Manuel Braun/Nils Reiter

Sangsprüche auf/in Wörterwolken oder: Vorläufige Versuche zur Verbindung quantitativer und qualitativer Methoden bei der Erforschung mittelhochdeutscher Lyrik

In unserem Alltag sind wir es gewohnt, Dinge aus verschiedenen Abständen zu betrachten, und wir wissen, dass wir jeweils etwas Anderes sehen, wenn wir sie aus geringer, aus mittlerer oder aus großer Distanz wahrnehmen. Wer unmittelbar vor einem Gründerzeithaus steht, dem fallen die Risse im Putz auf, die sich auf Höhe seiner Augen zeigen; wer es aus einem Abstand von 30 Metern anschaut, nimmt hingegen die Zahl und die Höhe der Stockwerke auf oder die Verzierung der Fassade durch vorgesetzte Erker und aufgesetzten Stuck; und wer es von einem drei Kilometer entfernten Hochhaus ausmacht, kann es im Kontext seines Viertels verorten, in dem es sich befindet. Solch unterschiedliche Blicke sollen im Folgenden auch auf den Sangspruch des 13. Jahrhunderts gerichtet werden, indem quantitative Analysen des Wortschatzes zum Ausgangspunkt für die qualitative Betrachtung einzelner Texte gemacht werden. Statistiken zum Lexikon dienen also als Werkzeug, mit dessen Hilfe Autorprofile ausgearbeitet und Literaturgeschichte überprüft werden kann.

Die Untersuchung legt dazu die folgenden sechs Schritte zurück: Erstens wird festgestellt, aus welcher Distanz die rezente Sangspruchforschung ihren Gegenstand wahrnimmt (Manuel Braun). Zweitens werden die Konzepte des ‚distant reading' und des ‚scalable reading' eingeführt (Manuel Braun). Drittens wird erklärt, wie sich Worthäufigkeit quantitativ analysieren lässt (Nils Reiter). Viertens werden das Korpus – es umfasst Walther von der Vogelweide, Bruder Wernher, Reinmar von Zweter, den Marner, Boppe, Rumelant von Sachsen sowie Frauenlob – und die Art seiner Aufbereitung vorgestellt (Manuel Braun/Nils Reiter). Fünftens geht es darum, statistische Verfahren als Heuristiken bei der Erstellung von Autorprofilen zu benutzen (Manuel Braun/Nils Reiter). Sechstens werden die Wortstatistiken als Verfahren eingesetzt, um literaturgeschichtliche Hypothesen zum Sangspruch zu testen (Manuel Braun/Nils Reiter).

1. Einzelne Autoren, einzelne Texte: zum Fokus der Sangspruchforschung

Gegenstand des folgenden Forschungsberichts sind 42 Aufsätze, die sich (nahezu) ausschließlich mit Sangspruchdichtung beschäftigen und die im Jahrzehnt 2005/2014 erschienen sind. Hält man sich an die gängigen Bibliographien, ist damit der Großteil der neueren Forschung zum Sangspruch jenseits der Monographien und Editionen ausgewertet. Wenn man dieses Sample daraufhin befragt,

wie viele Autoren und Strophen ein Aufsatz jeweils untersucht – als Kriterium hierfür gilt ein wörtliches Zitat im Haupttext –, ergibt sich folgendes Bild: 23 der 42 Aufsätze (55 %) widmen sich einem einzigen Autor (bzw. einem anonymen Korpus), bei den übrigen 19 (45 %) sind es zwischen zwei und 13, der Durchschnitt liegt hier bei 4,9. Auf die gesamte Probe gerechnet werden pro Publikation 2,8 Autoren (bzw. anonyme Korpora) ausgewertet. Die Zahl der zitierten Strophen liegt zwischen einer und 48, der Durchschnitt bei 9,8. Über diesem Durchschnitt liegen 15 Aufsätze (36 %), die übrigen 27 (64 %) behandeln zwischen einer und neun Strophen; 14 (33 %) beschränken sich auf eine Textbasis von einer bis drei Strophen.

Was die quantitative Aufbereitung der Sangspruchforschung zu Tage bringt, ist ein ‚Normalmodus' philologischer Praxis. Dieser nimmt eine überschaubare Anzahl von Strophen in den Blick, die von nur wenigen Autoren stammen. Der Zugang zum Sangspruch ist in der Regel ein analytisch-interpretatorischer, und der ist eben nur möglich, wenn die Zahl der betrachteten Texte begrenzt bleibt. Es steht außer Zweifel, dass die Mediävistik auf diese Weise Ergebnisse erzielt, die ihrem disziplinären Selbstverständnis als hermeneutische Wissenschaft entsprechen und die diesem zufolge relevant und valide sind. Es ist allerdings auch so, dass mit dieser Methode nur bestimmte Fragen aufgeworfen und beantwortet werden können, andere hingegen, etwa literatur- und kulturgeschichtliche oder historisch-semantische, ausgeblendet werden. Für diese bräuchte es die Bearbeitung deutlich größerer Textmengen, bräuchte es so etwas wie die Draufsicht aus der Distanz. Der Begriff der Distanz ist ein unmittelbares Zitat, denn in der Computerphilologie firmiert eine solche Forschungspraxis seit einiger Zeit unter dem Etikett des ‚distant reading'.

2. ‚Distant reading', ‚scalable reading': Konzepte einer datengetriebenen Literaturwissenschaft

Geprägt worden ist der Begriff des ‚distant reading' von Franco Moretti. Erstmals verwendet Moretti ihn 2000 in seinem Aufsatz „Conjectures in World Literature", und 2013 verleiht er ihm programmatischen Charakter, indem er ein Buch mit ihm betitelt. Dieses besteht aus einer Sammlung von Studien, die etwa in Auseinandersetzung mit Erich Auerbach und Ernst Robert Curtius Europa als Umwelt der verschiedenen Nationalliteraturen beschreiben, die das System der Weltliteratur mithilfe von Immanuel Wallersteins Modell von Zentrum und Peripherie ordnen und die in ihm ablaufende Prozesse der Diversifizierung und Unifizierung von Gattungen evolutionstheoretisch erklären, die die Durchsetzung der Detektivgeschichten Conan Doyles auf dem literarischen Markt als Effekt formaler Eigenschaften deuten, die die weltweite Verbreitung des Hollywood-Films nachzeichnen, die die Geschichte des europäischen mit der des chinesischen Romans vergleichen und so deren jeweilige Eigenheiten herausarbeiten, die die Titel von 7.000 britischen Romanen der Jahre 1740 bis 1850 linguistisch analysieren und die beobachteten Veränderungen auf solche des Buch-

markts beziehen oder die die Figurenkonstellationen literarischer Texte mittels Netzwerkanalysen darstellen.

Schon diese knappen Inhaltsangaben lassen einige Prinzipien des ‚distant reading' hervortreten, und da es sich bei ihm um eine Praxis literaturwissenschaftlichen Arbeitens, nicht um eine ausformulierte Theorie handelt, bleibt auch kein anderer Weg, als sie aus Morettis Aufsätzen zu abstrahieren. Demnach nimmt das ‚distant reading' nicht den einzelnen Text in den Blick, sondern eine große Zahl von Texten, die auch nur durch bibliographische Verzeichnisse oder literaturgeschichtliche Forschung repräsentiert sein können. Geschuldet ist dieses Vorgehen dem Interesse, zeitlich und räumlich weit ausgreifende literaturgeschichtliche Fragen zu beantworten, und es geht Hand in Hand mit dem Wunsch, die Schranken des Kanons – für Moretti: „secularized theology" (67) – zu durchbrechen und sich dem Ganzen der Überlieferung zuzuwenden.

Nahegelegt wird das ‚distant reading' auch durch die Digitalisierung:

> With digital databases, this is now easy to imagine: a few years, and we'll be able to search just about all novels that have ever been published, and look for patterns among billions of sentences. Personally, I am fascinated by this encounter of the formal and the quantitative." (164)

Digitale Archive machen der Literaturwissenschaft nicht nur enorme Textmengen zugänglich, sondern ermöglichen es ihr überhaupt erst, diese durch computergestützte Suchabfragen zu bearbeiten: „the width of the corpus and the speed of the search have increased beyond all expectations: today, we can replicate in a few minutes investigations that took a giant like Leo Spitzer months and years of work" (212). Diese Möglichkeiten haben auch zur Folge, dass die sprachliche Verfasstheit der Literatur wieder stärker in den Fokus der Forschung gerät. Denn die Werkzeuge der Computerlinguistik eignen sich besonders dazu zu erfassen, aus welchen Wörtern und Sätzen Texte bestehen: „When it comes to phenomena of language and style, we can do things that previous generations could only dream of." (212) Das ‚distant reading' operiert also nicht nur oberhalb der Ebene des Textes (auf der der Gattung oder eines sonst wie konstituierten Korpus), sondern auch unterhalb von ihr (auf der des Wortes oder des Motivs).

Sein Zugriff ist ein quantifizierender, und entsprechend verändern sich mit dem Übergang vom ‚close reading' zum ‚distant reading' auch die Darstellungs- und die Herangehensweise. Erstere bevorzugt „Graphs, Maps, Trees" (F. Moretti, 2005), Letztere bricht bewusst mit den überkommenen Methoden: „We do not need more interpretations, […], not because they have nothing to say, but because, by and large, they have already said what they had to" (F. Moretti, 2013, 154). Um zu lernen, „how not to read [texts]" (48), lehnt sich das ‚distant reading' an die Naturwissenschaften an, es arbeitet mit abstrakten Modellen und sucht nach kausalen Erklärungen. Außerdem veranstaltet es Experimente:

> It will be difficult, no doubt, because one cannot study a large archive in the same way one studies a text: texts are designed to ‚speak' to us, and so,

> provided we know how to listen, they always end up telling us something; but archives are not messages that were meant to address us, und so they say absolutely nothing until one asks the right question. And the trouble is, we literary scholars are not good at that: we are trained to listen, not to ask questions, and asking questions is the opposite of listening: it turns criticism on its head, and transforms it into an experiment of sorts: ‚questions put to nature' is how experiments are often described, and what I'm imagining here are questions-put to culture. Difficult; but too interesting not to give it a try. (165)

Auch wenn wir Morettis Aufruf „give it a try" aufnehmen und uns von seinen Überlegungen anregen lassen wollen, werden wir uns im Folgenden nicht allein dem ‚distant reading' verschreiben. Vielmehr folgen wir Martin Muellers Konzept des ‚scalable reading' (2012) und wechseln also zwischen Fern- und Nahsicht und kombinieren die neuen quantitativen Verfahren mit erprobten qualitativen.

3. Grundlagen der quantitativen Textanalyse

Wir werden im Folgenden einige zentrale Erkenntnisse und Ideen aus der quantitativen Textanalyse einführen. Zählt man Wörter, ist eine der ersten – und banalen – Beobachtungen, dass verschiedene Wörter unterschiedlich häufig vorkommen. Tatsächlich ist die Verteilung der Wörter und ihrer Häufigkeiten keineswegs zufällig – sie folgt einer Regelmäßigkeit, die als Zipf'sches Gesetz bekannt ist (G. K. Zipf, 1935): Sortiert man die Wörter in einem Text nach ihrer Häufigkeit und zeichnet sie in ein Koordinatensystem ein, ergibt sich eine Kurve, die ungefähr der Funktion $f(x) = 1/x$ entspricht (Abb. 1, gestrichelte Linie). Es gibt also ganz wenige Wörter, die extrem häufig (linkes Ende), und etwas mehr Wörter, die noch oft vorkommen. Dann jedoch nähert sich die Kurve asymptotisch der x-Achse an: Unendlich viele Wörter kommen ganz selten vor (der sog. ‚long tail').[1] Diese Gesetzmäßigkeit zeigt sich auch im Korpus der Sangsprüche (Abb. 1, durchgezogene Linie).

Schaut man sich die sehr häufigen Wörter an, stellt man fest, dass es sich bei diesen um inhaltlich wenig aussagekräftige Wörter handelt: Im zeitgenössischen Deutsch etwa kommen Artikel, Pronomina und Konjunktionen am öftesten vor. Sie bilden geschlossene Wortklassen, zu denen es nur selten Neuschöpfungen gibt. Da diese Wörter in maschinellen Anwendungen vor der weiteren Verarbeitung meist entfernt werden, werden sie auch als ‚stop words' bezeichnet.[2] Offene Wortklassen sind dagegen Nomina, Verben und Adjektive, in denen neue Wörter erscheinen (können).

1 Im Neuhochdeutschen, das eine sehr produktive Wortbildung aufweist, ist das auch intuitiv leicht verständlich: Das Wort „Bahnticketautomatlicht" z. B. ist zum Zeitpunkt des Schreibens dieses Artikels noch nicht im Internet vertreten.

2 Bei einer Google-Suche etwa werden sie im Normalfall ignoriert.

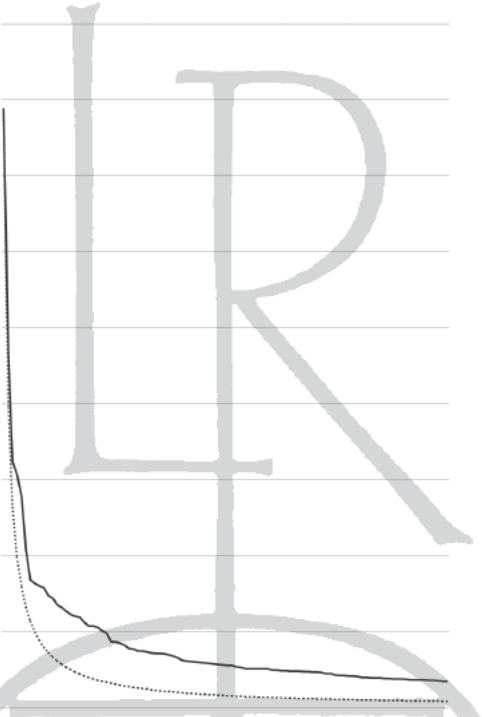

Abb. 1: Zipf-Verteilung im Sangspruchkorpus (durchgezogen) und theoretische Vorhersage (gestrichelt)

Eine Folge des Zipf'schen Gesetzes ist, dass in (sinnvollen) Texten einer gewissen Länge zwangsläufig Wörter mehrfach vorkommen. Ein Kriterium für sprachlichen Einfallsreichtum könnte nun sein, möglichst viele unterschiedliche Wörter zu verwenden. Quantifizieren lässt sich das über das Verhältnis aus der Zahl aller Wörter (Tokens) und der Zahl der verschiedenen Wörter (Types) in einem Text (oder Korpus). Das Verhältnis aus Types und Tokens ist dann eine Maßzahl zur Einschätzung der sprachlichen Variabilität (sog. Type-Token-Relation, TTR, vgl. B. J. Richards, 1987). Ein Verhältnis von z. B. 30 heißt, dass von 100 Wörtern 30 neu sind (also in diesem Text noch nicht verwendet worden sind). Aufgrund des Zipf'schen Gesetzes ist es zwangsläufig so, dass die TTR bei längeren Texten sinkt: Immer mehr Wörter werden wiederholt (Artikel etc.), was die Zahl der Tokens erhöht, ohne dass die Zahl der Types steigt. Es ist daher sinnvoll, die Type-Token-Relation zu normalisieren und in einen festen Bezugsrahmen zu setzen: Die standardisierte Type-Token-Relation (STTR) berechnet die TTR für 1000-Wörter-Fenster, die über den Text gelegt werden. Das arithmetische Mittel der TTR-Werte bildet dann die STTR.

Das reine Zählen von Wörtern nach absoluter Häufigkeit ist noch aus einem anderen Grund wenig zielführend, wenn es darum geht ‚interessante' Wörter zu finden: Die Texte sind unterschiedlich lang. Entsprechend wenig sagt es zunächst aus, wenn das Wort *meisterschaft* sowohl bei Boppe als auch beim Marner zweimal vorkommt. Wir gehen daher dazu über, relative Häufigkeiten zu zählen, d. h. wir normalisieren die absoluten Häufigkeiten über die Textlänge. Die relative Häufigkeit von *meisterschaft* beim Marner ist dann 0.000407083,[3] bei Boppe ist sie 0.000819001, also etwa doppelt so hoch – wären die Korpora gleich groß, würde *meisterschaft* bei Boppe etwa doppelt so oft auftreten wie beim Marner. Mit einem Computerprogramm lässt sich die relative Häufigkeit für alle (verschiedenen) Wörter (Types) berechnen. Dabei zeigt sich in aller Regel, dass auch die relativ häufigsten Wörter noch nicht sonderlich interessant sind. Typischerweise handelt es sich bei ihnen um allgemeine, polyseme Wörter, die keinen direkten Rückschluss auf den Inhalt erlauben. Das häufigste sinntragende Wort im Sangspruch ist z. B. *man*.

TF*IDF ist ein Maß aus der Informationsextraktion (K. Spärck Jones, 1972) und misst, wie relevant ein Wort für ein Dokument oder Korpus im Vergleich mit anderen Dokumenten oder Korpora ist. In die Berechnung von TF*IDF fließt also das Gesamtkorpus mit ein. Berechnet man den TF*IDF-Wert für das gleiche Wort aus dem gleichen Dokument, aber in einem anderen Korpus, erhält man einen anderen Wert. TF*IDF beschreibt das Verhältnis aus ‚term frequency' (TF) und ‚inverted document frequency' (IDF). ‚Term frequency' ist die relative Häufigkeit eines Wortes in einem Dokument, ‚document frequency' beschreibt, in wie vielen Dokumenten das Wort insgesamt vorkommt. Der TF*IDF-Wert berechnet sich dann durch Teilung der beiden Werte.[4] Ein hoher Wert für ein Wort in einem Dokument heißt, dass es in diesem Dokument häufig vorkommt und in den anderen Dokumenten nicht so häufig. Einen niedrigen Wert erhält man, wenn das Wort selten in diesem Dokument oder aber in allen Dokumenten vorkommt.

4. Zusammenstellung und Aufbereitung des Korpus

Hinter der Auswahl Walthers von der Vogelweide, Bruder Wernhers, Reinmars von Zweter, des Marners, Boppes, Rumelants von Sachsen und Frauenlobs stehen drei Überlegungen: Die Korpora sollten so groß sein, dass man mit ihnen überhaupt statistisch arbeiten kann; sie sollten wichtige Ausprägungen der Gattung exemplarisch abbilden; und sie sollten sich einigermaßen gleichmäßig über das 13. Jahrhundert verteilen, um Fragen nach literaturhistorischen Entwicklungen zu ermöglichen.

Aus pragmatischen Gründen werden jeweils die gängigen Editionen verwendet. Diese wurden, sofern nötig, digitalisiert und korrigiert. Anschließend

3 Die Zahlen werden sehr klein, das macht sie unhandlich, aber nicht unvergleichbar.

4 ‚Inverted' beschreibt die Tatsache, dass man mit der invertierten ‚document frequency' multipliziert, also durch sie teilt: TF*IDF = TF x (1/DF) = TF/DF.

wurde der Wortschatz über ‚stop word'-Listen auf sinntragende Wörter (Vollverben, Substantive, Adjektive) reduziert. Diese wurden schließlich von Hand elementar nach Lexer lemmatisiert. Da dabei nicht alle Wörter in ihrem jeweiligen Verwendungskontext aufgesucht werden konnten, ist mit einer gewissen Anzahl von Fehlern zu rechnen, etwa durch Homonyme. Statistisch sollten diese nicht allzu sehr ins Gewicht fallen, zumal das so generierte Material hier nur dazu dient, Tendenzen zu entdecken.

Zur Exploration dienen zum einen Wörterwolken, zum anderen interaktive Diagramme. Wörterwolken sind eine einfache und anschauliche Möglichkeit, um die Vorkommenshäufigkeit von Wörtern zu visualisieren. Wörter werden dabei umso größer (Schriftgröße) dargestellt, je höher der ihnen zugeordnete Zahlenwert ist. Da sich die Schriftgröße nur auf die vertikale Höhe auswirkt, kann bei sehr langen oder sehr kurzen Wörtern ein leicht verfälschter Eindruck entstehen. Die Positionierung im Raum erfolgt automatisch so, dass möglichst wenige Lücken entstehen; die Entfernung der Wörter voneinander hat also keine interpretierbare Bedeutung. Die Wörterwolken hier wurden mit dem Werkzeug ‚wordle' (http://www.wordle.net) erstellt; sie enthalten jeweils die nach Maßgabe des TF*IDF-Werts häufigsten 75 Wörter. Sie sollen es ermöglichen, Besonderheiten im Wortschatz der untersuchten Autoren zu entdecken, also Wörter, die bei diesen häufig vorkommen (und häufiger als bei anderen Autoren). In den Diagrammen ist für jedes Autorkorpus und für jedes Wort ablesbar (interaktiv, damit das Anzeigen aller Wörter nicht zur Unlesbarkeit führt), ob die relative Häufigkeit des Wortes im Korpus vom Durchschnitt des Gesamtkorpus abweicht und, wenn ja, um wie viel. Ein positiver Wert auf der y-Achse bedeutet also, dass ein Wort in dem auf der x-Achse bezeichneten Autorkorpus häufiger vorkommt als im Gesamtkorpus.

5. Worthäufigkeit als Weg zum Autorprofil

Am Beispiel Walthers von der Vogelweide, Bruder Wernhers und Reinmars von Zweter soll nun gezeigt werden, wie sich Wörterwolken nutzen lassen, um die Semantik von Sangspruchkorpora zu explorieren. Sieht man sich die wichtigen Wörter in Walthers Sangspruch-Œuvre an (Abb. 2), findet man etliche, die entweder für die Gattung oder für den Autor erwartbar sind. Gattungstypisch sind Wörter, die sich den Themen ‚Ethik' (*êre*, *guot*, *mâze*, *scham*, *staete*, *triuwe*) und ‚Religion' (*geist*, *heide*, *kristenheit*, *meit*, *muoter*) zuordnen lassen oder die auf die Existenz eines Fahrenden Bezug nehmen (*arm*, *gast*, *milt*, *silber*). Andere Wörter kommen zwar absolut gesehen seltener vor, doch hebt die Wörterwolke sie gleichwohl hervor, weil Walther sie öfter als seine Kollegen gebraucht. Sie sind demnach als autortypisch einzuschätzen. Das bekannte Bild des (religions-) politischen Sängers etwa bestätigt die Bedeutung des Wortes *pfaffe*, dem sich auch der Gegenbegriff *leie* beiordnen lässt. Auch die Prominenz der Wörter *hof* und *keiser* passt hierzu.

Abb. 2: Wörterwolke zu Walther von der Vogelweide

Die Wolke lenkt den Blick allerdings auch auf Wörter, die jenseits der bekannten Sangspruch- und Walther-Klischees stehen und die, wenn man sie als wichtig wahrnimmt, zu Nachfragen anregen. So führt Walther offenbar einen intensiven Diskurs über das Sprechen – *zunge* ist eines der von ihm am häufigsten benutzten Wörter, aber auch *rede* kommt vergleichsweise oft vor –, der sich in mehrere Richtungen verzweigt: in eine, die verschiedene Arten von Sprechakten benennt (*danken*, *heizen*, *klagen*/*klage*, *künden*, *loben*), in eine, die wahrheitsgemäßes Sprechen thematisiert (*liegen*, *triegen*), und in eine, in der es um kunstgemäßes Sprechen geht (*sanc*). Walthers Diskurs über die Sprache und das Sprechen aufzuarbeiten und ihn mit dem zu vergleichen, den andere Autoren über das Thema führen, könnte eine lohnende Aufgabe darstellen. Denn es ist ja denkbar, dass auch hier der Minnesang, der konstantive und performative Aussagen immer schon engführt (‚ich liebe, ich singe'), auch den Sangspruchdichter Walther in Richtung einer verstärkten Selbstbezüglichkeit gelenkt hat.

Da dieses Thema den Rahmen überschreitet, der diesem Aufsatz gesteckt ist, haben wir einen anderen quantitativen Befund qualitativ nachgearbeitet und sind einer Spur nachgegangen, auf die uns die Wörter *bluomen*, *brechen* und *grüen* geführt haben, die eher in Minneliedern – genauer: im Natureingang – zu vermuten wären als im Sangspruch. Diese Annahme bestätigt sich insofern, als die *Lehensbitte* (L 28,1) ja die Rückkehr zur Kunst des Minnesangs in Aussicht stellt und diesen durch die Elemente des *locus amoenus* – Vögel, Heide und eben Blumen – aufruft. Auch der Frauenpreis *Durchsüezet und geblüemet sind die reinen frowen* (L 27,17) operiert auf der Grenze zum Minnesang. Aber die Blume kann auch den Kontext wechseln, das Wort kann metaphorisch gebraucht werden und dem Fürstenlob dienen: So wird der Landgraf Hermann als der *Dürnge bluome* (L 35,15) apostrophiert, und so wird – deutlich raffinierter – der Herzog Leopold mit einer Heide verglichen, *dar abe man bluomen*

Abb. 3: Wörterwolke zu Bruder Wernher

brichet wunder (L 21,5). Es gibt also offenbar bei Walther eine Übertragung der Naturmetaphorik vom Frauen- auf das Fürstenlob. Diese bestätigt nicht nur das Wissen über den Austausch zwischen beiden Gattungen, sondern wirft auch die Frage auf, was aus Gendersicht eigentlich von Fürsten zu halten ist, die als Blumen(wiesen) metaphorisiert werden.

Die Wörterwolke zu Bruder Wernher (Abb. 3) rückt das Wort *vriunt* ins Blickfeld und regt dazu an, sich die entsprechenden Belege anzusehen und herauszufinden, welches Konzept der Freundschaft dessen Sprüche vertreten. Das Ergebnis dieser Probebohrung sieht so aus: Der Sprecher selbst scheint an Freundschaft zu glauben, denn er wendet sich an *wîse* (Nr. 13, V. 9) und *liebe vriunt* (Nr. 60, V. 4), um von ihnen Rat zu erbitten. Seine Lehren sind freilich weniger eindeutig. So erklärt er in einem Spruch, dass man von Freunden keine Hilfe zu erwarten habe (Nr. 14), in zwei anderen, dass man nur mit ihrer Hilfe im Kampf bestehen könne (Nr. 28f.). Er ruft dazu auf, dem Freund in der Not beizustehen, und warnt davor, dass aus diesem sonst ein Feind werden könne (Nr. 67). Auch falsche Freunde tauchen an einer Stelle auf (Nr. 70). Freundschaft spielt bei Bruder Wernher also durchaus eine wichtige Rolle, ihr Kern ist die gegenseitige Hilfeleistung, doch kann man sich auf diese, das klingt wiederholt an, eben nicht verlasen.

Diese Nachforschung haben wir nun im Korpus Reinmars von Zweter wiederholt (Abb. 4), wo das Wort *vriunt* ebenfalls überdurchschnittlich oft auftaucht, aber anders gebraucht wird. Es dient hier der Anrede des Adressaten (Nr. 92, V. 5; Nr. 122, V. 12), es meint den Geliebten (Nr. 49), und es referiert ironisch auf das Glück (Nr. 248, V. 3). Nicht so viel Bedeutung hat es demgegenüber als Bezeichnung einer Person, zu der eine frei gewählte, aber ethisch basierte Bindung besteht. Und wenn doch, dann wird sie öfter nicht um ihrer selbst wil-

Abb. 4: Wörterwolke zu Reinmar von Zweter

len angesprochen, sondern dient der Veranschaulichung anderer Themen: etwa der Themen ‚Wahrhaftigkeit' (Nr. 95) und ‚Klugheit' (Nr. 123), deren Bedeutung sich unter anderem daran zeigt, dass sich mit ihrer Hilfe Freunde gewinnen lassen; der des Themas ‚Patenschaft', die nur dann dauerhaft wirkt, wenn sie aus Freundschaft übernommen wird (Nr. 168); der des Themas ‚Boshaftigkeit', die nicht einmal davor zurückschreckt, Freunde einander zu entfremden (Nr. 174); oder der des Themas ‚Vergänglichkeit', die dazu führt, dass Freunde einander ins Grab folgen (Nr. 229). Die Strophen, die zur Beständigkeit bzw. zur Lauterkeit in der Freundschaft mahnen (Nr. 124, 243), stehen im Werk Reinmars recht vereinzelt da, und die anderen, die auf deren Zweckfreiheit abheben (Nr. 249f.), negieren das, was bei Wernher zentral war: die gegenseitige Hilfe. Letztere kommt nur einmal vor, und zwar als Beistand mittels Besitz (Nr. 244).

Die drei Beispiele haben gezeigt, was ein quantitativer Zugang zum Wortschatz für die Erfassung von Autorprofilen zu leisten vermag: Er lenkt den Blick auf Wörter, die in einem Korpus häufiger vorkommen als in einem anderen. Vielfach sind das Wörter, die das Vorwissen bestätigen, das man über einen Autor ohnehin hat. Immerhin erfährt man dann genauer, welche Semantik in dieses Vorwissen Eingang gefunden hat. Außerdem kann man unter den wichtigsten Wörtern des Autorkorpus recht rasch Gruppen semantisch verwandter Wörter ausmachen, und über solche Wortfelder dann auf Diskurszusammenhänge kommen. Vielleicht am interessantesten sind jene Fälle, in denen man sich zunächst einmal wundert, dass bestimmte Wörter bei einem Autor so oft vorkommen. Sie regen zum gezielten Nachforschen an. Statistische Verfahren sind hier Hilfsmittel einer ersten heuristischen Exploration einzelner Autorkorpora, an die sich tiefergehende Erkundungen anzuschließen haben. So wird aus dem ‚distant reading' ein ‚scalable reading'.