

Computergestützte Netzwerkanalyse in Biologie, Sprach- und Geschichtswissenschaft

Matthis Krischel, Frank Kressing, Heiner Fangerau

Institut für Geschichte, Theorie und Ethik der Medizin
Universität Ulm
Frauensteige 6 (Michelsberg)
89075 Ulm
matthis.krischel@uni-ulm.de
frank.kressing@uni-ulm.de
heiner.fangerau@uni-ulm.de

Abstract: Bei der computergestützten Stammbaum- und Netzwerkanalyse in Biologie, Sprach- und Geschichtswissenschaft treten einige besondere methodische Herausforderungen auf, die sich aus der Quellen- und Interpretationsproblematik ergeben. Mit unserem Beitrag vertreten wir die These, dass die Divergenz der Quellenprobleme allgemein überschätzt wird, während die Interpretationsprobleme allgemein unterschätzt werden und dass bei kritischer Beachtung dieser Prämisse die Entwicklung von Sprachen und Ideen mit ähnlichen rechnergestützten Methoden beschrieben werden können, wie sie in der Biologie zur Überprüfung von Genealogien zum Einsatz kommen.

Spezifische Probleme treten auf, wenn Forscher versuchen, große Datenmengen computergestützt zu analysieren und dann in ihren synchronen und diachronen Verhältnissen darzustellen, denn dies erfordert eine Abstraktion von den tatsächlichen Forschungsgegenständen. In unseren Beispielen zeigen wir, wie Datenauswahl, -verabreichung und -darstellung bestimmte Repräsentationen genealogisch-historischer Prozesse ermöglichen. Wie wir an Stammbäumen und Netzwerken aufzeigen, werden mit Methoden auch Metaphern und implizite Annahmen übernommen. Deshalb sollten sich Forscher bewußt sein, dass die Darstellungsform ihrer Daten die Ergebnisse mitkonstruiert und Interpretationen vorwegnehmen kann.

In der historischen Betrachtung von wissenschaftlichen Transferprozessen stellt die Ermittlung des jeweils transferierten Wissens ebenso eine Herausforderung dar wie die Rekonstruktion der Vermittlungswege. Es ist verlockend, die Analyse komplexer generationen-, kultur- oder disziplinübergreifender Austauschprozesse durch eine möglichst breite Datenbasis und deren computergestützte Verarbeitung argumentativ zu stützen. Allerdings bringt die rechnerbasierte Überprüfung diachroner Austauschprozesse zwischen wissenschaftlichen Disziplinen auch einige Probleme mit sich. Vorbild für eine Anwendung informationstechnischer Methoden zur Untersuchung von Entwicklungs- und Verwandtschaftsverhältnissen ist die rechnergestützte Analyse in der modernen Genetik. Die Übertragung dieser Verfahren auf kultur- und geisteswissenschaftliche Fragestellungen ist allerdings nicht ohne weiteres möglich, da die hier behandelten Gegenstandsbereiche nicht in gleicher Weise kodiert und dekodiert werden können wie beispielsweise der genetische Code. Dennoch erscheint die Adaption rechnerbasierter Verfahren auch nicht unmöglich, da die Quellen- und Interpretationsproblematik in der Biologie durchaus mit der in den Sprach- und Geschichtswissenschaften vergleichbar ist.

Mit unserem Beitrag vertreten wir die These, dass die Divergenz der Quellenprobleme allgemein überschätzt wird, während die Interpretationsprobleme allgemein unterschätzt werden und dass bei kritischer Beachtung dieser Prämisse die Entwicklung von Sprachen und Ideen mit ähnlichen computergestützten Methoden beschrieben werden können, wie sie in der Biologie zur Überprüfung von Genealogien zum Einsatz kommen. Zu diesem Zweck vergleichen wir Stammbaum- und Netzwerkansätze in Biologie, Sprach- und Geschichtswissenschaft und richten unser besonderes Augenmerk auf die Übertragbarkeit von Methoden und Metaphern, die bei der Analyse von Entwicklung in verschiedenen Disziplinen verwandt werden. Die hier beschriebene Methodenübertragung ist deshalb besonders relevant, weil mit ihr quantitative Methoden für die für gewöhnlich eher qualitativ orientierten Geisteswissenschaften nutzbar gemacht werden. Wir werden auf Traditionen quantitativer Forschung in Sprach- und Geschichtswissenschaft hinweisen und auf spezielle Probleme ihrer Anwendung eingehen.¹

1. Gemeinsame Ansätze in Biologie, Sprach- und Geschichtswissenschaft

Sowohl Biologen als auch Linguisten stellen die Entwicklung der Objekte ihrer Wissenschaften in phylogenetischen Bäumen dar. Dies ist zum Teil durch eine gemeinsame Vergangenheit der Disziplinen spätestens seit dem 19. Jahrhundert zu erklären. Darüber hinaus weisen übertragene Metaphern wie „Erbgut“ und „Verwandtschaft“ auf vielfältige Bezüge auch zu anderen Wissenschaften hin [W02, 06]. Die Beweggründe für die aktive Zusammenarbeit von Genetikern und Linguisten seit den 1980er Jahren sind jedoch eher in Arbeiten von Populationsgenetikern wie Luigi Luca Cavalli-Sforza [CF81] [CP88] und Guido Barbujani [Ba91] [Ba97] zu sehen, die sich bemühen, Korrelationen zwischen genetischen und sprachlichen Großgruppen zu finden und in

¹ Hier beziehen wir uns auf ein vom BMBF gefördertes Forschungsprojekt zu den Wechselwirkungen von Natur- und Geisteswissenschaften, in dem an den Universitäten Ulm und Düsseldorf die historisch-genealogischen Entwicklungen in Biologie, Linguistik und Wissenschaftsgeschichte untersucht werden.

Stammbäumen darzustellen. Im Kontext dieser Arbeiten ergibt sich die Frage der Methodenübertragung aus einer wissenschaftlichen Disziplin in eine andere. Das grundsätzliche Problem des Verhältnisses von biologischer und kultureller Entwicklung rückt ebenso in den Fokus wie die Frage, ob beide aus der Perspektive der Geschichtswissenschaft als „Evolution“ beschrieben werden können. Vorbereitet bereits in den 1970er Jahren durch Werke wie Campbells „Evolutionäre Erkenntnistheorie“ [Ca74] oder Stegmüllers Verständnis des Kuhn'schen Ansatzes zur Wissensentwicklung als evolutionär [St75] häufen sich etwa seit den 1980er Jahren Arbeiten zum Themenkomplex der Evolution von Wissenschaft und Technik. Der für unsere Fragestellung relevante Ansatz einer verallgemeinerten Evolutionstheorie ist vor allem von David Hull [Hu88] für die Entwicklung von Wissenschaft angedacht worden.

Seit der Mitte des 19. Jahrhunderts gehen sowohl Biologen als auch Linguisten von einer monophyletischen Entwicklung in ihrer Disziplin aus. Die Idee eines gemeinsamen genetischen Ursprungs aller menschlichen Sprachen bestand schon seit biblischer Zeit.² Im Einklang damit herrschte bis zur frühen Neuzeit die Idee eines gemeinsamen Ursprungs aller menschlichen Sprachen im Hebräischen vor. Daneben entwickelten sich seit dem 16. Jahrhundert Modelle eines polyphyletischen Ursprungs verschiedener Sprachfamilien – zunächst des Semitischen, dann des Finno-Ugrischen [Wi92] [Gy99] [Dr05] und des Indoeuropäischen [Bo47] [Jo86] [Sc08] [Ra18] [Gr19], wobei die Bezeichnung „Indoeuropäisch“ für diese Sprachfamilie im deutschsprachigen Raum ab 1823 von Franz Bopp popularisiert wurde [Bo33] [Hu27], [Kl23],[Dr05]. Seit der Mitte des 19. Jahrhunderts stellten Linguisten die Entwicklung der indoeuropäischen Sprachen zunehmend in einem Stammbaum-Modell dar.³ Monophyletische Bäume, die alle menschlichen Sprachen von einer gemeinsamen Ursprache ableiteten, existierten allerdings auch schon früher.

Für Pflanzen, verschiedene Tiere und den Menschen, also die Untersuchungsgegenstände der Biologie, werden in der christlichen Mythologie⁴ separate Schöpfungsakte postuliert. Dies fand in die Mehrzahl der frühen Evolutionstheorien, wie etwa die von Jean-Baptiste Lamarck, Eingang. Lamarck ließ zwar Veränderung innerhalb einer Art zu, schloss jedoch Artenwandel aus. Charles Darwin hingegen nennt es in der *Entstehung der Arten* „wahrlich eine großartige Ansicht, dass der Schöpfer den Keim alles Lebens, das uns umgibt, nur wenigen oder nur einer einzigen Form eingehaucht hat“ [Da09:691]. Seit der Mitte des 19. Jahrhunderts geht also die Biologie von einer einzigen Entstehung des Lebens und einer Verwandtschaft aller lebenden Wesen miteinander aus. Diese Idee der gemeinsamen Abstammung mündete in ihrer Darstellung in phylogenetischen Bäumen. Während die Unterstellung einer Verwandtschaft von Fischen und Reptilien wenige Gemüter erhitzte, sorgte die gemeinsame Abstammung von Menschen und Affen von gemeinsamen, affenähnlichen Vorfahren für angeregte Diskussionen im 19. Jahrhundert, die bis heute nicht gänzlich beigelegt sind.⁵ Die Verwandtschaft von Arten wurde bis zur Hälfte des 20. Jahrhunderts auf Grund von Homologie der Organe und Körperformen bestimmt. Karl-Ernst von Baer (1792-1876)

² Vgl. den Mythos vom Turmbau zu Babel.

³ Der erste derartige Sprachenstammbaum wird im Allgemeinen August Schleicher zugeschrieben, und stammt von 1853.

⁴ Vgl. Genesis.

⁵ Vgl. z.B. die Debatte zwischen Huxley und Wilberforce von 1860 und den sog. *Dover Trial* über die Lehre von „Intelligent Design“ von 2005.

und Ernst Haeckel (1834-1919) entwickelten auf dieser Grundlage ihre Embryologie und die These, dass die Ontogenese die Phylogenese nachvollzieht. Im Gegensatz dazu werden heute vor allem Ähnlichkeiten von Proteinsequenzen der DNS für Abstammungsvergleiche herangezogen.

In der Wissenschaftsgeschichte und –soziologie hingegen gibt es im Gegensatz zu den Phylogenien der Sprachen und Lebewesen eine lange Tradition, die Genealogie von Wissen mit Hilfe von synchronen und diachronen Verflechtungen der Akteure des Wissenschaftsbetriebes zu beschreiben. Ein Beispiel stellen Ludwig Flecks Denkkollektive dar [F180]. Betrachtet man beispielsweise die gegenseitigen Bezugnahmen von Biologen und Sprachwissenschaftlern im Zusammenhang mit der Problematik der Klassifikation und mit evolutionäre Vorstellungen nach 1850, so wird deutlich dass neues Wissen auf der Basis von und vernetzt mit bereits vorhandenen Kenntnissen entstand. Als Beispiel mag hier der Schwiegersohn Ernst Haeckels, Wilhelm Immanuel Bleek (1827-1875) gelten, zu dessen Buch *Über den Ursprung der Sprache* (1876) Haeckel das Vorwort verfasste und darin eine Methodenübertragung aus der Biologie in die Linguistik anregte. Kommunikation wurde also zu einem zentralen Element wissenschaftlichen Vorgehens [St94]. Diese Kommunikation erstreckte sich dabei nicht nur auf die sich immer weiter ausdifferenzierenden Einzelwissenschaften. Vielmehr umfasste sie auch disziplinübergreifende Austauschprozesse, die zur Übernahme von Metaphern und Methoden der einen Disziplin durch die andere beitrugen [We95]. Die Untersuchung derartiger Kommunikationsprozesse wird genutzt, um die Entwicklung von Ideen, Konzepten und Theorien zu rekonstruieren. Die hier angewandten Methoden sind in der Tradition des Zitationsspiegels [CE17] zu sehen, und insbesondere die Methoden der Zitationsanalyse und der historischen Netzwerkanalyse haben sich hier als fruchtbar erwiesen. [Er97] [FH09]

Während in Biologie und Linguistik also bislang Stammbaumannsätze dominieren, sind Netzwerke in der Geschichtswissenschaft und Soziologie deutlich stärker ausgeprägt, doch wirken auch sie ihrerseits wieder in sie Biologie und Sprachwissenschaft zurück, so dass der angesprochene Methoden- und Wissenstransfer in alle Richtungen erfolgt. Auf der Grundlage unterschiedlichen Quellenmaterials bietet der Netzwerkgedanke Interpretationsalternativen in allen drei Wissenschaften, die sich zum Teil erheblich von herkömmlichen Modellen unterscheiden und besonders in der Biologie und Sprachwissenschaft die Frage der Interpretation von Befunden neu stellen.

2. Quellenproblem

Das Quellenproblem stellt sich in den genannten Disziplinen verschieden dar. In jedem Fall müssen Daten in einer Form erhoben werden oder in eine Form gebracht werden, die sie für eine computergestützte Analyse nutzbar machen. Dies erscheint zunächst in der Genetik als Zweig der Biologie am einfachsten zu sein. Ein DNS-Strang kann im Labor analysiert und interpretiert werden. Zwei oder mehr auf die in ihnen enthaltenen Aminosäuresequenzen reduzierten DNS-Stränge können mit Hilfe des Computers verglichen werden, um ihre Ähnlichkeit festzustellen. Ein großes Problem bildet dabei die Varianz der Aminosäuresequenzen verschiedener Individuen der gleichen Art. Ein Beispiel hierfür ist das Bakterium *Escherichia coli*, bei dem zwei sequenzierte Varianten sich in 30% unterscheiden [DM07], ein Unterschied, der dem Vielfachen der

genetischen Varianz entspricht, der sich zwischen verschiedenen Arten von Säugetieren messen lässt. Hieraus wird deutlich, dass die Kriterien, nach denen zwischen verschiedenen Klassen von Lebewesen unterschieden wird, nicht nur durch die Varianz der DNS-Stränge determiniert wird. Ebenso wie die Laborprozesse, die zu DNS-Sequenzen führen, sind die Unterscheidungsmerkmale von Arten und Klassen durch menschliche Aushandlungsprozesse bestimmt. Dementsprechend beschreiben Janich und Weingarten die Gegenstände der Biowissenschaften als durch Unterscheidung voneinander vom Menschen konstituiert [JW99:78-9]. Dies gilt insbesondere für Lebewesen wie Bakterien, oder ihre Teile, wie Chromosomen oder Gene, die gemäß diesem Ansatz erst als Laborgegenstände entstehen. [JW99:192]

Die Idee des gesetzmäßigen Sprachwandels durch regelmäßige, ausnahmslose Lautverschiebungen wurde Ende des 19. Jahrhunderts vor allem durch die linguistische Schule der Junggrammatiker vertreten, welche bereits Vorläufer im 18. Jahrhundert hatte.⁶ Die Tradition der Anlehnung linguistischer Verfahren an die Naturwissenschaften setzte sich allerdings auch im 20. Jahrhundert in Form der Lexikostatistik und Glottochronologie fort. Grundlage der zeitlichen Datierung des Sprachwandels ist dabei die in der linguistischen Literatur stark umstrittene Annahme einer konstanten Retentionsrate des ererbten Lexikons einer Sprache, weshalb die so genannten Swadesh-Listen seit den 1950er Jahren immer wieder modifiziert und teilweise in kulturangepasste Varianten überführt wurden. In dieser modifizierter Form werden sie auch heute noch angewandt.⁷ Analog zur Anwendung von Netzwerken in der Genetik⁸ werden Verfahren wie die Split Decomposition auch auf die Linguistik angewandt: Lexikalischen Einheiten einer Sprache werden statusabhängige Kodierungen zugewiesen (z.B. unterschiedliche Codes für Kognaten⁹ und Lehnwörter), um so integrative Repräsentationen genetisch-sprachlicher Verwandtschaften in Gestalt von *neighbourNet*- und *splittree*-Modellen zu erzeugen [Pa09] [MM05:145-64], welche sowohl phylogenetische als auch Entlehnungsbeziehungen (horizontale Transmissionen) darstellen können. In der Geschichtswissenschaft hat der Ansatz der Rekonstruktion personengebundener Netzwerke durch Zitations- und Korrespondenzanalysen eine lange Tradition, aber nur begrenzte Verbreitung. Bei Zitationsnetzen wird gezeigt, wen ein Autor in einem bestimmten Werk oder über die Dauer seiner Karriere zitiert. Von besonderem Interesse sind hier Wechsel von Zitationsgewohnheiten, die auf eine Änderung wissenschaftlicher Theorien hindeuten können. In Korrespondenznetzen werden qualitativ und quantitativ die Briefkontakte einer Person oder einer Personengruppe nachvollzogen. Anders als Zitationsnetze enthalten Korrespondenznetzwerke häufig auch Ideen, die ein Autor auf Grund politischer, sozialer oder persönlicher Umstände nicht veröffentlichen wollte oder konnte.¹⁰ Ein prominentes Beispiel ist das „Darwin Correspondence Project“ der Universität Cambridge, das die 14.500 Briefe kategorisiert und erschließt, die Charles Darwin während seiner Lebenszeit mit fast 2.000 Korrespondenten ausgetauscht hat. Die

⁶ Vgl. Lambert ten Kate, z.B. [te23].

⁷ Vgl. Dyen, Kruskal, [Pa09].

⁸ Vgl. [MM05:140]: „[...] detecting and displaying gene transfer and convergence are directly relevant to the analysis of contact-induced change in linguistics, and both are tackled in computer programs based on networks.“

⁹ Bei Kognaten handelt es sich um angenommene natürliche Nachfolger einer gemeinsamen Proto-Form (vgl. [Pa09:408]).

¹⁰ Vgl. [WG62] zu deutsch-russischen Wissenschaftsbeziehungen im 18. Jahrhundert

Aufarbeitung eines solchen Netzwerkes sagt im Idealfall nicht nur etwas über die im Zentrum stehende Person aus, sondern auch über die Wissenschaftskultur seiner Zeit und ihre Protagonisten. Während das Erheben von Daten zur Konstruktion von Zitations- und Korrespondenznetzwerken relativ problemlos möglich ist, ist die Rekonstruktion von Ideen als im Computer zu verarbeitende Daten schwierig. Eine Möglichkeit ergibt sich durch die Erstellung von Feature-Listen, die eine Idee und ihre Bestandteile umfassend charakterisieren. Ein Transfer von Ideen oder ihren Bestandteilen lässt sich durch computergestützte Analyse ihrer Verteilung über verschiedene Dimensionen (diachron, geographisch, etc.) rekonstruieren. Ziel ist es, darzustellen wie ein Netzwerk von Ideen eine wissenschaftliche Theorie bildet und wie Wissenstransfer sich auf die Theoriebildung auswirkt. Ein Modell für auf diese Weise darstellbare Paradigmenwechsel hat Chaomei Chen in Form seines Programmes *Citespace* geliefert [Ch06], das aus Zitationsbeziehungen und Schlagworten multidimensionale Netzwerke erstellt, in denen Wissensverschiebungen durch Cluster, Knoten und Kantenkonstellationen repräsentiert werden.¹¹

Historiker, historische Linguisten und Populationsgenetiker versuchen alle die Interpretation ihrer Datenauswertung durch weitere Quellen abzusichern. Dabei haben es Historiker häufig leichter, da sich per Definition nur mit der Entwicklung innerhalb historischer Zeiträume beschäftigen. Gerade in der Wissenschaftsgeschichte sind aus den untersuchten Zeiträumen gelegentlich Parallelüberlieferungen zu finden, gegen die eine Analyse geprüft werden kann.¹² Dies ist zum Teil auch für Linguisten möglich, die versuchen den Zeitpunkt eines Sprachwandels an gewissen historischen Ereignissen festzumachen. In der Biologie hängt es stark von der Generationsdauer einer untersuchten Art ab, wie spekulativ oder überprüfbar eine Quellenbasis ist. Es ist festzuhalten, dass schon das Erheben der Daten einiges an Interpretation und Standardisierung verlangt.

3. Interpretationsproblem

Ebenso wie das Quellenproblem tritt das Interpretationsproblem in Biologie, Linguistik und Geschichte auf. Schon die Vorauswahl der Daten, die zur Analyse herangezogen werden, ist ein Interpretationsvorgang. Zur rechnergestützten Stammbaum- oder Netzwerkanalyse müssen die Daten außerdem in einer Form erhoben oder in eine Form gebracht werden, in der sie von einem Computer verarbeitet werden können. Hierfür müssen sie in die Form von standardisierten, diskreten Stücken gebracht werden.

In der Molekularbiologie wird die lange Kette der Erbinformationsträger als Reihe von Nukleinbasen dargestellt. Während die Desoxyribose nicht weiter beachtet wird, wird die Folge, in der die in der DNS enthaltenen Aminosäuren Adenin, Cytosin, Guanin und Thymin auftreten als „genetischer Code“ in den Fokus der Betrachtung gerückt und als Abfolge der vier Buchstaben A, C, G und T kodiert. Liegen mehrere dieser Ketten vor, können ihre Längen und einzelne Teilstücke verglichen werden. Da die Erbinformation zum Teil zwischen zwei Varianten einer Art mehr Unterschiede aufweisen kann als

¹¹ Konkrete Anwendungen dieser Methoden für die Geschichtswissenschaft stellen [FH09] vor.

¹² Eine so komfortable Quellenlage ist jedoch stark von der Prominenz und Epoche des Forschungsobjektes abhängig.

zwischen zwei anderen Arten, stellt sich die Frage nach der Relevanz des Artbegriffs in bisheriger Definition und der Anordnung der Arten in einem Stammbaum. Ein weiteres Problem stellt der laterale Gentransfer dar. Bei einigen Bakterien können im Verlauf der Entwicklung bis zu 30% der Gene lateral erworben sein [DM07]. Phylogenetische Bäume sehen keinen lateralen Transfer vor und sind somit gänzlich ungeeignet, Prozesse darzustellen, bei denen dieser auftritt. Dies verlangt eine Adaptation der informationstechnischen Analyseverfahren an die neue Interpretation des Materials. Parallelen für solche Rekonstruktionsprobleme lassen sich ebenfalls in den geisteswissenschaftlichen Disziplinen finden, in denen eine horizontale Transmission kultureller Merkmale, z. B. Entlehnungen in der Linguistik oder Aneignung von Technologien, zu finden ist. [AG05] Netzwerkansätzen (*reticulated networks*) sind eine Methode, horizontalen Transfer und multiplen Ursprung zu erfassen. Diese Interpretationsansätze werden verwendet, um Genealogien sowohl in der Biologie [Hu98] [BM04] [HB06] [DM07] als auch in den Geisteswissenschaften [BHB98] [MW03] [NML06] zu verdeutlichen.

Seit den 1960er Jahren versuchen Linguisten mit den Methoden des statistischen Massenvergleichs von Sprachen [Gr00] und der long range comparison [Do98] [St89] genetische Sprachverwandtschaften bis in eine Zeittiefe von 10-20.000 Jahren zurückzuvorforschen und die sprachliche Vielfalt der Menschheit auf nur wenige sprachliche Großfamilien zurückzuführen. [CF81] [Ru94] Diese Ansätze entsprechen allerdings keineswegs der *mainstream linguistic*. Es ist recht fragwürdig, ob die erhobenen Daten eine Analyse für so lange Zeiträume gestatten, da die Zuverlässigkeit der Aussagen abnimmt, je weiter die Analyse sich zurück erstreckt. Ähnliche Probleme ergeben sich in der Biologie. Obwohl die Zeiträume, über die Biologen Aussagen treffen, deutlich länger sind, erkennen auch sie an, dass ihre Aussagen unzuverlässiger werden, je weiter sie zurückreichen. [DM07]

Insgesamt ergibt sich somit das Bild, dass in der Linguistik mit Lexemen und Morphemen zwar exakt zu bestimmende Analyseeinheiten zur Verfügung stehen, die etwa mit Genomen in der Biologie vergleichbar sind, dass aber die Entwicklung der Struktur [Sa16a] dieser distinkten Einheiten in einem semiotischen System sowohl in Bezug auf die Bestimmung von Protoformen als auch auf die Chronologie ihrer Entwicklung erhebliche und in der Linguistik sehr kontrovers diskutierte Probleme bereitet.

Die Frage, ob ein gerichteter, binär verzweigter Baum die richtige Interpretation von Sprachvergleichen ist, ist nicht neu. Das in der komparativistischen Linguistik des 19. Jahrhunderts entwickelte Modell des rein phylogenetisch dargestellten Sprachwandels wurde schon früh in Zweifel gezogen – zunächst durch die Wellentheorie [Sc72], dann durch den Nachweis von Sprachbündeln, die sich über mehrere Sprachzweige hinweg ausbilden [Ja31], und schließlich durch die Untersuchung von Pidgins, Kreol- und Mischsprachen. [Bi81] [Mo03] Es zeigte sich, dass, ausgehend vom Baum-Modell der Sprachentstehung, auch horizontaler Austausch von Lexemen und morphologischen Strukturen zwischen Sprachen und Sprachfamilien stattfindet, der ebenso wie sprachimmanente Faktoren (Laut- und Bedeutungswandel) zum Sprachwandel beiträgt und durchaus auch Teile des sprachlichen Grundwortschatzes umfassen kann. Dieses differenziertere Bild von Sprachwandelprozessen wird durch moderne Netzwerkanalysen genetischer Sprachverwandtschaften repräsentiert [Hu98] [BM04] [MM05:155], welche einerseits etablierte Ergebnisse der Indogermanistik bestätigen, andererseits aber

auch zeigen, dass parallel zu Prozessen wie Endosymbiose und horizontalem Gentransfer in der Biologie [Sc83] [Me05] [MH98] [DM07] [GW08] auch in der Linguistik lexikalisch-morphologische Entlehnungsprozesse als Ursachen des Sprachwandels zunehmend größeres Gewicht zugebilligt wird.

Das hermeneutische Problem wird in der Geschichtswissenschaft traditionell als am stärksten ausgeprägt angesehen. Dieser Ansicht wollen wir nicht fundamental widersprechen, wir halten jedoch den Unterschied nicht für unvergleichlich groß. Wie in den anderen Disziplinen wird ein Teil der Interpretation auch hier schon durch die Vorauswahl der Quellen vorweggenommen. Der eigentliche Grund für den größeren Interpretationsspielraum ist jedoch, dass Historiker die durch Datensammlung ermittelten Netzwerke selbst wieder interpretieren und in ihren historischen Narrativ einordnen wollen. Zeigt sich z.B. in einem Korrespondenznetzwerk das Abreißen eines Briefkontaktes, so kann das viele Ursachen haben, vom Tod eines Korrespondenten, über persönlichen oder wissenschaftlichen Streit bis zu einem Forschungswechsel eines Korrespondenten.

4. Gemeinsame methodische Herausforderungen

Bei der rechnergestützten Stammbaum- und Netzwerkanalyse treten einige besondere methodische Herausforderungen auf, die sich aus der Quellen- und Interpretationsproblematik ergeben. Die Verwendung gerichteter binär verzweigender Graphen (*directed bifurcating graphs*) als phylogenetische Stammbäume für die Beschreibung evolutionärer Ereignisse zwingen die Daten in eine Baumstruktur, bei der jeder Knoten im Graph über genau zwei Töchter verfügt und alle Vererbungen als Eltern-Kind-Relation beschrieben werden. Die Darstellung von Entwicklung in einem sich binär verzweigenden Baum folgt nicht zwingend aus einer Evolutionstheorie. Der Baum aus Darwins *Herkunft der Arten* (1859) z.B. verzweigt sich nicht binär. Es stellt sich also die Frage, inwiefern die Informationswissenschaft den phylogenetischen Baum als Binärbaum konstituiert hat.

Wie wir zeigen konnten, treten spezifische Probleme auf, wenn Forscher versuchen, große Datenmengen computergestützt zu analysieren und dann in ihren synchronen und diachronen Verhältnissen darzustellen. Zum einen ergibt sich ein Quellenproblem, d.h. die Ausgangsdaten müssen in eine Form gebracht werden, in der sie vom Rechner zu verarbeiten sind. Dies erfordert eine Abstraktion von den tatsächlichen Forschungsgegenständen. In unseren Beispielen haben wir gezeigt, wie Datenauswahl, -verabreichung und -darstellung entweder phylogenetische Baum- oder Netzwerkrepräsentationen genealogisch-historischer Prozesse ermöglichen.

Den Diskursen des Informationszeitalters folgend scheinen sich momentan Netzwerkinterpretationen in weiten Teilen der Biologie, Sprach- und Geschichtswissenschaft zur Beschreibung von Entwicklungsprozessen durchzusetzen. Wir möchten jedoch darauf hinweisen, dass auch die Darstellung von Entwicklung in Netzen nicht frei von potentiellen Problemen ist. Wie wir an Stammbäumen gezeigt haben, werden mit Methoden auch Metaphern und implizite Annahmen übernommen. Auch bei der Darstellung von Daten in Netzwerken sollten sich Forscher deshalb bewusst sein, dass die Darstellungsform ihrer Daten die Ergebnisse mitkonstruiert und Interpretationen vorwegnehmen kann.

Literaturverzeichnis

- [AB49] Arnold, R, Libby, W F 1949: Age Determinations by Radiocarbon Content. Checks with Samples of Known Age. *Science*. Washington DC 1949 no. 12, 23. Dez.
- [AG05] Atkinson, Q.; Gray, R. 2005: Curious Parallels and Curious Connections – Phylogenetic Thinking in Biology and Historical Linguistics. In: *Systematic Biology* 54 (4), p. 513 – 526.
- [Au90] Auroux, S: Representation and the place of Linguistic Change before Comparative Grammar. In: De Mauro, T; Formigari, L (ed.): *Leibniz, Humboldt, and the Origins of Comparativism*. Amsterdam Studies in the Theory and History of Linguistic Science Series III: Studies in the History of the Language Sciences. John Benjamins Publication Company, Amsterdam, 1990, pp. 231-238.
- [Ba91] Barbujani, G. 1991: What do Languages tell us about Human Microevolution? In: *Trends in Ecology and Evolution* 6(5), p. 151 – 156
- [Ba97] Barbujani, G. 1997: DNA Variation and Language Affinities. In: *American Journal of Human Genetics* 61, p. 1011 – 1014
- [BHB98] Barbrook, A.C.; Howe, C.J.; Blake, N.; Robinson, P. 1998: The phylogeny of the Canterbury tales. In: *Nature*. 394:839.
- [Bi81] Bickerton, D: *Roots of Language*. Karoma Publishers, Ann Arbor, 1981.
- [BM04] Bryant, D.; Moulton, V. 2004: Neighbor-Net: Aagglomerativemethod for the construction of phylogenetic networks. *Mol. Biol. Evol.* 21: 255-265.
- [Bo16] Bopp, F: *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache*. Frankfurt/Main, 1816.
- [Bo33] Bopp, F: *Vergleichende Grammatik des Sanskrit, Zend, Armenischen, Griechischen, Lateinischen, Lithauischen, Altslawischen, Gothischen und Deutschen*. Frankfurt/Main, 1833.
- [Bo47] Boxhorn, M Z van: *Antwoord van Marcus Zuerius van Boxhorn op de Vraaghen, hem voorgesteld over de Bediedinge van de tot noch toe onbekende Afgodinne Nehalennia, onlanx uytgegeven. In welke de ghemeine herkomste van der Griecken, Romeinen, ende Duytschen Tale uyt den Scythen duydelyck bewesen, ende verscheiden Oudheden van dese Volckeren grondelijck ontdeckte ende verklaert*. Willem Christiaens van der Boxe, Leyden, 1647 c.

- [Bo47] Boxhorn, M Z van: *Bediedinge van de tot noch toe onbekende Afgodinne Nehalennia, over de duseent ende ettelicke hondert Jahren onder het Sandt begraven, dan onlanx ontdeckt op het Strandt van Valcheren in Zeelandt.* Willem Christiaens van der Boxe, Leyden, 1647 a.
- [Bo47] Boxhorn, M Z van: *Vraagen voorghestelt ende Op Leyden: ghedraaghen aan de Heer Marcus Zuerius van Boxhorn over de Bediedinge van de tot noch toe onbekende Afgodinne Nehalennia, onlanx by Hem uytgegeven.* Willem Christiaens van der Boxe, Leyden, 1647 b.
- [BS05] Baptiste E, Susko E, Leigh J, MacLeod D, Charlebois R L, Doolittle W F: Do orthologous gene phylogenies really support tree-thinking? *BMC Evol. Biol.* 5 (1), 2005, S. 33.
- [Ca01] Cavalli-Sforza, L L: *Genes, Peoples, and Languages.* Penguin Press, London, 2001.
- [Ca74] Campbell, Donald T.: *Evolutionary Epistemology.* In: Schilpp, P.A. (Hrsg.). *The philosophy of Karl R. Popper*, LaSalle, IL, 1974: Open Court: 412-463
- [CE17] Cole, F. J. and Eales, N. B. 1917: "The History of Comparative Anatomy." *Science Progress* 11: 578 -596.
- [CF81] Cavalli-Sorza, L L; Feldman, M W: *Cultural Transmission and Evolution: A quantitative approach.* Princeton, 1981.
- [Ch06] Chen, C. (2006): CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359-
- [CP88] Cavalli-Sforza, L L; Piazza, A; Menozzi, P; Mountain, J: Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences* 85 (August) 1988: S. 6002-6006.
- [Da09] Darwin, C.: *Über die Abstammung der Arten, Zweitausendeins*, Frankfurt, 2009.
- [DM07] Dagan T, Martin W: Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc Natl Acad Sci USA* 104, 2007, S. 870-5.
- [DB92] Dyen, I; Kruskal, J B; Black, P: *An Indoeuropean Classification: A Lexicostatistical Experiment.* Transactions of the American Philosophical Society 82/5. American Philosophical Society, Philadelphia, 1992.
- [Do98] Dolgopolsky, A: *The Nostratic Macrofamily and Linguistic Palaeontology.* McDonald Institute for Archaeological Research, Oxford 1998.
- [Dr05] Driem, G van: Sino-Austronesian vs. Sino-Caucasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory. In (Prasada, Y; Yadava, G Bh; Lohani, R R; Prasain, B; Parajuli, K Hrsg.): *Contemporary Issues in Nepalese Linguistics.* Linguistic Society of Nepal, Kathmandu, 2005, S. 285-338.
- [Er97] Erickson, B. H. 1997. "Social networks and History." *Historical Methods* 30(3): 149-157.
- [FH09] Fangerau, H.; Halling, T.: *Netzwerke: Allgemeine Theorie einer Universalmetapher in den Wissenschaften? Ein transdisziplinärer Überblick,* Transcript Verlag, Bielefeld, 2009.
- [Fl80] Fleck, L.: *Entstehung und Entwicklung einer wissenschaftlichen Tatsache,* Surkamp, Frankfurt 1980.
- [Gi94] Gimbutas, M: *Das Ende Alteuropas. Der Einfall von Steppennomaden aus Südrussland und die Indogermanisierung Mitteleuropas.* Institut für Sprachwissenschaft, Innsbruck 1994.
- [GI95] Gamkrelidze, T; Ivanov, V: *Indo-European and the Indo-Europeans.* Mouton de Gruyter, Berlin, New York 1995.

- [Gr00] Greenberg, J H: Indo-European and Its Closest Relatives: The Eurasiatic Language Family. Volume I: Grammar. Stanford University Press 2000.
- [Gr19] Grimm, J: Deutsche Grammatik. Göttingen, 1819-1834
- [Gr63] Greenberg, J H: The Languages of Africa. Mouton, Indiana University Center, Bloomington, Den Haag 1963.
- [Gr71] Greenberg, J H: The Indo-Pacific Hypothesis. In (Seboek, Th A et al. Hrsg.): Current Trends in Linguistics. Vol. 8: Linguistics in Oceania. Mouton de Gruyter, The Hague, 1971, p. 807-871.
- [Gr87] Greenberg, J H: Language in the Americas. Stanford University Press 1987.
- [GW08] Gould S B, Waller R F, McFadden G I: Plastid Evolution. *Annu. Rev. Plant Biol.* 59, 2008, S. 491-517.
- [Gy99] Gyarmathi, S: *Affinitas linguae Hungaricae cum linguis Fennicae originis grammaticae demonstrata.* Göttingen, 1799.
- [HB01] Henze K; Martin W 2001: How are mitochondrial genes transferred to the nucleus? *Trends in Genetics* 17: 383-387.
- [HB06] Huson, D.H.; Bryant, D. 2006: Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23: 254 - 267.
- [HB95] Henze K; Badr A; Wettern M; Cerff R; Martin W 1995. A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc. Natl. Acad. Sci. USA* 92, 1995, P. 9122-9126.
- [Hu30] Humboldt, W von 1830 [1827]: Ueber den Dualis [gelesen in der Akademie der Wissenschaften am 26. April 1827]. *Abhandlungen der historisch-philologischen Klasse der Königl. Akademie der Wissenschaften zu Berlin aus dem Jahre 1927, 1930*, p. 161-187.
- [Hu88] Hull, D. L.: *Science as a Process – An Evolutionary Account of the Social and Conceptual Development of Science.* The University of Chicago Press, Chicago, 1988.
- [Hu98] Huson, D H: SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*: 14(1), 1998, S. 68-73.
- [Ja31] Jakobson, R 1931: Über die phonologischen Sprachbünde. In (Jakobson, R Hrsg.): *Selected Writings 1991 [repr].* Bd. 1: *Phonological Studies.* 's-Gravenhage, 1931, S. 137-143.
- [Jo86] Jones, W [1786]. The third anniversary discourse. On the Hindus. In (Lehmann, W P Hrsg.): *A Reader in Nineteenth Century Historical Indo-European Linguistics.* Indiana University Studies in the History and Theory of Linguistics, Bloomington, London 1967, S. 7-20.
- [JR99] Jain R, Rivera M C, Lake J A: Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 96 (7), 1999, S. 3801–3806.
- [JW99] Janich P., Weingarten M.: *Wissenschaftstheorie der Biologie,* Wilhelm Fink Verlag, München, 1999.
- [K123] Klapproth, H J von: *Asia Polyglotta.* Paris 1823.
- [Le53] Lees, R B: The basis of glottochronology. In: *Language* 29, 1953, S. 113-127.
- [Ma00] Matisoff, J A 2000: On Sino-Bodic and other symptoms of neosubgroupitis. *Bulletin of the School of Oriental and African Studies* 63(3), 2000, S. 356-369.
- [Ma32] Malte-Brun, C [posthumus]: *Précis de la géographie universelle, ou description de toutes les parties du monde sur un plan nouveau, d'après les grandes divisions naturelles du globe: Nouvelle édition, revue, corrigée, mise dans un nouvel ordre et augmentée de toutes les nouvelles découvertes par J J N Huot* (6 vols). Th Lejeune, Bruxelles 1832.

- [Ma78] Matisoff, J A: Variational Semantics in Tibeto-Burman. The ‘Organic’ Approach to Linguistics Comparison. Occasional Papers of the Wolfden Society on Tibeto-Burman Linguistics vol. VI. Institute for the Study of Human Issues, Philadelphia 1978.
- [Ma96] Martin W 1996: Is something wrong with the tree of life? *BioEssays* 18, 1996, S. 523-527.
- [Me05] Mereschkowsky C: Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. In: *Biol. Centralb.* 25, 1905, S. 593-604.
- [MH98] Martin W, Herrmann RG 1998: Gene transfer from organelles to the nucleus: how much, what happens, and Why? *Plant Physiol.* 118:9-17.
- [MM05] McMahon, A & R: *Language Classification by Numbers*, Oxford University Press 2005.
- [Mo03] Mous, M 2003: *Ma’a or Mbugu. The Making of a Mixed Language*. Amsterdam 2003.
- [MS01] Margulis, L; Sagan, D: Marvellous microbes. *Resurgence* 206, 2001, S. 10–12.
- [MW03] [Minet, Wang. Language Acquisition, Change and Emergence: essays in evolutionary linguistics. City University of Hong Kong Press, 2003.](#)
- [NML06] Nunn, C.L.; Mulder, M.B.; Langley, S. 2006: Comparative methods for studying cultural trait evolution: a simulation study. In: *Cross-Cultural Res.* 40:177-209.
- [Pa09] Pagel, M 2009: Human Language as a culturally transmitted replicator. *Nature Reviews/ Genetics* 10 (June), 2009, S. 405-415.
- [Ra18] Rask, R: *Undersøgelse om det gamle nordiske eller islandske Sprogs Oprindelse*. Kopenhagen, 1818.
- [Ra31] Rafinesque, C: *Mémoires sur l’origine des nation Nègres, ou Introduction à l’histoire des nègres indigènes d’Asie, d’Afrique Polynesie, Amérique et Europe*. Ms, American Philosophical Society, Philadelphia, 1831.
- [Re87] Renfrew, C: *Archaeology and Language: The Puzzle of Indo-European Origins*. Cambridge University Press, 1987 [1990].
- [RH99] Race H L; Herrmann, R G; Martin W: Why have organelles retained genomes? *Trends in Genetics* 15, 1999, S. 364-370.
- [RM01] Rujan T, Martin W 2001: How many genes in Arabidopsis come from cyanobacteria? An estimate from 386 protein phylogenies. *Trends in Genetics* 17, 2001, S. 113-120.
- [Ro04] Roggenbuck, S: *Die Wiederkehr der Bilder. Arboreszenz und Raster in der interdisziplinären Geschichte der Sprachwissenschaft*. Gunter Narr, Tübingen 2004.
- [Ru94] Ruhlen, M: *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford University Press, 1994.
- [Sa16a] Saussure, F de: *Cours de linguistique générale*. Payot, Paris, Lausanne, 1916.
- [Sa16b] Sapir, E: *Time Perspective in Aboriginal American Culture: A Study in Method*. Memoir 90, Anthropological Series No. 13. Geological Survey, Dept. of Mines, Ottawa, 1916.
- [Sa73] Sankoff, D: Mathematical Developments in Lexicostatistical Theory. In (Seboek, Th A Hrsg.): *Current Trends in Linguistics XI*. Mouton, Den Haag, 1973, S. 93-114.
- [Sa79] Saussure, F de: *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*. B. G. Teubner, Leipzig, 1879.
- [Sc08] Schlegel, F W 1808: *Über die Sprache und Weisheit der Indier*. Köln, 1808.
- [Sc50] Schleicher, A: *Die Sprachen Europas in systematischer Übersicht*. Bonn, 1850.

- [Sc53] Schleicher, A: Die ersten Spaltungen des indogermanischen Urvolkes. Allgemeine Monatsschrift für Wissenschaft und Literatur, 1853, S. 786-787
- [Sc72] Schmidt, J: Die Verwandtschaftsbeziehungen der indogermanischen Sprachen. Böhlau, Weimar, 1872.
- [Sc83] Schimper A F W: Über die Entwicklung der Chlorophyllkörner und Farbkörper. Botanische Zeitung 41, 1883, S. 105–14, 121–31, 137–46, 153–62.
- [SK02] Syvanen, M; Kado, C (Hrsg.): Horizontal Gene Transfer. Academic Press, London 2002 (2. Aufl.).
- [SM09] Shijulal, N S; Martin, W; Dagan, T : A network approach to study vertical inheritance and lateral transfer during the evolution of Indo-European languages. Vortrag beim Startup-Meeting des BMBF-Projekts Evolution und Klassifikation in in Biologie, Linguistik und Wissenschaftsgeschichte. Schloss Mickeln, Düsseldorf, 12.06. 2009.
- [St75] Stegmüller, W.: Hauptströmungen der Gegenwartsphilosophie. 5. Aufl., Kröner, Stuttgart, 1975.
- [St89] Starostin, S A 1989: Nostratic and Sino-Caucasian. In (Shevoroshkin, V Hrsg): Explorations in Language Macrofamilies. Brockmeyer, Bochum 1989.
- [St94] Stichweh, R.: Wissenschaft, Universität, Professionen: Soziologische Analysen. Suhrkamp, Frankfurt, 1994.
- [Sw55] Swadesh, M: Towards Greater Accuracy in Lexico-Statistic Dating. In: International Journal of American Linguistics 21, 1955, S. 121-137.
- [Sw59] Swadesh, M: Linguistics as an Instrument of Prehistory. In: Southwestern Journal of Anthropology 15, 1989.
- [te23] ten Kate, L H.: Aenleiding tot de Kennisse van het Verhevene Deel der Nederduitsche Sprake waer in Hare zekerste Grondslag, edelste Kagt, nuttelijkste Onderscheiding, en geregeldste Afleiding overwogen en nagespoort, en tegen het Allervoornaemste der Verouderde en Nog-levende Taelverwanten, als 't Oude Moeso-Gotthisch, Frank-Duitsch, en Angel-Saxisch, beneffens het Hedendaegsche Hoog-Duitsch en Yslandsch, vergeleken word (2 vols.). Rudolph en Gerard Wetstein, Amsterdam, 1723.
- [We02] Weigel, S.: Genealogie und Genetik. Schnittstellen zwischen Biologie und Kulturgeschichte. Akademie, Berlin, 2002.
- [We06] Weigel, S.: Genea-Logik. Generation, Tradition und Evolution zwischen Kultur- und Naturwissenschaft. Fink, München, 2006.
- [We95] Weingart, P.: Grenzüberschreitungen in der Wissenschaft = Crossing boundaries in science. Nomos Verlagsgesellschaft, Baden Baden, 1995.
- [WG62] Winter, E; Grau, C; Hoffmann, P; Lemke, H (Hrsg.): Lomonosov – Schlözer – Pallas. Deutsch-russische Wissenschaftsbeziehungen im 18. Jahrhundert. Quellen und Studien zur Geschichte Osteuropas Bd. XII. Akademie-Verlag, Berlin, 1962.
- [Wi92] Witsen, N 1692: Noord en oost Tatarie, ofte Bondig Ontwerp van eenige dier Landen en Volken, welke voormaels bekend zijn geweest, beneffens verscheide tot noch toe onbekende, en meest nooit voorheen beschreven Tartersche en Nabuurige Gewesten, Landstreeken, Steden, Rivieren, en Platetzen, in de Norder en Oostelyke Gedeelten van Asia en Europa. François Halma, Amsterdam, 2 Bd., 1692.