt linear ist, sondern Reflexe der älteren Lesart immer wieder deutlich Platz einnehmen. Ebenso scheint sich die stark nachdrückliche Lesart zu verhalten. Ob es sich bei dem Muster um die Reflektion des tatsächlichen Sprachwandels handelt oder ob Artefakte der Auszeichnung, des Samplebias, usw. dazu führen, liesse sich nur in einer umfassenderen Studie ermitteln. Will man ein 'umkippen' der Hauptlesart des Lexems annehmen, deutet die Visualisierung auf die Zeit um die Mitte des 19. Jahrhunderts hin.

Bei vorhandener Annotation (oder automatischer Disambiguation) der Bedeutungen konkreter Belege jeglicher Lexeme könnte diese Art Visualisierung erzeugt werden.

2.2 Lexikographie

Für die Lexikographie ergeben sich vor allem auf den Wörterbuchinhaltsstatistiken, den Quellennachweisen und den Gebrauchsstatistiken beruhende vielfältige Möglichkeiten der Visualisierung. Die Anzahl der Belege pro Jahr, die pro Jahr konsultierten Quellen, ein Graph der gegenseitigen Verknüpfungen innerhalb der Quellen und Artikel, das alles sind denkbare lexi-

kographische Visualisierungsformen. Die Lexikologie steht besonders bei nicht akademischen Nutzern jedoch sicherlich im Vordergrund. Demgegenüber stehen die relativ leicht aus den momentan vorliegenden Daten heraus zu gewinnenden lexikographischen Visualisierungen vor allem bzgl. des Wörterbuchvergleichs.

2.3 Individuelle Visualisierungen

Eine Komplexitätsschranke für die Wortgeschichtenkomplexität postuliert [Hoenen \(2018a\)](#), der als einen wenn auch vorraussichtlich nur selten anwendbaren Ausweg hochindividuelle Wortgeschichtencomics als mögliche komplexe Visualisierung präsentiert (Konferenzvortrag), siehe Abb. 10. Personifikation als visualisierender Parameter wurde auch im künstlerischen Umfeld im Zusammenhang mit Zeit verwendet (vgl. den Januskopf der Mythologie, sowie [Rosenberg and Grafton \(2013, S.208\)](#)).

Die Gefahr visueller Überladung besteht besonders dann, wenn Zusammenhänge sehr komplex werden und sehr viele nicht relatierte oder korrelierte Einzelaspekte dargestellt werden müssen. Umsomehr wird dann eine automatische Visualisierungsmaschinerie Schwierigkeiten haben. Das Dilemma der Lexikologie ist aber nunmal, dass die besonders komplexen Zusammenhänge auch die besonders interessanten sind. Es ist möglich gute komplexe Visualisierungen automatisch zu erschaffen und mehr noch, Visualisierungen, die Datenmengen verarbeiten, welche manuell nicht verarbeitbar sind leisten einen besonderen Beitrag der Komplexitätsreduktion. Allerdings existieren Visualisierungen, welche ggf. künstlerische oder ästhetisierende Anteile haben oder sehr komplexe kognitive Prozesse wie Personifikation, Ironie und vor allem Metaphorik, welche noch immer computational nicht völlig zufriedenstellen auflösbar scheint, ausnutzen. Als Zugangsweg kann eine solche Form besonders attraktiv wirken und auch die Eintönigkeit der automatisch erzeugten immer gleichen Visualisierungen einer digitalen Repräsentation aufbrechen. Voraussetzung ist jedoch ein hoher Aufwand sowie das Vorhandensein und Erkennen geeigneter Kandidatenwörter.

3 Welche Aufgaben und welche Visualisierung

Neben der konkreten Visualisierungsaufgabe spielen bestimmte Parameter wie Wortfrequenz eine Rolle dafür, welche Art Visualisierung am geeignetesten ist. Dies soll am Beispiel der Kombination Beleghäufigkeit- Anzahl Bedeutungen exemplifiziert werden:

	selten belegt	mitteloft belegt	häufig belegt
wenig Bedeutungen	Belegliste (Konkordanz)	Bezeugungskurve(n)	Bezeugungskurve(n)
mittelviele Bedeutungen	Timeline mit Bedeutungsspannen	proportional stacked area (s.o.)	Timeline Rohrdatz et al. (2011)
viele Bedeutungen	? (Mosaikplot,...)	2-d Wege	"

Im Idealfall könnte ein Decision Tree entwickelt werden, der von individuellen Datenlagen zu Vorschlägen für geeignete Visualisierungen führt.

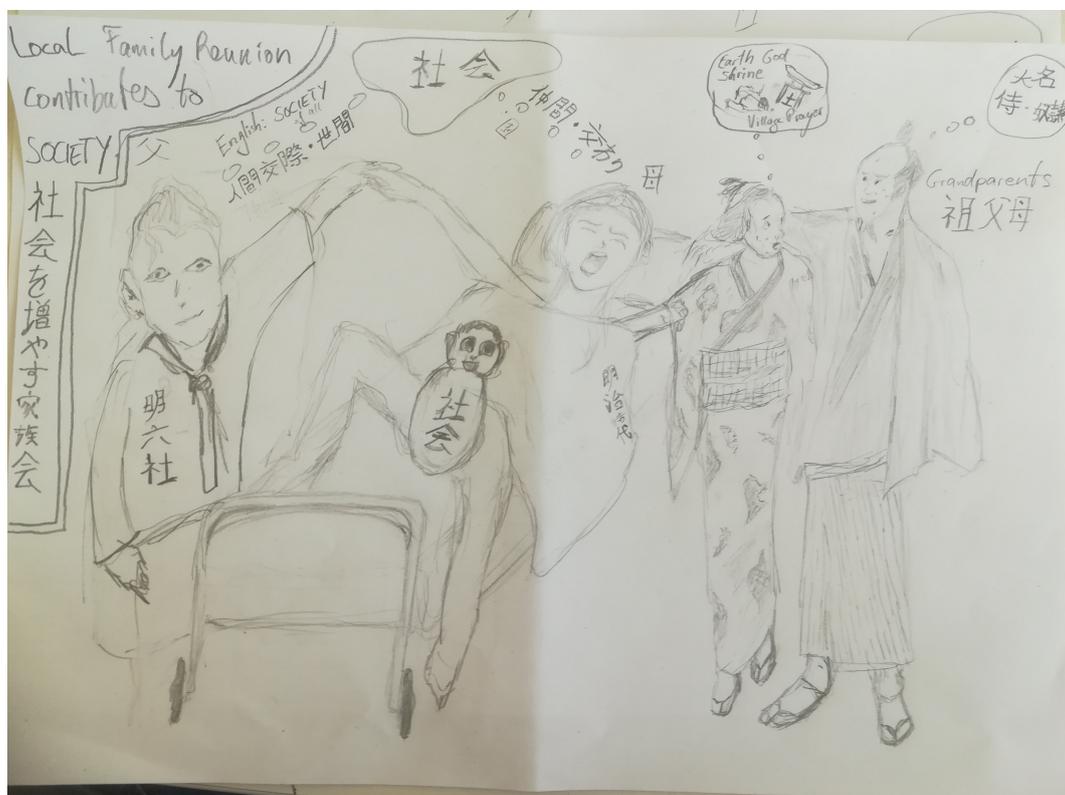


Abbildung 10: Ein Manga zur Exploration der Etymologie bzw. Entstehung des Japanischen Wortes SHAKAI, für Gesellschaft frei nach [Yanabu \(1991\)](#). Das Wort entstand als Neologismus aus existierenden Elementen um die Meiji Zeit (1868-1912) nach der Öffnung Japans für westliche Einflüsse. Daher ist eine Geburtsmetapher passend. Ein Konzept *Gesellschaft* im westlichen Sinne verallgemeinernd wie z.B. im DWDS als 1. Bedeutung angegeben fehlte im feudal geprägten Japan jener Zeit, das durch die Großeltern repräsentiert ist, welche an Worte *denken*, die feudale und lokale Kontexte evozieren, welche in den beiden Bestandteilen SHA und KAI immanent sind. Die Eltern sind die Meijizeit (Mutter) und die Meirokusha (Vater), eine Gruppe Intellektueller, welche ein einflussreiches Journal zur Reflektion westlicher Konzepte herausgegeben haben. Siehe auch [Hoenen \(2018a\)](#). Die Verortung bedeutungstragender Elemente im Bild erfordert vom Betrachter einige Arbeit, was durch eine filmische Darstellung sicher vereinfacht würde (Etymologie-Anime). Es ist allerdings hoch fragwürdig, ob Geschichten oder sogar ähnliche Einzelbilder zur Verdeutlichung wissenschaftlicher Sachverhalte für jede mögliche Konstellation erzeugbar sind und ob es genug interessierte Betrachter gäbe den Herstellungsaufwand zu rechtfertigen.

4 Konklusion

In diesem Artikel wurde versucht einen Überblick über Aspekte und viele Beispiele für Visualisierungen in der historischen Lexikographie und Lexikologie zu geben. Dabei wurde eine Kategorie text-dominanter Visualisierungen angenommen, welche solche visuellen Darstellungen beschreibt, die einerseits für rein sequentiell schriftliche Darstellungen nicht nötig sind und deren Primärzugang das Lesen als kognitiver Prozess ist. Tabellen sind Teil solcher Visualisierungen und im Rahmen der Lexikographie können sie u.a. z.B. zur systematisch-strukturierenden Darstellung von Wortfamilien dienen. Schließlich wurden 2 neue automatisch generierbare Visualisierungen vorgestellt, sowie ein künstlerisches Herangehen an sehr komplexe Visualisierungen. Insgesamt hat die Untersuchung ergeben, dass sich Visualisierungen im Kontext historischer Lexikographie und Lexikologie trotz einiger eventueller Schwierigkeiten wie einer ggf. lückenhaften Beleglage in Zukunft als Zugriffsform und Anreicherung für Artikel im Digitalen Raum gewinnbringend nutzen lassen.

Literatur

- Abromeit, F., Chiarcos, C., Fäth, C., and Ionov, M. (2016). Linking the tower of babel: modelling a massive set of etymological dictionaries as rdf. In *Proceedings of the 5th Workshop on Linked Data in Linguistics (LDL-2016): Managing, Building and Using Linked Language Resources, Portoroz, Slovenia*, pages 11–19.
- De Melo, G. (2014). Etymological wordnet: Tracing the history of words. In *Proceedings of LREC 2014*, pages 1148–1154. ELRA.
- Dixit, C. and Karrfelt, F. (2016). Visualizing etymology: A radial graph displaying derivations and origins.
- Gould, S. J. (1993). *Eight little piggies: Reflections in natural history*. W.W.Norton.
- Hamilton, W. L., Leskovec, J., and Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*.
- Heyer, G., Wiedemann, G., and Niekler, A. (2018). *Topic-Modelle und ihr Potenzial für die philologische Forschung*, pages 351–368. Berlin, Boston: De Gruyter.
- Hoenen, A. (2018a). Attempts at visualization of etymological information. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).
- Hoenen, A. (2018b). Recurrence analysis function, a dynamic heatmap for the visualization of verse text and beyond. *Visualisierung sprachlicher Daten: Visual Linguistics–Praxis–Tools. Heidelberg: Heidelberg University Publishing*.
- Hoenen, A. (2019). Annotated timelines and stacked area plots for visualization in lexicography. *Proceedings of the Elexis workshop, collocated with EADH Galloway*.

- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.
- Moretti, F. (2013). *Distant reading*. Verso Books.
- Ong, W. J. (2013). *Orality and literacy*. Routledge.
- Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.
- Piper, P. et al. (1883). *Die Schriften Notkers und seiner Schule*, volume 3. JCB Mohr.
- Rehbein, M. (2017). Informationsvisualisierung. In *Digital Humanities*, pages 328–342. Springer.
- Richter, M. I. (2014). *Scientific Visualisation: Epistemic Weight and Surpluses*. Peter Lang GmbH.
- Rohrdantz, C., Hautli, A., Mayer, T., Butt, M., Keim, D. A., and Plank, F. (2011). Towards tracking semantic change by visual analytics. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 305–310. Association for Computational Linguistics.
- Rosenberg, D. and Grafton, A. (2013). *Cartographies of time: A history of the timeline*. Princeton Architectural Press.
- Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages, VL '96*, pages 336–, Washington, DC, USA. IEEE Computer Society.
- Splett, J. (2002). Bedingungen des Aufbaus, Umbaus und Abbaus von Wortfamilien. In Hundsnurscher, F., Augst, G., Splett, J., Gruaz, C., and Haßler, G., editors, *Lexikologie/Lexicology. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen.*, Handbücher zur Sprach- und Kommunikationswissenschaft 21.1. de Gruyter.
- Tufte, E. (1983). *The Visual Display of Quantitative Information*. Graphics Press.
- Yanabu, A. (1991). *Modernisierung der Sprache: eine kulturhistorische Studie über westliche Begriffe im japanischen Wortschatz*. Iudicium-Verlag.
- Zhang, W., Geeraerts, D., and Speelman, D. (2015). Visualizing onomasiological change: Diachronic variation in metonymic patterns for woman in chinese. *Cognitive Linguistics*, 26(2):289–330.