

Wie viele Schwalben machen einen Sommer?

Multiples Testen oder Signifikanzen-Fischen

Andreas Günther, Erdmannhausen

Zusammenfassung

In diesem Artikel möchte ich der Frage nachgehen, wann die an einem Text beobachteten sprachlichen Merkmale aussagekräftig genug sind, um Hypothesen über die Textgenese zu stützen. Es gibt bei allen statistisch betrachtbaren Phänomenen ab und zu „dumme“ Zufälle, die zwar der Intuition bedeutsam erscheinen, hinter denen aber keine planende Gestaltung oder ursächliche Beziehung steht. Man wird daher nach mehreren Hinweisen Ausschau halten, aber wie viele müssen es sein? Mit anderen Worten gefragt: „Wie viele Schwalben machen einen Sommer?“

Scholastica Deck ist in dieser Zeitschrift¹ der Frage nachgegangen, ob man aus Beobachtungen am Wortbestand eines Textes irgendwelche Schlüsse ziehen dürfe, wenn jene Lexeme selten sind und der Referenz-Kanon klein und sehr speziell ist. Ich möchte diese Überlegungen hier erweitern und fragen: Besteht nicht immer noch die Möglichkeit, auf eine falsche Fährte zu geraten, selbst wenn die Mathematik eine klare Sprache spricht? Statistische Argumente werden zwar selten alleine vorgebracht; um sie aber einmal *per se* betrachten zu können, werde ich hier von weiteren Implikationen absehen.

Bei einem statistischen² Vergleich geht man von einer Standard-Hypothese³ aus und formuliert, was von dem Text zu erwarten wäre, wenn diese zuträfe. Z. B. erwartet man zunächst eine jesajanische Sprache im Jesaja-Buch. Dann stellt man gewisse Abweichungen von dieser Erwartung fest und zieht den Schluß, der Text könne nicht zur ersten Schicht des Buches gehören. In einem (vom ersten möglichst zu trennenden) zweiten Schritt findet man vielleicht eine andere *Herkunft* (Schicht, Quelle, Redaktion, Sprachepoche etc., z. B. die dtr. Sprache), zu deren Stil der Text besser passen würde als zu seiner Umgebung.

Implizit macht man dabei folgende Annahmen:

1. Die Verteilung der betrachteten sprachlichen Merkmale hängt im Wesentlichen von der Herkunft des Textes ab. Die Auswirkungen der Herkunft sind deutlich und werden nicht etwa durch andere Einflüsse zu sehr überlagert (wie das Thema des Textes oder die aktuelle Stimmung des Autors).

¹ Scholastica Deck, Wortstatistik - ein immer beliebter werdendes exegetisches Handwerkszeug auf dem (mathematischen) Prüfstand, BN 60 (1991), 7-12.

² Wobei es egal ist, ob dieser als solcher bezeichnet wird oder ob Zahlen zum Einsatz kommen.

³ In der Mathematik meist als „Nullhypothese“ bezeichnet.

2. Die Verteilung dieser Zufallsvariable ist mathematisch faßbar als hypergeometrische Verteilung. Der Wortschatz einer Herkunft wird als ein großes Reservoir aufgefaßt, aus dem der Autor beim Schreiben „schöpft“ und dabei mit einer bestimmten Wahrscheinlichkeit ein bestimmtes Merkmal „erwischt“.

Der zum Vergleich herangezogene Gesamttext einer Herkunft gibt uns mit ausreichender Genauigkeit über die tatsächliche (historische) Beschaffenheit dieser Verteilung Auskunft. Wir nehmen an, daß z. B. Deuteronomisten immer so geschrieben haben, wie wir das (in den Texten, die deuteronomistisch zu nennen wir uns geeinigt haben) heute vorfinden.⁴

3. Verschiedene Herkünfte unterscheiden sich durch unterschiedliche Verteilungen der Merkmale. Während Schicht A Lexem L nur selten benutzt, kommt es in Quelle B häufiger vor.
4. Unsere Annahmen darüber, was zu Schicht A und Quelle B usw. gehört, sind halbwegs zutreffend. Diese Einteilung muß unabhängig von dem Sprachgebrauch der Herkünfte erfolgt sein, sonst sind Zirkelschlüsse unvermeidlich.⁵

Es ist sehr fraglich, ob diese Annahmen (insbesondere die erste) überhaupt zutreffen. Texte zu schreiben und Worte zu wählen hat nichts mit Lotterie zu tun.⁶ Die Argumentation mit Wortwahl und Wortschatz macht aber diese Annahme. Einmal angenommen, sie träfe zu. Dann könnte man die Herkunft eines Textes z. B. daran erkennen, wie oft welches Lexem benutzt wird, sofern man das nötige Vergleichsmaterial hat.

Bei jeder Zufallsvariable gibt es neben vielen typischen Ausfällen auch einige wenige extreme. Zwei Pässe hintereinander im Mensch-ärgere-dich-nicht sind immer ein Anlaß zur Verwunderung bei den Mitspielern, aber es kommt vor. Schnee im Sommer und warme Tage im Winter sind unwahrscheinlich, aber nicht ausgeschlossen. Genauso gibt es von einem Künstler viele für ihn typische Werke und daneben einige ausgefallene. Und letztere laufen eben Gefahr, zu Unrecht als unecht eingestuft zu werden.

Schlüsse aufgrund einer Zufallsvariablen hängen also auch immer davon ab, ob man das Pech hatte, eine untypische Stichprobe zu ziehen. Sie unterliegen also

⁴ Wahrscheinlich bietet uns die Bibel aber keine unverzerrte Stichprobe aus den jeweiligen Sprachprofilen (vgl. Deck [s. Anm. 1], 8; Christof Hardmeier, *Jesajaforschung im Umbruch*, *VuF* 31 (1986), 3-31, 15). Dies könnte dahin führen, daß ein Text falsch eingeordnet wird, nur weil er sich z. B. der Umgangssprache bedient.

⁵ Vgl. Robert Polzin, *Late Biblical Hebrew. Toward an Historical Typology of Biblical Hebrew Prose*, Missoula, Montana 1976, 20.

⁶ Anders verhält es sich mit formalen Kriterien wie Satzbau, Satzlänge, Häufigkeit der Wortarten. Auch dort, wo „linguistischer Kontrast“ besteht, kann die Wortwahl aussagekräftig sein (vgl. Avi Hurvitz, *A Linguistic Study of the Relationship between the Priestly Source and the Book of Ezekiel*, Paris 1982, 124).

selbst wieder dem Zufall. Man kann dem etwas wehren, indem man vorher festlegt, welches Risiko man höchstens eingehen will, sich gegen die Standardhypothese zu entscheiden, obwohl sie richtig ist.⁷

Wenn es z. B. darum geht, ein neues Medikament zuzulassen, werden umfangreiche Versuche über dessen Wirksamkeit gemacht. Man geht zunächst davon aus, das neue Präparat sei unwirksam, eingetretene Effekte seien dem Zufall zu verdanken. Erst, wenn dies nicht mehr haltbar ist, weil eine Zufälligkeit zu unwahrscheinlich wäre, nimmt man die Gegenthese an. Übliche Signifikanzniveaus sind Irrtumswahrscheinlichkeiten (für den Fehler erster Art) von 5 oder 1%. Nach diesem Wert wählt man dann die nötige Berechnung (Test) und führt sie aus. Das heißt aber auch, daß man im Durchschnitt von 100 in Wirklichkeit unwirksamen Medikamenten eines bzw. fünf irrtümlich doch zuläßt.

Die Annahme, daß biblische Bücher von Gott reden, ist plausibel. Man kann auch nicht sagen, daß Gottesbezeichnungen (GB) in der Bibel selten sind; dennoch gibt es ein Buch, das gerade einmal **יה** nennt, noch dazu versteckt in **שלהבתיה** (Hld 8,6). Im Buch Rut (1287 Wörter) sind es immerhin 22 GB. Die Wahrscheinlichkeit, daß ein beliebiges Wort eine GB ist, berechnet sich durch den Quotienten $\frac{\text{GB im Text}}{\text{Wörter insgesamt in diesem Text}}$. Für Rut ist dieser Wert $\frac{1}{58.5}$, im ganzen AT $\frac{1}{31.6}$.

Testen wir die Hypothese, das Buch Rut sei ein at. Text, aufgrund obiger Annahme. Wir überprüfen, ob in diesem Buch GB anders verteilt (seltener oder häufiger) sind als im übrigen AT. Im AT gibt es 9647 Gottesbezeichnungen⁸, davon 161 **אל**, 2614 **אלהים**, 46 **יה** und 6826 **יהוה**. Nun betrachten wir das gesamte AT als Textvorrat, aus dem zufällig ein Buch von 1287 Wörtern gezogen wird. Wie wahrscheinlich ist es, dabei höchstens 22 GB zu erhalten?

Abbildung 1 zeigt die entsprechende Verteilung. Auf der X-Achse sind die möglichen Ausfälle der Anzahl der GB aufgezeichnet. Je höher der Graph, desto wahrscheinlicher; also sind Werte um die 42 herum am wahrscheinlichsten. Durch die senkrechten Striche ist der Bereich eingegrenzt, der 99% aller Ausfälle abdeckt: wir erhalten mit einer Wahrscheinlichkeit von 99% einen Wert zwischen 26 und 57 GB.

Rut liegt außerhalb dieses Bereiches und ist deshalb kein at. Text (auf einem Signifikanzniveau von 1%). 22 GB sind zwar nur vier zu wenig, aber das scheint nur der Intuition wenig.⁹ Unter der Annahme, at. Texte würden sich durch eine

⁷ Neben diesem „Fehler erster Art“ gibt es noch den „Fehler zweiter Art“, sich für die Standardhypothese zu entscheiden, obwohl sie falsch ist. Der braucht uns hier aber nicht so sehr zu kümmern.

⁸ Falls Ihr Bibelprogramm ein paar mehr oder weniger zählt (z. B. 72), mag das daran liegen, daß z. B. theophore Eigennamen mitgezählt sind oder nicht. Es ändert nichts am Ergebnis, demonstriert aber, wie Vorentscheidungen auch hier Einfluß nehmen.

⁹ Es ist hierbei jedoch zu beachten, daß man mit solchen Tests zum einen nichts beweisen

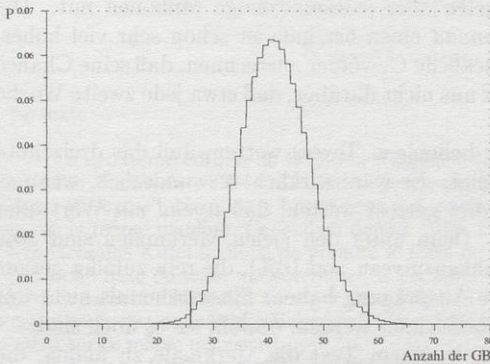


Abbildung 1: Hypergeometrische Verteilung $H(304634, 9647, 1287)$

einheitliche Häufigkeit von GB auszeichnen (im Sinne der Statistik, also mit der gewählten Toleranz), würde Rut in hohem Bogen herausfliegen. – Das unsinnige Ergebnis zeigt, daß die Prämisse falsch war; ein einzelnes Merkmal kann nicht über die Zuordnung von Texten entscheiden. Es war ja auch nur ein Rechenbeispiel.

In diesem Fall war die Frage genau gestellt, und es war nur eine Frage. Der Exeget jedoch stellt viele Fragen an den Text, an seine verschiedenen Elemente und auf seinen verschiedenen Ebenen (Lexeme, Wendungen, Syntax, Inhalt). In den Augen des Statistikers führt er also nicht nur einen „Test“ durch, sondern sehr viele. Und das nicht methodisch-mathematisch, sondern meist intuitiv. Das ist, wie wenn man zwei Würfel solange wirft, bis der benötigte Pasch fällt; oder ob man zwei Dutzend Würfel wirft und dann die passenden Augenzahlen herausucht. Wir könnten Rut ohne Schwierigkeiten wieder repatriieren, wenn wir nach anderen Merkmalen fragen würden. Anders gewendet: wenn man nur lange genug sucht, findet man einzelne Hinweise in die eine wie in die andere Richtung. Von den vielen Merkmalen gibt es sicher einige in extremer Ausprägung und mit zufälliger Ähnlichkeit zu Merkmalen einer anderen Herkunft, man muß sie nur mit Geduld finden.

Ähnlich verhält es sich mit Lexemen, die nur in späten Texten überliefert sind. Gerade wenn diese selten sind, kann der Befund dem Zufall zu verdanken sein, obwohl sie in früheren Zeiten durchaus in Gebrauch waren. Zudem ist hier zu berücksichtigen, daß wahrscheinlich der größte Teil der Bibel erst in exilischer Zeit und später geschrieben wurde (Schichtungseffekte).

kann (im Sinne der Logik) und zum andern nur falsifizieren kann. In diesem Fall haben wir die Gegenthese, Rut sei kein at. Text, nahegelegt, indem wir die These als unwahrscheinlich erkannt haben.

Wichtig ist, Vorher und Nachher zu unterscheiden. Die Wahrscheinlichkeit, daß ich bei „6 aus 49“ im Lotto gewinne (wenn ich denn spielen würde), ist knapp 1 zu 14 Millionen. Es spielen aber jedesmal einige Millionen mit. Die Wahrscheinlichkeit, daß *irgend jemand* einen 6er hat, ist schon sehr viel höher. Nach der Ziehung kann sich der glückliche Gewinner ausrechnen, daß seine Chance nur 1:x war. Dennoch wundern wir uns nicht darüber, daß etwa jede zweite Woche jemand einen 6er hat.

Ich kann *vor* dem Lesen eines beliebigen Textes wetten, daß das dreizehnte Wort im Text eine Auffälligkeit zeigt. Es wäre wirklich verwunderlich, wenn es ausgerechnet so wäre. Ich kann aber getrost wetten, daß *irgend ein* Wort oder Merkmal des Textes auffällig ist. Denn unter den vielen Merkmalen sind 10% (bedingt durch das gewählte Signifikanzniveau von 10%), die rein zufällig gerade im betrachteten Text eine extreme Ausprägung haben. Sie würden als nicht mit der Standardhypothese vereinbar bezeichnet werden, obwohl sie es doch sind.

Die Kunst besteht dann darin, in einem Text die Merkmale zu finden, die zufälligerweise in die gerade verfolgte Hypothese passen, und die übrigen nicht zur Kenntnis zu nehmen. Letztere stehen dann in der vom Kollegen vorgelegten nächsten Arbeit zum Thema.

Man kann nicht eine fixe Zahl angeben, wie viele sprachliche Hinweise auf die Herkunft eines Textes es pro 100 Wörter sein müssen. Darauf kommt es auch nicht an. Wichtig ist, im Hinterkopf zu behalten: je weniger Hinweise man hat, desto eher sind sie durch blanken Zufall erklärbar. Auch muß die absolute Zahl dieser Zufälle größer werden, je umfangreicher (und vielseitiger) der Text wird.

Eine Schwalbe macht noch keinen Sommer, und ein als herkunftstypisch eingestuftes Merkmal weist einen Text noch nicht automatisch zu. Solange man nicht plausibel machen kann, daß eine Herkunft ein „Copyright“ für ein bestimmtes Phänomen besitzt, müssen mehrere Merkmale zugunsten der Hypothese angeführt werden. Insbesondere muß auch sorgfältig nach Hinweisen in andere Richtungen gefahndet werden und nichtsprachliche Kriterien zu Rate gezogen werden.¹⁰ Ferner sind nicht alle Berechnungen, die sich mathematisch durchführen lassen, auch sinnvoll – die Prämissen müssen offengelegt und sorgfältig geprüft werden.¹¹

¹⁰ Vielleicht ist es für unseren Bereich sinnvoller, einfach der These, die den Befund am besten erklärt, den Vorzug zu geben, auch wenn diese nur ein wenig besser erklärt als die anderen Thesen. Dann würden sich aber die hier dargestellten Probleme noch verschärfen: wenn der Text nicht deutlich genug seine Herkunft verrät, so findet man erst recht Belege für die eine wie für die andere Meinung. Am besten ist es bei solchen Fällen, einzugestehen, daß man mit solchen Methoden hier nichts entscheiden kann.

¹¹ Ich danke Herrn Dipl.-Math. Gerald van den Boogart (Freiberg) herzlich für seine fachliche Unterstützung. Für Anregungen, Kritik und Fragen zu diesem Artikel: Andreas.Guenther@Uni-Tuebingen.de