

Kurze Zusammenfassung zu organisatorischen Fragen

Was man für Beteiligungs nachweise tun muss

Für einen BN wird eine regelmäßige Bearbeitung von Aufgaben vorausgesetzt. Diese bestehen meist in Form eines Kurzessays von maximal einer halben Seite Länge (Seitenränder: rechts, links, oben, unten jeweils 3cm, Schriftgröße 12pt, keine Spezialfancyfonts, entweder Times oder Gnu FreeFonts), können aber auch variieren. Meist werden sich die Aufgaben auf die Lektüre, welche Online bereitgestellt wird, beziehen. Jeder Teilnehmer, der einen BN bekommen möchte, hat das Recht **drei mal** auf die Abgabe der Hausaufgabe zu verzichten. Die Aufgaben werden per Email (mattis.list@uni-marburg.de) an mich gesendet, wobei die Frist jeweils der Montag (23:59) in der Woche vor der Sitzung ist. Wird diese Frist überschritten, gelten die Aufgaben als nicht abgeliefert.

Beim Versenden der Aufgaben sind die folgenden wichtigen Richtlinien zu beachten

- In der Betreffszeile der Email steht hausaufgabe (klein geschrieben), danach folgt ein Bindestrich, und danach die Nummer der Hausaufgabe, wobei einstelligen Nummern der Hausaufgaben eine 0 vorangestellt wird, also beispielsweise
 - hausaufgabe-01
 - hausaufgabe-10
 - usw.
- In der Emails sollte ein unverfänglicher Satz stehen, um zu verhindern, dass sie im Spamordner landet.
- In den Anhang der Email wird die Aufgabe eingefügt. Hierbei muss es sich um eine PDF-Datei handeln! Alles andere (Word, Open Office) wird NICHT berücksichtigt. Die PDF-Datei wird wie folgt benannt:
 - hausaufgabe-NR-nachnamekleinohnesonderzeichen.pdf
 - Konkretes Beispiel: hausafgabe-08-schaeuble.pdfAuch hier werden NUR die Einsendungen akzeptiert, die sich an dieses Format halten!
- In der Datei selbst, welche auf beliebige Art (Word, LaTeX, OpenOffice) erstellt werden kann, steht links in der Kopfzeile immer der Name des Bearbeiters, in der Mitte steht die Nummer der Hausaufgabe (nach dem Schema "Hausaufgabe Nr. X"), und rechts steht das Datum, also beispielsweise:

Johann-Mattis List	Hausaufgabe 8	15.10.2013
<p>Heute war wieder mal schönes Wetter, ich habe keinen Bock gehabt, mich mit der Linguistik zu beschäftigen...</p>		
blablabla		
Literaturangaben		
<ul style="list-style-type: none">• Saussure, F. (1916): Hab vergessen, wie das Buch hieß. Irgendwer: Irgendwo.		

Das Einhalten dieser Formatvorlagen ist extrem wichtig, weil nur so eine realistische Bearbeitungszeit der Aufgaben von meiner Seite gewährleistet werden kann! Abweichungen von den Vorgaben werden daher NICHT toleriert!

Was man für benotete Leistungsnachweise tun muss

Einen benoteten Leistungsnachweis erhält, wer die Abschlussklausur besteht. In dieser Klausur werden verschiedene Aufgaben gestellt, welche im Zeitraum von 90 Minuten bearbeitet werden müssen. Inhaltlich richten sich die Aufgaben nach Übungsaufgaben und Themen, welche im Seminar besprochen werden. Die Klausur selbst wird aus einem Mix von verschiedenen Fragen bestehen, die einerseits inhaltlich (also im Stile von Laberaufgaben aus der Schule), andererseits eindeutig (also im Stile von Fragen beim PISA-Test) beantwortet werden müssen.

Wo es die Literatur und die Übungsaufgaben zum Seminar gibt

Für das Seminar habe ich unter dem Link <http://lingulist.de/quahl> eine Seite errichtet, unter der sowohl die Übungsaufgaben als auch die Texte zu den jeweiligen Sitzungen heruntergeladen werden können. Bei dieser Seite handelt es sich um eine geschützte Seite. Benutzername und Passwort werden im Seminar genau JETZT bekanntgegeben.

Wie das Seminar strukturiert ist

Derzeit plane ich, die Sitzungen wie folgt ablaufen zu lassen:

- Besprechung der Übungsaufgaben und Vorstellen von Musterlösungen
- Besprechung der neuen Themen
- Bearbeitung von Übungsaufgaben oder allgemeine Diskussion

Wie lang die unterschiedlichen Blöcke jeweils dauern wird von der jeweiligen Thematik abhängen.

„Once Upon a Time...“

1 Allgemeine Punkte vorweg

Sprachvergleich

Wenn man sich mit den historischen Aspekten der historischen Sprachwissenschaft befasst, so stellt sich zu allererst die Frage, was man eigentlich unter historischer Sprachwissenschaft versteht. Fasst man den Rahmen sehr eng, so beginnt die Geschichte der historischen Sprachwissenschaft eigentlich erst ab dem 19. Jahrhundert, parallel mit der Entwicklung der Evolutionstheorie und der Erforschung der Geschichte der Erde in der Geologie, denn davor wurden Sprachen nicht aus einem explizit historischen Blickwinkel betrachtet. Fasst man den Rahmen weiter, so kann man die Frage nach den Wurzeln des historischen Denkens in Bezug auf Sprachen stellen, und erforschen, wie Sprachen im Laufe der Geschichte verglichen wurden, wann genau die „moderne“ historische Sprachwissenschaft begann, und worin die entscheidenden Unterschiede zur „vormodernen“ Sprachforschung liegen.

In der Sprachtypologie werden Sprachen generell nicht-historisch betrachtet. Weshalb lässt sich die historische Sprachbetrachtung dennoch auch in der synchronen Typologie nicht gänzlich ignorieren?

Periodisierung

Eine gängige grobe Periodisierung der Geschichte der historischen Sprachwissenschaft teilt diese in drei Phasen ein: (a) Antike und Mittelalter, (b) die Zeit vom 16. bis zum 18. Jahrhundert, und (c) den Beginn der historischen Linguistik im 19. Jahrhundert.

Was mag der Grund sein, dass gerade die Zeit vom 16. bis zum 18. Jahrhundert als eine gesonderte Phase angesetzt wird? Am 26. Juli 1581 schlug ein Meteorit in Thüringen ein (Wikipedia: "1581"). Welche weiteren "makrohistorischen" Ereignisse fallen in diese Zeit?

2 Antike und Mittelalter

Kein Bock auf Fremdsprachen

Die Antike, vor allem das alte Griechenland, zeichnet sich durch ein nahezu vollständiges Desinteresse an fremden Sprachen und Sprachvergleich aus (Pedersen 1983: 10, Robins 1973: 5). Etymologische Betrachtungen wurden zuweilen ange stellt, waren jedoch meist mit der Absicht verbunden, die „wahre“ Bedeutung der Wörter zu finden (Diderichsen 1974: 278-280), und nicht mit einem Interesse an der Bedeutungsentwicklung sprachlicher Zeichen an sich.

Woran mag es liegen, dass die so geistverliebten Griechen und Römer kein Interesse zeigten, sich mit dem Vergleich von Sprachen zu befassen?

Die **physei-thesei**-Debatte

Grundlage der *physei-thesei*-Debatte, die ausführlich in Platons *Krátulos* beschrieben wird, ist die Frage nach der Beziehung von Zeichenform und Zeichenbedeutung. Der grundsätzliche Streitpunkt der *physei-thesei*-Debatte ist dabei, ob die Struktur der Sprache durch die Natur (altgr. φύση *physei*) oder durch die Konvention (altgr. θέση *thesei*) bestimmt sei. Die Beantwortung dieser Frage hat unmittelbare Folgen für die Etymologie: Wenn die Beziehung zwischen Bezeichnendem und Bezeichnetem natürlich ist, so stellt die ursprüngliche Bedeutung der Wörter deren „wahre“ Bedeutung dar. Die Etymologie könnte also also den wahren Kern eines Wortes enthüllen. Wenn die Beziehung nicht natürlich ist, so sollte sie schleunigst aus allen Diskussionen verbannt werden, in denen die „wahre Bedeutung“ als Argument für eine bestimmte Überzeugung angeführt wird.

Es heißt oft, die *physei-thesei*-Debatte wurde in gewisser Weise von Ferdinand de Saussure (1857 -- 1913) wieder aufgegriffen, als er sein berühmtes Postulat von der „Arbitrarität des sprachlichen Zeichens“ machte, und der Debatte zugunsten der „Konvention“ ein Ende bereitete. Unten sind zwei Zitate, eins von Saussure selbst, und eins von Roman Jakobson (1896 -- 1982) abgedruckt. Bestätigen diese Zitate die Behauptung, dass Saussure sich ausnahmslos auf die Seite der „Konventionalisten“ geschlagen habe?

Das Wort beliebig erfordert hierbei eine Bemerkung. Es soll nicht die Vorstellung erwecken, als ob die Bezeichnung von der freien Wahl der sprechenden Person abhinge [...]; es soll besagen, daß es unmotiviert ist, d.h. beliebig im Verhältnis zum Bezeichneten, mit welchem es in Wirklichkeit keinerlei natürliche Zugehörigkeit hat. (Saussure 1916: 80)

Der Zusammenhang zwischen einem **signans** und einem **signatum**, den Saussure willkürlichweise arbiträr nennt, ist in Wirklichkeit eine ge-wohnheitsmäßige, erlernte Kontiguität, die für alle Mitglieder der gegebenen Sprachgemeinschaft obligat ist. Aber neben dieser Kontiguität behauptet sich auch das Ähnlichkeitsprinzip, **la ressemblance**. [Dieses] Prinzip [spielt] eine gewaltige Rolle in der Frage der Derivation, in der Frage der Wortsippen, wo die Ähnlichkeit der Wörter einer gemeinsamen Wurzen so entscheidend ist, und wo man schon gang und gar nicht mehr vom Willkürlichen sprechen darf. (Jakobson 1962[1971]: 272f)

Das hebräische Paradigma

Die Verbreitung des Christentums führte zur Literalisierung vieler bis dahin nicht verschriftlichter Sprachen und weckte damit auch allmählich das Interesse der Europäer an diesen Sprachen (Pedersen 1972: 4). Die Verbreitung des Christentums führte gleichzeitig dazu, dass das Hebräische entsprechend der biblischen Geschichte von *Turm* zu *Babel* als die älteste Sprache, oder „Mutter aller Sprachen“ angesehen wurde (Arens 1955: 72-80, Klein 1999, Klein 2004).

Unten ist ein Zitat aus Isidors *Etymologiae sive origines* wiedergegeben, dass als frühes Beispiel für das „hebräische Paradigma“ dienen kann. Inwiefern mag Isidors Argumentation typisch sein für die damalige Zeit?

„Die lateinischen und griechischen Buchstaben sind offensichtlich aus den Griechischen entstanden. Dort nämlich steht das Aleph an erster Stelle. Aus diesem wurde dann aufgrund ähnlicher Aussprache von den Griechen das Alpha abgeleitet, und daraufhin von den Lateinern das A. Der Übersetzer hat nämlich aufgrund des ähnlichen Klangs der anderen Sprache die Buchstaben erschaffen,

woraus ersichtlich wird, dass die hebräische Sprache die Mutter aller Sprachen und Schriften ist." (*Etymologiae*: 1.3.4)¹

Sprachbeziehungen

Während vor dem Beginn der historischen Linguistik eine genealogische Perspektive auf Sprachbeziehungen nur sporadisch angenommen wurde, waren sich die Gelehrten recht früh der Tatsache bewusst, dass Sprachen einander auf weit andere Art beeinflussen können. Dies ist sicherlich der Tatsache geschuldet, dass die Folgen von Sprachkontakt sehr viel einfacher zu erkennen sind als die Folgen von Sprachverwandtschaft, nicht nur für diejenigen, die sich besonders für Sprachen interessieren, sondern auch für die „normalen“ Sprecher, die in Kontakt mit Sprechern anderer Sprachen sind. Ähnlichkeiten zwischen Sprachen als Folge von Sprachkontakt und nicht als Folge von genetischer Sprachverwandtschaft anzusehen, war eine Perspektive, die von der Mehrheit der Gelehrten in der Antike bis hinein ins Mittelalter eingenommen wurde (Allen 1953: 57). Sprachbeziehungen wurden also weitestgehend als *derivationell* angesehen.

Was mag der Grund sein, dass derivationelle Sprachbeziehungen problemlos von Gelehrten erkannt und anerkannt wurden, während man genealogische Beziehungen meist ignorierte? Inwiefern können die beiden unten wiedergegebenen Zitate in diesem Zusammenhang als „typisch“ für antikes Denken bezeichnet werden?

Ich denke nämlich, daß die Hellenen, zumal die in der Nähe der Barbaren wohnenden, gar viele Worte von den Barbaren angenommen haben. [...] Wenn nun einer aus der hellenischen Sprache erklären will, inwiefern diese mögen richtig gebildet sein, und nicht aus jener, der das Wort wirklich angehört, so siehst du wohl, daß er nichts schaffen wird. (*Krátulos*: 409d-e)

In continuation, that I may follow the course which I prescribed to myself, let me repeat that words are either Latin or foreign. Foreign words, like men, and like many of our institutions, have come to us, I might almost say, from all nations. [...] But this division of mine is intended to refer chiefly to the Greek language, for it is from thence that the Roman language is, in a very great degree, derived, and we use even pure Greek words where our own fail, as they also sometimes borrow from us. (*Institutio oratoria*: I.5.55-58)

3 Vom 16. bis zum 18. Jahrhundert

Sprachvergleich

Ab dem 16. Jahrhundert wurden erste Grammatiken semitischer Sprachen in Europa veröffentlicht. Hebräisch wurde dadurch neben Latein und Griechisch zur dritten GelehrtenSprache. Beflügelt durch die Erfindung der Druckkunst rasch anwachsende grammatische und lexikalische Sprachmaterial (Pedersen 1972: 6), versuchten viele Forscher nun auch, den etymologischen Nachweis für das Alter des Hebräischen und die Abstammung aller Sprachen von diesem zu erbringen.

¹Meine Übersetzung, Originaltext: „Litterae Latinae et Graecae ab Hebraeis videntur exortae. Apud illos enim prius dictum est aleph, deinde ex simili enuntiatione apud Graecos tractum est alpha, inde apud Latinos A. Translator enim ex simili sono alterius linguae litteram condidit, ut nosse possimus linguam Hebraicam omnium linguarum et litterarum esse matrem.“

Begriffe aus der hebräischen Grammatik beeinflussten auch die Sprachforschung. So scheint auch das Konzept der „Wurzel“, dem als Vergleichsgegenstand in der historisch-vergleichenden Sprachwissenschaft nach wie vor zentrale Bedeutung zukommt, auf die Erforschung der semitischen Sprachen und der traditionellen semitischen Grammatiken zurückzugehen (Campbell and Poser 2008: 95).

Die Art und Weise, wie der Nachweis der Verwandtschaft des Hebräischen mit allen anderen erbracht wurde, war allerdings – aufgrund der Tatsache, dass er ja durch die biblische Geschichte ohnehin schon „bewiesen“ war – noch sehr weit von dem entfernt, was man heutzutage wissenschaftlich nennen würde. Einzelne Wörter wurden aus Wortlisten herausgepickt und verglichen, wobei meist schon eine geringe Ähnlichkeit Vokalen und Konsonanten ausreichte, um die These der Sprachenharmonie als bestätigt anzusehen.

Unten sind einige Beispielhafte Zitate zu Wortvergleichen, sowie Wortvergleiche selbst in Tabellen wiedergegeben. In der vorherigen Sitzung wurden kurz Beispiele für „richtige“ Wortvergleiche gegeben. Worin besteht der grundlegende Unterschied?

Hebräisch	Deutsch	Bedeutung
ʔēm	Am	„Mutter“
māhār	Morn	

[E]inige hebräische Wörter sind (auch) in unserer deutschen Sprache verblieben. (Münster 1523: 27f)²

Was die Ableitung der Wörter durch Addition, Subtraktion, Transposition und Inversion der Buchstaben anlangt, so steht fest, dass man so verfahren kann und muss, wenn wir bedenken, dass die Hebräer von rechts nach links schreiben und die Griechen und die übrigen von links nach rechts. (Guichard 1606, zitiert nach Arens 1955: 76)

Webb-Chinesisch	Mittelchinesisch	Griechisch	Bedeutung
keūen	k ^h wen ²	kūōn	„Hund“
niu gin	ŋjo ² -nin	gunē	„Frau“
yuen xue	ŋjon ³ -yæwk	fa-iēn	„wollen (lernen, sprechen)“

But was not γυνη compounded of *niu gin* under the customary licence of transposition? (Webb 1787: 50)

Die Skythenhypothese

Im Zuge der etymologischen Forschungen unter dem hebräischen Paradigma stießen verschiedene Forscher zwangsläufig auch auf „tatsächliche“ Ähnlichkeiten zwischen den europäischen Sprachen. Es war in bewisser Weise also nur eine Frage der Zeit, bis die Forscher bemerkten sollten, dass auch ohne abstruse Vergleiche auffällige Ähnlichkeiten zwischen verschiedenen Sprachen entdeckt werden konnten, die weit zwingender waren, als die Vergleiche mit dem Hebräischen. Die Skythenhypothese (vgl. insbes. Metcalf 1974: 234-240, Muller 1986: 9-12, Campbell and Poser 2008: 18-23), die in diesem Zusammenhang besondere Beachtung verdient, kann dabei als Vorläufer der „Indogermanen-Hypothese“ angesehen werden.

²Meine Übersetzung, Originaltext: „[...] aliquot Hebraeas voces in nostram Germanicam receptas esse linguam.“

Im unten wiedergegebenen Zitat wird der Skythenhypothese ein ‚abstrakter‘ Charakter zugewiesen. Lässt sich der Grund für diese Argumentation aus dem Zitat selbst erschließen? Wenn ja, worin besteht er? Wenn nein, worin könnte er bestehen?

Since the early 17th century there had been formulated, first in the Netherlands, a theory of the common origin of the main languages of Europe from the somewhat mythical language of the Scythians. This principle of a linguistic unity of the European languages, classical and modern, reaching far into the East, soon gained the abstract character of a prototype. (Muller 1986: 10)

4 Der Beginn der historischen Linguistik

Lautkorrespondenzen

Lautkorrespondenzen zwischen miteinander verwandten Sprachen wurden bereits vor dem 19. Jahrhundert für eine Reihe von Sprachen entdeckt. Ihre Bedeutung für den Nachweis von Sprachverwandtschaft erlangten sie jedoch erst durch die Werke von Jacob Grimm (1785 – 1863, vgl. Grimm (1822)), und Rasmus Rask (1787 – 1832, vgl. Rask 1818) zu Beginn des 19. Jahrhunderts. Mit den Lautkorrespondenzen ist gleichzeitig ein Übergang von einer auf oberflächlichen Ähnlichkeiten beruhenden „phänotypischen“ Betrachtungsweise hin zu einer „genotypischen“ Betrachtungsweise von Kognaten verbunden (vgl. Lass 1997: 130). Dies heißt, dass Ähnlichkeit nicht mehr als Kriterium für das Auffinden von kognaten Wörtern in miteinander verwandten Sprachen verwendet wird, sondern Regelmäßigkeit der Korrespondenz.

Das unten wiedergegebene Zitat zur Wichtigkeit der Regelmäßigkeit der Lautkorrespondenzen ist in seiner Radikalität typisch, auch für heutige historische Linguistik. Worin besteht die Radikalität, und warum wird sie von vielen historischen Linguisten befürwortet?

Das bedeutet, dass bei Vergleichen der Form und bedingt der Vorzug gegeben werden muss. Wenn zwei Formen sich genau - oder den Regeln nach - entsprechen, wiegt das auch gewisse Abweichungen in der Bedeutung auf. Szemerényi 1970: 15f

Bäume und Sterne

Nach Jacob Grimm und Rasmus Rask war August Schleicher (1821 – 1868) eine der einflussreichsten Persönlichkeiten der Linguistik des 19. Jahrhunderts. Er prägte durch seine Arbeiten wie kein anderer den wissenschaftlichen Stil der Forschung in der historischen Sprachwissenschaft. Schleicher entwickelte das Verfahren der **linguistischen Rekonstruktion**, indem er **Protoformen** an die Stelle von **Lautkorrespondenzen** setzte und damit die „Grundlosigkeit der noch immer nicht ganz verschollenen Annahme, daß auch die nicht indischen indogermanischen indogermanischen Sprachen vom altindischen (Sanskrit) abstammen“ (Schleicher 1861[1866]: 8) untermauerte. Schleicher etablierte die neue Fassung der Verwandtschaftsmetapher, die ihren Ausdruck in dem berühmten **Stammbaummodell** fand, welches er erstmals im Jahr 1853 publizierte (Schleicher 1853a, Schleicher 1853b).

Für Schleichers Ansichten finden sich oben viele Zitate. Welche Überschrift könnte man jedem von diesen am besten geben, und wie könnte man Schleichers Auffassung von „Vererbung“, „Wandel“ und „Ursprung“ charakterisieren?

Der Verfall ist wirklich ein allmählicher wie die geschichtliche Entwicklung, er ist in Perioden theilbar, wie diese, je nach dem grösseren oder geringeren Grade der Entfernung vom Ursprünglichen und er verläuft bei allen Sprachen in analoger Weise, wie die Geschichte. Aus letzterem Satze folgt die Möglichkeit und der Vortheil einer vergleichenden Behandlung der Sprachengeschichte. (Schleicher 1848: 25)

Dennoch würde es eben so falsch sein das Lettische eine slawische Sprache zu nennen, als das Ossetische mit den Mingrelischen, Suarischen, u.s.w. zu einer Klasse zu rechnen. Dass dergleichen Färbungen von einer zur anderen Sprache sich verpflanzen können, scheint daher durch die Erfahrung gerechtfertigt und ist auf diese Erscheinung bei der Eintheilung der Sprachen gebührende Rücksicht zu nehmen. (ibid.: 30)

Das Wichtigste und Entscheidende aber ist, dass das Ossetische die das Iranische charakterisierenden, es von den andern indogermanischen Familien unterscheidenden Lautgesetze aufzuweisen hat. (ibid.: 67)

Von Sprachsippen, die uns genau bekannt sind, stellen wir eben so Stammbäume auf, wie diess Darwin (S. 121) für die Arten von Pflanzen und Thieren versucht. (Schleicher 1863: 14)

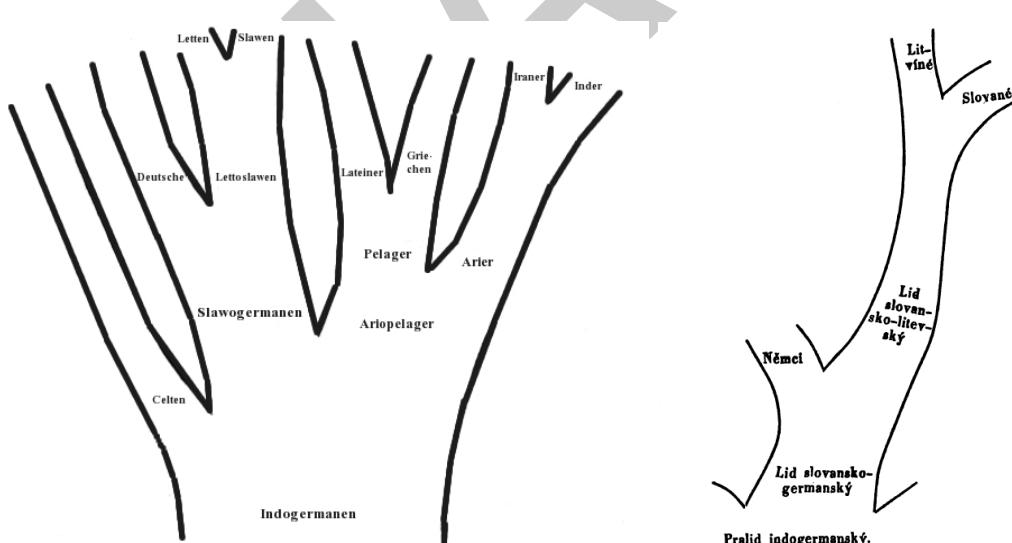


Abbildung 1: Sprachstammbäume in Schleicher (1853a und 1853b)

Literatur

- Allen, W. S. (1953). "Relationship in comparative linguistics". In: *Transactions of the Phyleological Society*, 52–108.
 Arens, H. (1955). *Sprachwissenschaft. Der Gang ihrer Entwicklung von der Antike bis zur Gegenwart*. Freiburg: Alber.

- Campbell, L. and W. J. Poser (2008). *Language classification: History and method*. Cambridge: Cambridge University Press.
- Diderichsen, P. (1974). "The foundation of comparative linguistics: Revolution or Continuation?" In: *Studies in the history of linguistics: Traditions and paradigms*. Ed. by D. H. Hymes. Indiana University studies in the history and theory of linguistics. Bloomington: Indiana University Press, 277306.
- Grimm, J. (1822). *Deutsche Grammatik*. 2nd ed. Vol. 1. Göttingen: Dieterichsche Buchhandlung. Google Books: MnsKAAAAIAAJ.
- Guichard, E. (1606). *L'Harmonie étymologique des Langues: Hebraïque, Chaldaïque, Syrique, Greque, Harmonie étymologique des langues hébraïque, chaldaïque, syriaque, grecque, latine, françoise, italienne, espagnole, allemande, flamande, angloise, où se démontre que toutes les langues sont descendues de l'hébraïque*. Paris: Le Noir.
- Jakobson, R. (1962 [1971]). *Zeichen und System der Sprache. Diskussionsbeitrag*. In: *Selected writings*. Vol. 2: *Word and language*. Berlin: Mouton, 272–279.
- Klein, W. P. (1999). "Die ursprüngliche Einheit der Sprachen in der philologisch-grammatischen Sicht der frühen Neuzeit". In: *The Language of Adam*. Proceedings of a conference held at the Herzog August Bibliothek (Wolfenbüttel, May 30–31, 1995). Ed. by A. P. Coudert. Wolfenbütteler Forschungen 84. Wiesbaden: Harrassowitz, 25–56.
- (2004). "Was wurde aus den Wörtern der hebräischen Ursprache? Zur Entstehung der komparativen Linguistik aus dem Geist etymologischer Spekulation". In: *Gottes Sprache in der philologischen Werkstatt. Hebraistik vom 15. bis zum 19. Jahrhundert*. Proceedings of the Symposium "Die Geburt der Philologie aus dem Geist der Hebraistik" (Wittenberg, Oct. 6–6, 2002). Ed. by G. Veltri and G. Necker. Studies in European Judaism 11. Leiden: Brill, 3–23.
- Lass, R. (1997). *Historical linguistics and language change*. Cambridge: Cambridge University Press.
- Metcalf, G. J. (1974). "The Indo-European hypothesis in the sixteenth and seventeenth centuries". In: *Studies in the history of linguistics: Traditions and paradigms*. Ed. by D. H. Hymes. Indiana University studies in the history and theory of linguistics. Bloomington: Indiana University Press, 233–257.
- Muller, J.-C. (1986). "Early stages of language comparison from Sassetti to Sir William Jones (1786)". In: *Kratylos* 31, 1–31.
- Münster, S. (1523). *Dictionarium Hebraicum, nunc primum editum et typis excusum, adiectis Chaldaicis vocabulis non parum multis*. Latin. apud Frobenium. Gallica: ark:/12148/bpt6k929970.
- Pedersen, H. (1972). *The discovery of language. Linguistic science in the nineteenth century*. 5. print. Vol. 40. Indiana University studies in the history and theory of linguistics. Bloomington: Indiana University Press.
- (1983). *A Glance at the History of Linguistics, with Particular Regard to the Historical Study of Phonology*. John Benjamins.
- Rask, R. K. (1818). *Undersögelse om det gamle Nordiske eller Islandske sprogs oprindelse* [Investigation of the origin of the Old Norse or Icelandic language]. Copenhagen: Gyldendalske Boghandlings Forlag. GoogleBooks: cWgJAAAQAAJ; English translation: - (1993). *Investigation of the origin of the Old Norse or Icelandic language*. Trans. by N. Ege. Travaux du Cercle Linguistique de Copenhague 26. Copenhagen: The Linguistic Circle of Copenhagen.
- Robins, R. H. (1973). "The history of language classification". In: *Diachronic, areal and typological linguistics*. Ed. by H. M. Hoenigswald and R. H. Langacre. Current Trends in Linguistics 11. The Hague and Paris: Mouton, 3–41.
- Saussure, F. de (1916). *Cours de linguistique générale*. Ed. by C. Bally. Lausanne: Payot; German translation: - (1967). *Grundfragen der allgemeinen Sprachwissenschaft*. Trans. from the French by H. Lommel. 2nd ed. Berlin: Walter de Gruyter & Co.
- Schleicher, A. (1848). *Zur vergleichenden Sprachgeschichte*. Bonn: König.
- (1853a). "Die ersten Spaltungen des indogermanischen Urvolkes". In: *Allgemeine Monatsschrift für Wissenschaft und Literatur* 3, 786–787.

- Schleicher, A. (1853b). "O jazyku litevském, zvláště na slovanský. Čteno v posezení sekcí filologické královské Společnosti Nauk dne 6. června 1853". In: *Časopis Českého Museum* 27, 320–334.
- (1863). *Die Darwinsche Theorie und die Sprachwissenschaft. Offenes Sendschreiben an Herrn Dr. Ernst Haeckel*. Weimar: Hermann Böhlau. ZVDD: urn:nbn:de:bvb:12-bsb10588615-5.
- (1861 [1866]). *Compendium der vergleichenden Grammatik der indogermanischen Sprache. I: Kurzer Abriss einer Lautlehre der indogermanischen Ursprache, des Altindischen (Sanskrit), Alteranischen (Altbaktrischen), Altgriechischen, Altitalischen (Lateinischen, Umbrischen, Oskischen), Altkeltischen (Altirischen), Altslawischen (Altbulgarischen), Litauischen und Altdeutschen (Gotischen)*. 2nd ed. Weimar: Böhlau.
- Szemerényi, O. (1970). *Einführung in die vergleichende Sprachwissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Webb, D. (1787). *Some reasons for thinking, that the Greek language was borrowed from the Chinese*. In: *Notes on the Grammatica Sinica of Mons. Fourmont*. London: Dodsley.
- Wikipedia. Wikipedia. The Free Encyclopedia. URL: <http://en.wikipedia.org>; German edition: Wikipedia. Die freie Enzyklopädie. URL: <http://de.wikipedia.org>.

Quellen

<i>Etymologiae</i>	<i>Etymologiae sive origines</i> [Etymologies or origins] (ca. 623 AD). By Isidorus Hispalensis (ca. 560–636 AD); Critical edition: W. M. Lindsay, ed. and comm. (1911). <i>Isidori Hispalensis episcopi etymologiarvm sive originvm libri XX</i> [The twenty books on the origins and etymologies of the bishop Isidorus Hispalensis]. Vol. 1: <i>Libri I-XX</i> [Books 1 - 10]. Scriptorum classicorum bibliotheca Oxoniensis. Oxford: Clarendon Press.
<i>Institutio oratoria</i>	<i>Institutio oratoria</i> [Institutes of oratory] (ca. 95 AD). By Quintilian (35–100); English translation: Quintilian (2006). <i>Institutes of oratory</i> . Ed. by L. Honeycutt. Trans. by J. S. Watson. Iowa. URL: http://honeyell.public.iastate.edu/quintilian/ .
<i>Krátylos</i>	Krátylos (after 399 BC). By Plátón (424–348 BC); English translation: <i>Cratylus</i> (1921). In: <i>Plato in twelve volumes</i> . Vol. 4: <i>Cratylus. Parmenides. Greater Hippias. Lesser Hippias</i> . Trans. from the Greek by H. N. Fowler. London: William Heinemann Ltd.; German translation: <i>Kratyllos</i> (2001). In: <i>Platon. Werke in acht Bänden. Griechisch und Deutsch</i> . Vol. 3: <i>Phaidon. Das Gastmahl. Kratylos</i> . Ed. by G. Eigler. Red. by D. Kurz. Greek text by L. Robin and L. Méridier. Trans. from the Greek by F. Schleiermacher. Darmstadt: Wissenschaftliche Buchgesellschaft.

„Die Basis entscheidet!“

1 Gegenstand

1.1 Sprachen

Was ist eine Sprache?

Was als Sprache gilt, d. h. welche Sprechtraditionen zu einer Sprache gezählt werden, hängt nicht von rein linguistischen, sondern insbesondere auch von sozialen und kulturellen Kriterien ab (Barbour and Stevenson 1998: 8). Demzufolge geht man bspw. davon aus, dass die Menschen in Shanghai, Peking und Meixian alle Dialekte des „Chinesischen“ sprechen, während die Menschen in Skandinavien eine der Sprachen „Norwegisch“, „Schwedisch“ oder „Dänisch“ sprechen. Dies bedeutet jedoch nicht, dass sich die chinesischen Varietäten weniger voneinander unterscheiden als die skandinavischen, wie Tabelle 1 deutlich macht:

Beijing Chinese	1	iou ²¹	i ⁵⁵	xuei ³⁵	pei ²¹ fəŋ ⁵⁵	kən ⁵⁵	t ^h ai ⁵¹ iəŋ ¹¹	tʂəŋ ⁵⁵	tsai ⁵³	naɔ̯ ⁵¹	tʂəŋ ⁵⁵ luən ⁵¹
Hakka Chinese	1	iu ³³	it ⁵⁵	pai ³³ a ¹¹	pet ³³ fʊŋ ³³	t ^h uŋ ¹¹	jɪt ¹¹ t̥eu ¹¹	hɔk ³³	e ⁵³	au ⁵⁵	
Shanghai Chinese	1	fi ²²		t ^h ā ⁵⁵ tsj ²¹	po? ³ fɔŋ ⁴⁴	ta? ⁵	t ^h a ³³ fiā ⁴⁴	tsəŋ ³³ hɔ ⁴⁴		lə? ¹ lə ²³ tsa ⁵³	
Beijing Chinese	2	ʂei ³⁵		də ⁵⁵	pəŋ ³⁵ lɪŋ ²¹	ta ⁵¹					
Hakka Chinese	2	man ³³	ʃin ¹¹		kʷɔ ⁵⁵	vɔi ⁵³					
Shanghai Chinese	2	sa ³³	ʃin ⁵⁵	fia? ²¹	pəŋ ³³ zj ⁴⁴	du ¹³					
Norwegian	1	nu:ravɪn ¹ ŋ	ɛ	sʊ:lŋ					kraŋlət	ɔm	
Swedish	1	nu:ðanvɪndən	ɛ	sʊ:lən		ty:stadə	ən gɔj			ɔm	
Danish	1	nøðanven ¹ ŋ	ʌ	so: ¹ ŋ	k ^h ʌm	en ^g ɑŋ	i sðøið ²			ʌm ²	
Norwegian	2	vem	a	dem	smj	va:	dŋ		stæfkəstə		
Swedish	2	vem	av	dɔm	sɔm	va	dŋ		starkast		
Danish	2	vem ²	a	bɔm	ɖ	va	dŋ		sðæʌŋəsðə		

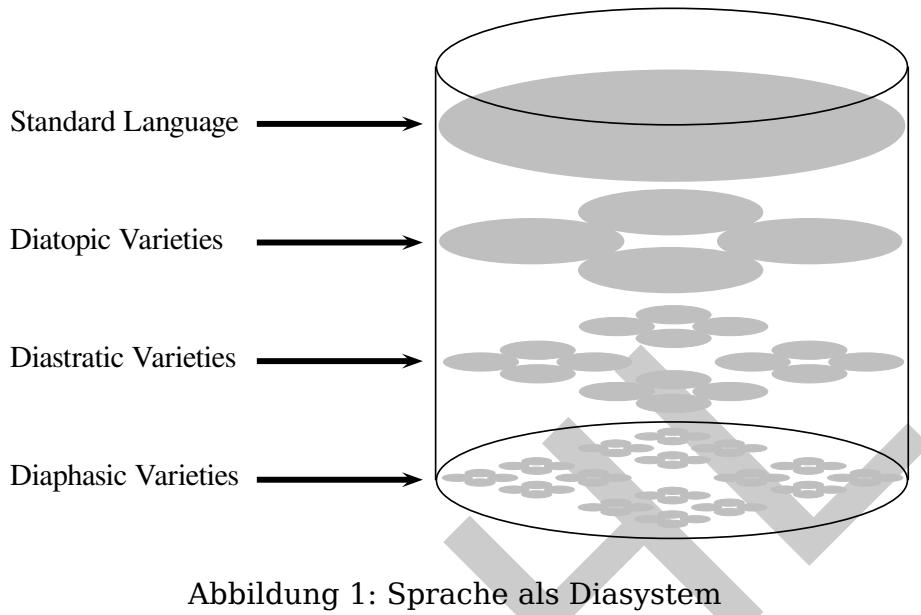
Tabelle 1: „Der Nordwind und die Sonne“ in verschiedenen Sprachvarietäten

Die Tabelle zeigt phonetische Transkriptionen der Übersetzung des Satzes „Der Nordwind und die Sonne stritten sich, wer von ihnen der stärkere sei“ in sechs verschiedenen sprachlichen Varietäten. Leider gibt es keine weiteren Angaben zur Struktur der Tabelle. Wie lässt sich diese erklären? Welche Schlussfolgerungen in Bezug auf die Einteilung der chinesischen Varietäten in Dialekte und der skandinavischen Varietäten in Sprachen lassen sich aus der Tabelle ziehen?

Das Konzept des Diasystems

Um den komplexen, heterogenen Charakter von Sprachen im Rahmen der Linguistik realistischer widerzuspiegeln wird in der Soziolinguistik gewöhnlich vom Modell des *Diasystems* (Bussmann 1996: 312) Gebrauch gemacht. Gemäß diesem Modell sind Sprachen komplexe Aggregate verschiedener linguistischer Systeme, „die miteinander koexistieren und sich gegenseitig beeinflussen“ (Coseriu 1973: 40). Ein wichtiger Aspekt ist dabei das Vorhandensein einer sogenannten *Dachsprache*, d. h. einer linguistischen Varietät, die als Standard für interdialektale Kommunikation dient (Goossens 1973: 11). Die unterschiedlichen linguistischen

Varietäten (Dialekte, Soziolekte), welche durch einen solchen Standard verbunden werden, bilden den *Varietätenraum* einer Sprache (Oesterreicher 2001), wie in Abbildung 1 dargestellt.



Wie lässt sich mit Hilfe des Modells des Diasystems die unterschiedliche Einteilung der chinesischen und der skandinavischen Varietäten erklären?

Vereinfachtes Sprachmodell der historischen Linguistik

In der historischen Linguistik wird ein sehr vereinfachtes Sprachmodell zugrunde gelegt. Es interessiert dabei weniger, was eine Sprache tatsächlich „ist“ als vielmehr, wie eine Sprache sich verändert. Dabei wird Sprache als System angesehen. In einem weiten Sinne besteht ein System aus einer Menge von Elementen und einer Menge von Relationen, die sich auf die Menge der Elemente beziehen (Marchal 1975: 462f). Für das Sprachmodell in der historischen Linguistik bedeutet dies, dass sprachliche Systeme *Laute* (Phone, Phoneme) und *Zeichen* (Wörter, Morpheme) als Elemente aufweisen und *phonotaktische* und *syntaktische Regeln* als Relationen.

Reicht ein dermaßen vereinfachtes Sprachmodell für eine Behandlung der wichtigen Probleme der historischen Linguistik aus?

1.2 Zeichen

Das klassische Zeichenmodell

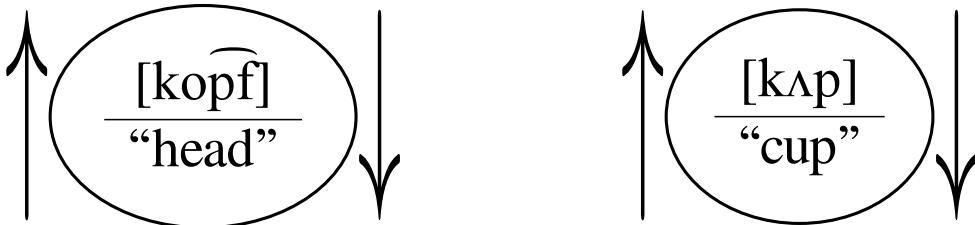
Sprachliche Zeichen werden in der historischen Linguistik gewöhnlich im Rahmen des traditionellen Zeichenmodells von Saussure (1916) beschrieben. Es wird also die Ausdrucksseite von der Inhaltsseite des sprachlichen Zeichens unterschieden, oder, wie Roman Jakobson es ausdrückt:

The sign has two sides: the sound, or the material side on the one hand, and meaning, or the intelligible side on the other. Every word, and more generally every verbal sign, is a combination of sound and meaning, or to put it another way, a combination of signifier and signified [...]. (Jakobson 1976[1978]: 3)

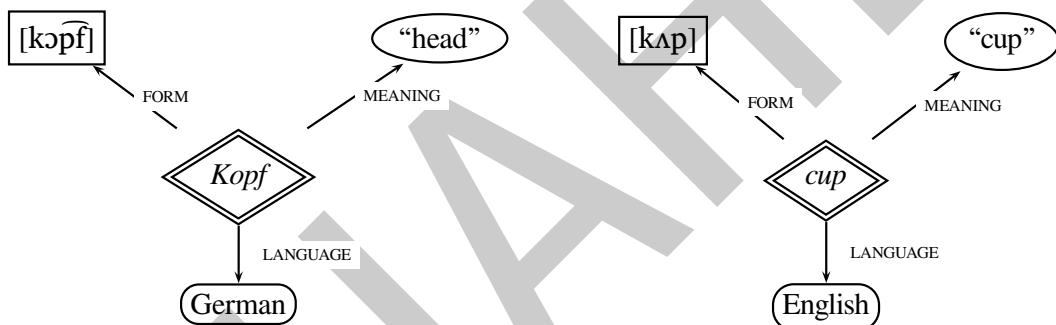
Was mein Jakobson mit „material“ und „intelligible“?

Nachtrag

Normalerweise wird das klassische Zeichenmodell Saussure's wie folgt verbildlicht:



Konstituierend für das sprachliche Zeichen ist neben der *Form* (Signifikant) und der *Bedeutung* (Signifikat) aber auch das sprachliche *System*, in dem das Zeichen seine Geltung hat. Eine genauere Verbildlichung des Zeichenmodells sollte daher auch dem System Rechnung tragen:



Betrachtet man die Struktur von Zeichenform und Zeichenbedeutung, so lassen sich grundlegende Unterschiede zwischen den beiden feststellen. Bei der Zeichenform handelt es sich um eine (phonetische) Sequenz, also eine Abfolge distinktiver Laute. Diese Laute sind materiell, insofern als sie bspw. als Schwingungen in der Luft gemessen werden können, oder als Spuren von Tinte auf einem Blatt Papier. Wichtig ist für die Zeichenform ferner ihre Linearität, denn nicht die bloße Anhäufung an Lauten, sondern ihre Abfolge ist entscheidend für die Unterscheidung von Zeichenformen. Man kann daher sagen, dass die Zeichenform (a) substantiell, (b) segmentierbar, und (c) linear ist. Wie aber verhält es sich mit der Zeichenbedeutung? Trage die Entsprechungen in der jeweiligen Spalte in der Tabelle ein.

No.	Form	Bedeutung
(a)	substantiell	
(b)	segmentierbar	
(c)	linear	

2 Wandel

Wandel als Katastrophe

Schon früh in der Geschichte der Linguistik war den Forschern in Europa bewusst, dass Sprachen sich wandeln können. Vorherrschend war dabei jedoch die Ansicht,

dass alle Formen des Wandels „katastrophisch“ abliefen, dass Wandel also im Rahmen eines unberechenbaren, chaotischen „Verfalls“ vor sich ginge. Erst spät (zu Beginn des 19. Jahrhunderts) wurde erstmals klar, dass sich bestimmte Phänomene des Sprachwandels, insbesondere der Lautwandel, durch eine beachtliche Regelmäßigkeit auszeichnen.

In welchem Zusammenhang ist uns die katastrophische Theorie des Sprachwandels bereits begegnet und wie haben wir sie damals genannt?

Wandel als Prozess

Dass Lautwandel gerade nicht zwangsläufig katastrophisch verläuft, kann schnell gezeigt werden, wenn man Wörter des Lateinischen mit Wörtern einer seiner Nachfolgersprachen (wie bspw. des Italienischen) vergleicht. Dabei stellen die Daten in Tabelle 2 jeweils nur Beispiele dar, die sich nahezu beliebig erweitern lassen.

Bedeutung	Latein	Italienisch
„Feder“	plu:ma	pjuma
„flach“	pla:nus	pjano
„Platz“	plate:a	pjats:a

Bedeutung	Latein	Italienisch
„Zunge“	lĩngua	lĩngwa
„Mond“	lu:na	luna
„langsam“	lentus	lento

Tabelle 2: Lateinische und Italienische Wörter

Wenn die Daten in Tabelle 2 tatsächlich nur einen Ausschnitt für eine viel größere Anzahl von Beispielen von Wortvergleichen zwischen dem Lateinischen und dem Italienischen darstellen, welche grundlegenden Eigenschaften des Lautwandels lassen sich dann daraus ableiten?

Wandel als Gesetz

Die Erkenntnis, dass Lautwandel zu großen Teilen regelmäßig verläuft, wurde enthusiastisch von den Sprachforschern zu Beginn des 19. Jahrhunderts aufgenommen, und führte von der Regularitätsannahme schnell zum Terminus des *Lautgesetzes*. Das Phänomen des Lautwandels wurde dabei bewusst mit einem den Naturgesetzen vergleichbaren gemeingültigen Prozess in Beziehung gesetzt. Die stärkste Hypothese in Bezug auf den Lautwandel wurde dabei von den sogenannten Junggrammatikern (einer Gruppe von Linguisten in Leipzig, die durchschnittlich jünger als andere Linguisten in Deutschland waren) formuliert, welche als „Junggrammatische Hypothese“ in die Geschichte einging:

Aller lautwandel, soweit er mechanisch vor sich geht, vollzieht sich nach ausnahmslosen gesetzen, d.h. die richtung der lautbewegung ist bei allen angehörigen einer sprachgenossenschaft, ausser dem Fall, dass dialektspaltung eintritt, stets dieselbe, und alle wörter, in denen der der lautbewegung unterworfenen laut unter gleichen verhältnissen erscheint, werden ohne ausnahme von der änderung ergriffen. (Osthoff and Brugmann 1878: XIII)

Was genau besagt die junggrammatische Hypothese in ihren Einzelheiten?

Ausblick

Wenn man sich mit Sprachwandel beschäftigt, lassen sich verschiedene Ebenen unterscheiden. Die bisherigen Beispiele orientieren sich primär am **Lautwandel**. Lautwandel an sich lässt sich jedoch wiederum im Hinblick auf verschiedene Aspekte untersuchen.

Die **prozedurale Perspektive** beschäftigt sich mit dem Prozess und dessen Geltungsbereich, wobei unterschiedliche **Mechanismen** des Lautwandels untersucht werden. Diese Perspektive wird in der Sitzung „Und sie ändert sich doch!“ (Sprachwandel, Lautwandel und lexikalischer Wandel) im Vordergrund stehen.

Die **substantielle Perspektive** beschäftigt sich mit dem Wandel hinsichtlich der lautlichen **Substanz**, wobei unterschiedliche **Typen** des Lautwandels unterschieden werden. Diese Perspektive wird in der Sitzung „Regelmäßige Korrespondenz“ (Sequenzalinierung, Identifizierung von Lautkorrespondenzen) genauer behandelt werden.

Lautwandel ist jedoch nicht die einzige Form von Wandel, die in der historischen Linguistik untersucht wird, denn auch die Bedeutungsseite des sprachlichen Zeichens kann sich im Laufe der Zeit wandeln. Phänomene des **semantischen Wandels** werden in der Sitzung „Das macht doch alles keinen Sinn mehr“ (Quantitativer Umgang mit Bedeutungswandel) genauer untersucht werden.

Eine weitere und allgemeinere Form des sprachlichen Wandels, die eine wichtige Rolle in der historischen Linguistik spielt, ist der **lexikalische Wandel**. Dieser wird vor allem in der Sitzung „Traue keiner Statistik“ (lexikostatistischer Sprachvergleich) wieder aufgegriffen.

3 Relationen

Wandel und Resultate des Wandels

Ausgehend von der Annahme, dass Lautwandel regelmäßig verläuft, lassen sich bestimmte Rückschlüsse in Bezug auf mögliche Resultate des Lautwandels postulieren. Tabelle 3 zeigt Beispiele für diese Resultate anhand der genetisch verwandten Sprachen Sanskrit, Latein, Griechisch und Gotisch.

Nummer	Sanskrit	Latein	Griechisch	Gotisch	Bedeutung
1	pad-	pe:d-	pod-	fo:t-	Fuß
2	pi'tar-	pater	pa'ter	fadrs	Vater
3	pardate:	ø	'perdomai	*firtan	furzen
4	ta'no:ti	tendere	'teinein	θanjan	dehnen
5	'trajas	tres	treis	θreis	drei
6	tu	tu	ty	θu	du
7	*ka'put-	kaput	ø	hawbiθ	Kopf
8	*kaln-	ø	ø	hals	Hals
9	kra'vis	kruor	'kreas	*hre:wa	Fleisch, Blut
10	ka-	k'woj	po-tʰen	hʷas	wie?, was?
11	tʃ'i'-ra-	k'wies	ø	hʷi:-la	Ruhe
12	ø	s-kʷalus	a's-palos	*hʷalas	Wal, Fisch

Tabelle 3: Lautkorrespondenzen verschiedener indogerman. Sprachen

Bezeichnend für die Anfänge der historischen Linguistik, aber auch nach wie vor für viele neuere Darstellungen, ist ein Vermischen der *Phänomene* mit den

Resultaten des Lautwandels. Während sich die Phänomene mit Hinblick auf einen Anfangs- und Endpunkt beschreiben lassen, müssen Anfangs- und Endpunkt aus den Resultaten des Lautwandels erst erschlossen werden. Wie der von Jacob Grimm (1785 – 1863) geprägte Terminus „Lautverschiebung“ jedoch schon andeutet, war diese strikte Trennung alles andere als selbstverständlich für die Pioniere der historischen Linguistik. So interpretiert Grimm seine Daten direkt historisch als Prozess, ohne deutlich zu machen, dass der Prozess nicht zwangsläufig aus den Daten ersichtlich wird:

Noch merkwürdiger als die einstimmung der liq. und spir. ist die abweichung der lippen- zungen- und kehllaute nicht allein von der gothischen, sondern auch der alth. einrichtung. Nämlich genau wie das alth. in allen drei graden von der goth. ordnung eine stufe abwärts gesunken ist, war bereits das goth. selbst eine stufe von der lateinischen (griech. indischen) herabgewichen. Das goth. verhält sich zum lat. gerade wie das alth. zum goth. (Grimm 1822: 584)

Das Gleiche gilt für Rasmus Rask (1787 – 1832), der die „Lautverschiebung“ bereits vor Grimm entdeckte:

Aber nicht nur in den Endungen, auch im Wort selbst gehen vielfältige Veränderungen vor sich. Es ist vielleicht nicht unangebracht, auf die häufigsten von diesen Übergängen vom Griechischen und Lateinischen zum Isländischen hinzuweisen. (Rask 1818: 169)¹

Ähnliche Beispiele für die Resultate des Lautwandels wie die in Tabelle 3 dargestellten lassen sich auch in den Werken von Grimm und Rask finden. Worin besteht die grundlegende Vermengung der Resultate mit den Phänomenen des Lautwandels in den Werken der beiden Forscher?

Ausblick

Die Unterscheidung zwischen „Wandel“ und den „Resultaten des Wandels“ spiegelt sich in unterschiedlichen, grundlegenden **Relationen** zwischen Zeichen und Sprachen wieder. Mit Relationen zwischen Zeichen werden wir uns in der Sitzung „Ist doch alles arbiträr“ (Sprachliche Zeichen und Zeichenrelationen) genauer befassen. Relationen zwischen Sprachen werden eine wichtige Rolle in den Sitzungen „Of Waves and Trees“ (Formen der Sprachverwandtschaft) und „Vaterschaftstest für Sprachen“ (Nachweis von Sprachverwandtschaft) spielen.

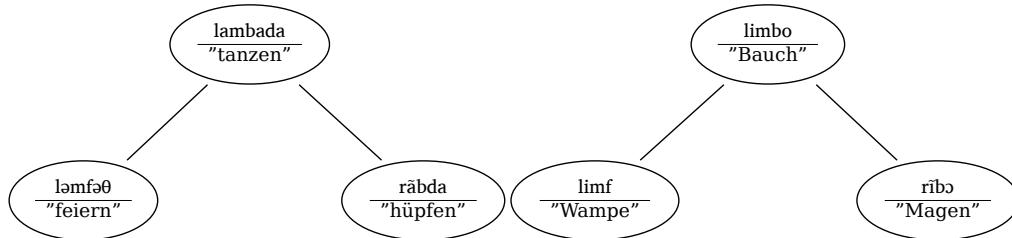
4 Ähnlichkeiten

Strukturelle Ähnlichkeiten und Lautkorrespondenzen

Wenn, wie die Arbeitshypothese der historischen Linguistik dies besagt, Sprachen sich hinsichtlich ihres Lautsystems regelmäßig ändern, so muss diese regelmäßige Änderung sich in Sprachen, die genetisch verwandt sind, in Form von Ähnlichkeiten in Lexemen widerspiegeln. Wenn bspw. eine hypothetische Sprache X die Worte [lambada] „tanzen“ und [limbo] „Bauch“ aufweist und sich dann im Verlaufe ihrer Geschichte in zwei Tochtersprachen Y und Z aufspaltet, und in diesen Tochtersprachen beide Wörter noch erhalten sind, dann kann man davon ausgehen, dass die

¹ Meine Übersetzung, Originaltext: „Men ikke blot i Endelserne, ogsaa i Ordene selv foregas mangfoldige Forandringer, det vil maaskje ikke være af Vejen her at mærke sig de hyppigste af disse Overgange fra Græsk og Latin til Islandsk.“

Wörter in den Tochtersprachen sich zwar unterscheiden, weil sich ihre Lautgestalt gewandelt hat, sie aber dennoch strukturelle Ähnlichkeiten aufweisen. In Sprache Y könnten die Wörter bspw. als [ləmfəθ] „feiern“ und [limf] „Wampe“ erhalten sein und in Sprache Z als [rābda] „hüpfen“ und [rībɔ] „Magen“, wie die folgende Abbildung zeigt:



In Sprache Y hätte entsprechend Lautwandel von [a] zu [ə], von [b, d] zu [f, θ], stattgefunden, und ein Schwund von Vokalen in Endsilben. In Sprache Z hätte ein Verlust von [m] nach Vokal mit gleichzeitiger Nasalierung des Vokals, ein Wandel von [l] zu [r], und ein Schwund aller Vokale in Mittelsilben stattgefunden. In einem solchen Fall würden die beiden Wörter in den beiden Tochtersprachen zwar sehr unterschiedlich aussehen, sie würden aber strukturelle Ähnlichkeiten aufweisen, die sich am besten aufzeigen lassen, indem man die Wörter aliniert, d.h. indem man die Segmente, die sich entsprechen, einander gegenüberstellt:

Sprache	Wort	Bedeutung	Alinierung
Y	ləmpət	"feiern"	l ə m f ə θ -
Z	rābda	"hüpfen"	r ã - b - d a
Y	limp	"Wampe"	l i m f -
Z	rībɔ	"Magen"	r ī - b ɔ

Wenn man sich diese Tabelle genauer anguckt, kann man sehr leicht erkennen, dass bestimmte Segmente miteinander „korrespondieren“: obwohl sie unterschiedlich sind, so wie [l] und [r], treten sie doch an den gleichen Stellen in den Wörtern auf, und zwar nicht nur in einem Fall. Die Wörter sind somit strukturell ähnlich. Die Lautsegmente, die in diesen Wörtern korrespondieren, werden **Lautkorrespondenzen** genannt. Wörter unterschiedlicher Sprachen, die auf eine gemeinsame Vorform in einer gemeinsamen Vorgängersprache zurückgehen, nennt man **Kognaten**.

Dt. <Zeh> und en. <toe> sind Kognaten, genauso wie dt. <Zahn> und en. <tooth> und dt. <Zaun> und en. <town>. Welche Lautkorrespondenz lässt sich aufgrund der Beispiele postulieren?

Ausblick

Ähnlichkeiten zwischen Zeichen und Sprachen werden eine große Rolle in den Sitzungen „Regelmäßige Korrespondenz“ (Sequenzalinierung, Identifizierung von Lautkorrespondenzen) und „Finding Cognates“ (automatische Kognatenerkennung) spielen.

5 Beweise

Schlussverfahren

Schlussverfahren sind wichtig, um „zu einem Schluss zu kommen“, also etwas zu beweisen. Sie sind daher auch wichtig für die historische Linguistik, da diese

schließlich wie alle Wissenschaften irgendetwas beweisen will. Die drei grundlegenden Schlussverfahren **Deduktion**, **Induktion** und **Abduktion** lassen sich am besten unterscheiden, indem man sie als eine Kombination von Regel, Ereignis, und Resultat auffasst:

Deduktion: „All bunnies have long ears, and the thing that brings the Easter eggs is a bunny. Therefore, the thing that brings the Easter eggs has long ears.”

Induktion: „The thing that brings the Easter eggs is a bunny, and the thing that brings the Easter eggs has long ears. Therefore, all bunnies have long ears.”

Abduktion: „All bunnies have long ears, and the thing that brings the Easter eggs has long ears. Therefore, the thing that brings the Easter eggs is a bunny.”

Das abduktive Schlussverfahren wird gemeinhin als das schwächste der drei Verfahren angesehen. Warum ist das so, und warum findet es dennoch vor allem in historischen Wissenschaften Verwendung?

Ausblick

Mit Beweisen, Indizien und Schlussverfahren werden wir uns explizit in der Sitzung „Vaterschaftstest für Sprachen“ Nachweis von Sprachverwandtschaftin auseinandersetzen.

Literatur

- Barbour, S. and P. Stevenson (1998). *Variation im Deutschen. Soziolinguistische Perspektiven*. Berlin: de Gruyter.
- Bussmann, H., ed. (1996). *Routledge dictionary of language and linguistics*. Trans. from the German by G. Trauth and K. Kazzazi. London and New York: Routledge.
- Coseriu, E. (1973). *Probleme der strukturellen Semantik. Vorlesung gehalten im Wintersemester 1965/66 an der Universität Tübingen*. Tübingen: Narr.
- Goossens, J. (1973). *Niederdeutsch. Sprache und Literatur. Eine Einführung*. Neumünster: Karl Wachholtz.
- Grimm, J. (1822). *Deutsche Grammatik*. 2nd ed. Vol. 1. Göttingen: Dieterichsche Buchhandlung. Google Books: MnsKAAAAIAAJ.
- Jakobson, R. (1976 [1978]). *Six lectures on sound and meaning*. Trans. from the French by J. Mepham. With an intro. by C. Lévi-Strauss. Cambridge and London: MIT Press.
- Marchal, J. H. (1975). “On the concept of a system”. English. In: *Philosophy of Science* 42.4, 448–468. JSTOR: 187223.
- Oesterreicher, W. (2001). “Historizität, Sprachvariation, Sprachverschiedenheit, Sprachwandel”. In: *Language typology and language universals. An international handbook*. Ed. by M. Haspelmath. Berlin and New York: Walter de Gruyter, 1554–1595.
- Osthoff, H. and K. Brugmann (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*. Vol. 1. Leipzig: Hirzel.
- Rask, R. K. (1818). *Undersögelse om det gamle Nordiske eller Islandske sprogs oprindelse* [Investigation of the origin of the Old Norse or Icelandic language]. Copenhagen: Gyldendalske Boghandlings Forlag. GoogleBooks: cWgJAAAQAAJ; English translation: – (1993). *Investigation of the origin of the Old Norse or Icelandic language*. Trans. by N. Ege. Travaux du Cercle Linguistique de Copenhague 26. Copenhagen: The Linguistic Circle of Copenhagen.
- Saussure, F. de (1916). *Cours de linguistique générale*. Ed. by C. Bally. Lausanne: Payot; German translation: – (1967). *Grundfragen der allgemeinen Sprachwissenschaft*. Trans. from the French by H. Lommel. 2nd ed. Berlin: Walter de Gruyter & Co.

„Und sie ändert sich doch!“

1 Sprachwandel

Reim dich oder ich fress dich...

Dass Sprachen sich im Laufe ihrer Geschichte verändern können, war keine notwendige Entdeckung, welche die Gelehrten im Laufe der Geschichte machen mussten. Dies gilt insbesondere für die sich weniger offensichtlich verändernden Bereiche der Sprache, wie bspw. die Laute. Die folgende Tabelle gibt ein Beispiel für die Folgen des Lautwandels: Das chinesische Gedicht, entnommen aus dem *Buch der Oden* (ca. 1050–600 v. Chr., *Shījīng*: 28.3) und wiedergegeben in moderner Pīnyīn-Transliteration mitsamt einer Übersetzung von Bernhard Karlgren 1950, reimt nicht durchgängig in allen Reimwörtern, die grau unterlegt sind.

燕	燕	于	飛,	下	上	其	音。	The swallows go flying, falling and rising are their voices;
yān	yān	yú	fēi	xià	shàng	qí	yīn	
之	子	于	歸,	遠	送	于	南。	This young lady goes to her new home, far I accompany her to the
zhī	zǐ	yú	guī	yuǎn	sòng	yú	nán	south
瞻	望	弗	及,	實	勞	我	心。	I gaze after her, can no longer see her, truly it grieves my heart.
zhān	wàng	fú	jí,	shí	láo	wǒ	xīn	

Angesichts der Tatsache, dass zwischen der ursprünglichen Fassung des Gedichts und der jetzigen Form über 2000 Jahre vergangen sind, ist dies nicht verwunderlich. Für chinesische Gelehrte war es jedoch lange Zeit nicht die offensichtlichste Erklärung, warum die alten Gedichte zuweilen komische Reime aufwiesen. So wurden die komischen Reime zunächst alternativ erklärt als Folge von

- a) laxen Reimkonventionen der Vorfahren (Baxter 1992: 153-157), oder
- b) von sogenannten „Lautharmonisierungen“ (*xiéyīn* 叶音), welche die Vorfahren aus ästhetischen Gründen vorgenommen hätten (ibid.).

Ohne irgendeine Vorstellung von Sprachwandel zu haben, begannen die chinesischen Gelehrten die verschiedenen „komischen“ Reime zu systematisieren. Das Ergebnis waren erste Rekonstruktionen eines – wenn auch nicht als solchen erkannten – abstrakten Systems der Reime des Altchinesischen. Erst der chinesische Gelehrte Chén Dì (1541 – 1606) stellte die Hypothese auf, dass die komischen Reime tatsächlich eine Folge von Sprachwandel seien:

The writings of scholars must be made of adequate sounds. Even in the rural areas everybody orders the sounds harmonically. Can it be that the ancients solely did not have rhymes? One can say that in the same way in which ancient times differ from modern times, and places in the North differ from places in the South, characters change and sounds shift. This is a natural tendency. Therefore, it is inevitable that reading the ancient writings with modern pronunciation will sound improper and wrong.
(Máoshī Gǔyīnkǎo: 原序)¹

In China entdeckte man den Sprachwandel recht spät, wie sah es im Westen aus?

¹Meine Übersetzung, Original: 故士人篇章，必有音節，田野俚典，亦名諧聲，豈以古人之詩而獨無韻乎？蓋時有古今，地有南北，字有更革，音有轉移，亦孰所必至。故以今之音讀古之作，不免乖刺而不入。

Wandel und Regelmäßigkeit

Während den Menschen im Verlaufe der Geschichte bereits relativ lange bewusst war, dass Sprachen sich ändern können, war es eine radikal neue Erkenntnis, die sich zu Beginn des 19. Jahrhunderts herauskristallisierte, dass Sprachen sich in Prozessen ändern, von denen bestimmte sogar regelmäßig verlaufen können. Mit der Entdeckung der Regelmäßigkeit einher festigte sich ebenfalls die Erkenntnis, dass Sprachen miteinander verwandt sein können, wobei Verwandtschaft von Sprachen dadurch definiert ist, dass miteinander verwandte Sprachen aus einer gemeinsamen Vorgängersprache entstanden sind, wie bspw. das Englische und das Deutsche, die beide aus dem Protopfälzischen hervorgegangen sind.

Warum ist die "Entdeckung der Regelmäßigkeit" so wichtig für das Postulieren von Verwandtschaftsbeziehungen zwischen Sprachen (oder ist sie das überhaupt?)?

2 Lautwandel

Wandel als Gesetz

Rufen wir uns zunächst noch mal in Erinnerung, was die These der Junggrammatiker in Bezug auf den Lautwandel war:

Aller lautwandel, soweit er mechanisch vor sich geht, vollzieht sich nach ausnahmslosen gesetzen, d.h. die richtung der lautbewegung ist bei allen angehörigen einer sprachgenossenschaft, ausser dem Fall, dass dialektspaltung eintritt, stets dieselbe, und alle wörter, in denen der der lautbewegung unterworfenen laut unter gleichen verhältnissen erscheint, werden ohne ausnahme von der änderung ergriffen. (Osthoff and Brugmann 1878: XIII)

Wang (1969) trifft, wie wir inzwischen wissen, die Unterscheidung zwischen verschiedenen Perspektiven auf den Lautwandel. Reicht dieses Zitat aus, die junggrammatische Position eindeutig zu bestimmen? Begründe Deine Meinung mit Rückgriff auf den Text.

Von Wörtern und Geschichten

Nicht alle Linguisten waren der Meinung der Junggrammatiker. Besonders Dialektologen folgten dem berühmten Slogan „*chaque mot a son histoire*“ („jedes Wort hat seine Geschichte“), der gewöhnlich Jules Gilliéron (1854 – 1926) zugeschrieben wird (Campbell 1999: 189). Die Bedenken der Dialektologen standen jedoch streng genommen nicht direkt im Widerspruch zur junggrammatischen Doktrin, schließlich besagte die junggrammatische Theorie ja nicht, dass sich zwangsläufig *alle* Wörter einer Sprache regelmäßig änderten, sondern lediglich, dass idiosynkratischer Wandel „could be accounted for [...] by certain less obvious mechanisms of borrowing and analogy“ (Kiparsky 1988: 368). Dennoch begann die linguistische Gemeinschaft sich zu spalten, während die Diskussion selbst in eine Sackgasse geraten war, da keines der beiden Lager vollkommen überzeugende Argument für sich in Anspruch nehmen konnte.

Was meint Herr Kiparsky in diesem Zusammenhang mit "Analogie"?

Von der Diffusion

Die Situation änderte sich erst in den 60er Jahren des 20. Jahrhunderts, als neue Forschungen – die vorwiegend im Bereich der chinesischen Dialekte stattfanden – zur Beschreibung eines neuen Wandelmechanismus führten, der gewissermaßen das Gegenteil der junggrammatischen Hypothese darstellte. Diese nämlich hatten angenommen, dass Lautwandel *lexikalisch abrupt* und *phonetisch graduell* verläuft:

Regarding the lexicon [they assumed] that a change always affects the whole lexicon, and can therefore be seen as an abrupt change. Regarding the sounds [they assumed] that the change proceeded step by step, and can therefore be seen as a gradual change. (Wang 2006: 109)²

Die Ergebnisse der chinesischen Dialektologen jedoch legten den Schluss nahe, dass ein bestimmter Mechanismus des Lautwandels, der später *lexikalische Diffusion* genannt wurde, genau umgekehrt verläuft, nämlich *lexikalisch graduell* und *phonetisch abrupt*:

Phonological change may be implemented in a manner that is phonetically abrupt but lexically gradual. As the change diffuses across the lexicon, it may not reach all the morphemes to which it is applicable. If there is another change competing for part of the lexicon, residue may result. (Wang 1969: 9)

Die folgende Tabelle gibt ein Beispiel für die Phänomene der lexikalischen Diffusion. Die Tabelle listen Zeichenpaare mit identischer mittelchinesischer Lesung (ca. 600 n. Chr.), die mit ihren modernen Reflexen im Shuāngfēng Dialekt, der zur Gruppe der Min-Dialekte gehört, kontrastiert worden sind.

Character	Pīnyīn	Meaning	Middle Chinese	Shuāngfēng
步	bù	„to walk”	bo ³	bu ³³
捕	bǔ	„to grasp”	bo ³	p ^h u ²¹
刨	páo	„to dig”	bæw ¹	bə ³³
跑	páo	„to scrape”	bæw ¹	p ^h ə ²¹
盜	dào	„to rob”	daw ³	də ³³
導	dǎo	„to lead”	daw ³	t ^h ə ³⁵

Warum sind die mittelchinesischen Zeichenlesungen mit ihren Reflexen im Shuāngfēng-Dialekt ein Beispiel für lexikalische Diffusion? Wie könnten die Beispiele alternativ erklärt werden?

In Dubio pro Diffusione

Als Einzelfall sind Beispiele wie die in der Tabelle oben nur Beispiele für wie auch immer motivierten „irregulären“ Lautwandel, die leicht mit Hilfe externer Faktoren erklärt werden können. Dies scheint im Shangfeng-Dialekt jedoch nicht der Fall zu sein, wie eine umfangreiche Studie Chen (1972) der 616 Zeichen, deren Lesung im Mittelchinesischen einen stimmhaften Initial aufwies, zeigt: Zwar zeigen sich viele Beispiele für die Prozesse der Anlautverhärtung (*devoicing*) und Aspiration der stimmhaften Initiale, jedoch gibt es ebenfalls viele Fälle, in denen die stimmhaften

²Meine Übersetzung, Original: „作為詞彙，要變就都變，因而是一種突變。作為語音，變化是逐漸的，因而是—種漸變”。

Initiale bewahrt wurden. Dies spricht eher für eine Tendenz hin zur Anlautverhärtung und Aspiration im Shuāngfēng-Dialekt, denn für ein ausnahmsloses Gesetz. Eine derartige Tendenz kann recht gut erklärt werden, wenn man annimmt, dass Lautwandel nicht notwendigerweise das ganze Lexikon simultan erfasst, sondern in unterschiedlichen Geschwindigkeiten von Wort zu Wort „springt“:

When a phonological innovation enters a language it begins as a minor rule, affecting a small number of words [...]. As the phonological innovation gradually spreads across the lexicon, however, there comes a point when the minor rule gathers momentum and begins to serve as a basis for extrapolation. At this critical cross-over point, the minor rule becomes a major rule, and we would expect diffusion to be much more rapid. The change may, however, reach a second point of inflection and eventually taper off before it completes its course, leaving behind a handful of words unaltered.

Ist die lexikalische Diffusion eine notwendige Schlussfolgerung aus den Daten?

Tertium datur

Die lexikalische Diffusion ist dem junggrammatischen Lautwandelkonzept nicht nur in chronologischer Hinsicht entgegengesetzt, sondern greift auch dessen wichtigste Implikation für die linguistische Rekonstruktion an: Lautwandel verläuft der Theorie zufolge nicht ausnahmslos. Während einige Forscher daraufhin die junggrammatische „Hypothese“ vollständig verworfen, wies Labov (1981) jedoch nach, dass bestimmte Formen von Lautwandel phonologisch graduell und lexikalisch einheitlich verlaufen, dass also lexikalische Diffusion und „junggrammatisches Lautgesetz“ zwei verschiedene *Mechanismen* von Lautwandel darstellen:

There is no basis for contending that lexical diffusion is somehow more fundamental than regular, phonetically motivated sound change. On the contrary, if we were to decide the issue by counting cases, there appear to be far more substantially documented cases of Neogrammarian sound change than of lexical diffusion. (Labov 1994: 471)

Warum wohl wurde in diesem Zusammenhang das Wort "Mechanismus" gewählt, um den Geltungsbereich von lexikalischer Diffusion und junggrammatischem Lautwandel abzustecken? Welche "Aspekte des Lautwandels" werden hierbei hervorgehoben?

3 Bedeutungswandel

It's the Meaning, Stupid!

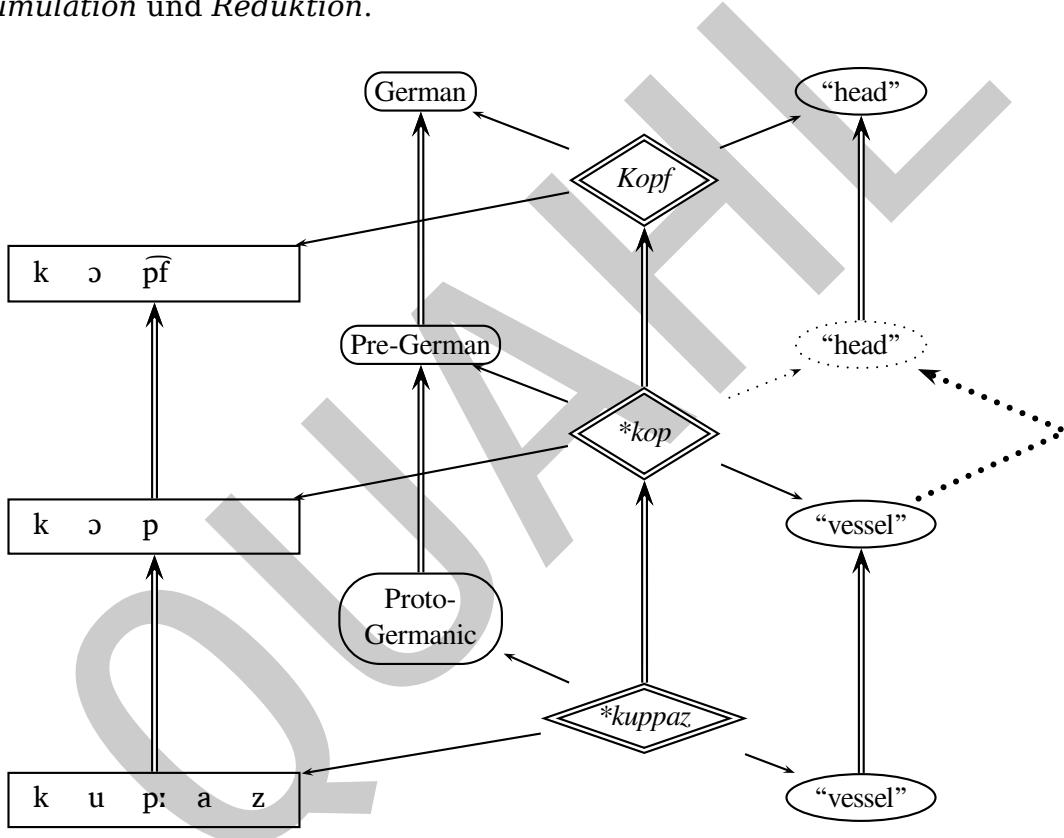
Während die Erforschung von Lautwandel sich großer Beliebtheit erfreut, und die Linguistik hier auch große Erfolge vorweisen kann, sieht es um die Erforschung des Bedeutungswandels weit weniger gut aus. Dies hängt sicher vor allem damit zusammen, dass man zwar weiß (und zuweilen sogar beobachten kann), dass Bedeutungen von Wörtern sich ändern, dass die Regelmäßigkeit, mit der sich die Lautgestalt von Wörtern ändert, jedoch keine Entsprechung in der Semantik findet:

There is [...] little in semantic change which bears any relationship to regularity in phonological change' (Fox 1995: 111).

Welche Beispiele für semantischen Wandel kennst Du? Welche Regelmäßigkeiten (wie marginal sie auch sein mögen) lassen sich zuweilen antreffen?

Alternation, Akkumulation, Reduktion

Ein Grund für die Unterschiede zwischen Lautwandel und Bedeutungswandel in Bezug auf die Regelmäßigkeit lässt sich in der grundlegend unterschiedlichen Struktur der Form- und der Bedeutungsseite des sprachlichen Zeichens finden. Während die Formseite des sprachlichen Zeichens eine sequenzielle Struktur aufweist, und Lautwandel durch die *Alternation* von Segmenten gekennzeichnet ist, ist die Bedeutungsseite wohl eher als konzeptuelles „Netzwerk“ zu beschreiben, und Bedeutungswandel ist nicht durch Alternation gekennzeichnet, sondern durch *Akkumulation* und *Reduktion*.



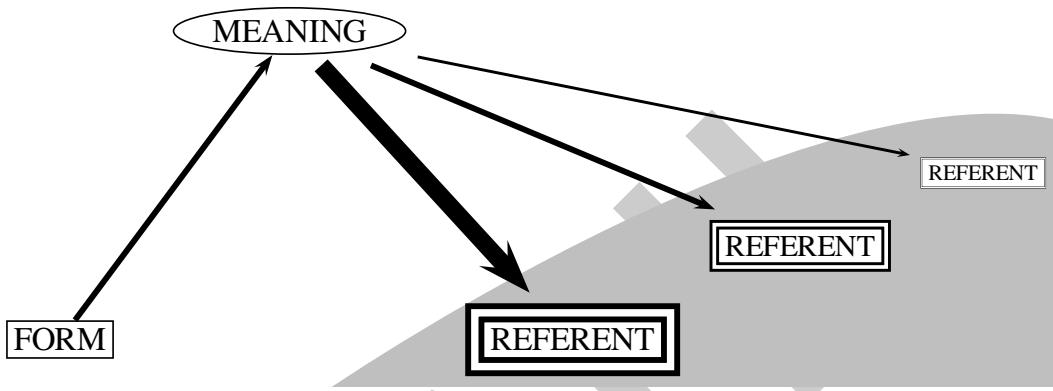
In der oben abgebildeten Grafik ist versucht worden, die Unterschiede zwischen Lautwandel und Bedeutungswandel anhand eines Beispiels zu kontrastieren. Wie lässt sich das Beispiel näher erläutern? Welche weiteren Beispiele gibt es für die Prozesse der Reduktion und der Akkumulation?

Referenzpotential

Im Gegensatz zur Form des sprachlichen Zeichens mangelt es der Zeichenbedeutung und Substanz und Linearität. Es mangelt ihr an Linearität, weil sie nicht sinnlich erfassbar ist, und es mangelt ihr an Linearität, da sie nicht von der Zeit abhängt. Ein weiteres Problem ist, dass – aufgrund der Arbitrarität der Verbindung zwischen Zeichenform und Zeichenbedeutung – „meaning is inherently fuzzy and

non-systematic" (Hock and Joseph 1995[2009]: 206). Bis heute gibt es daher leider noch keine semantische Theorie, die von einem breiten Teil der linguistischen Gemeinschaft unterstützt würde.

Saussures Zeichenmodell ist indifferent hinsichtlich der Frage, ob ein Zeichen verwendet wird, um auf die „reale Welt“ Bezug zu nehmen. Es wird lediglich betont, dass der Bedeutungsteil nicht mit dem Objekt, dass er denotiert, verwechselt werden sollte (Saussure 1916: 98). *Triadische Zeichenmodelle* versuchen, diese Lücke zu schließen, indem die *Bedeutung* eines Wortes (*meaning*) von dessen *Referenz* (*reference*) unterschieden wird, wobei erstere eine Kategorie und letztere ein mögliches Referenzobjekt determiniert (Löbner 2003: 257).



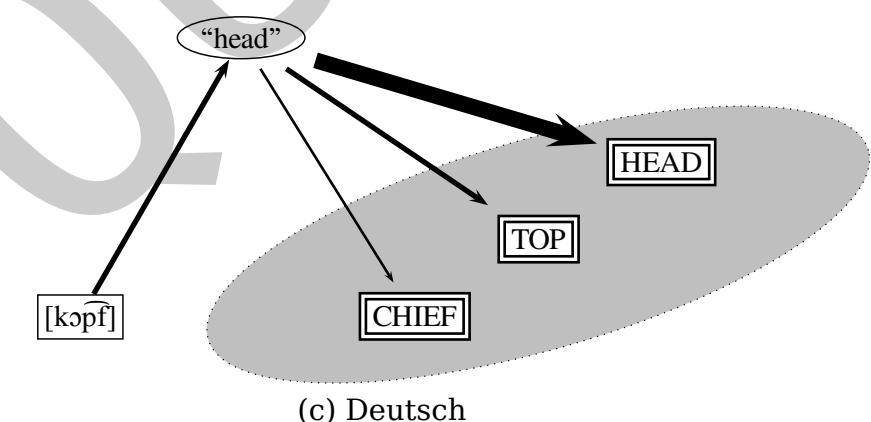
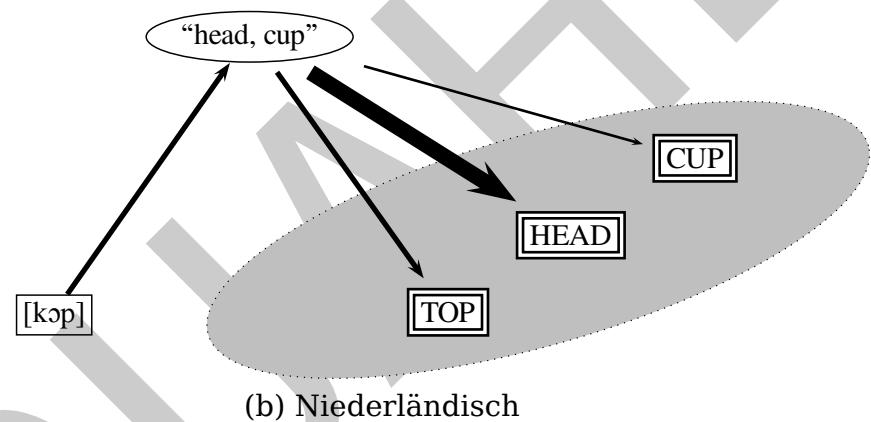
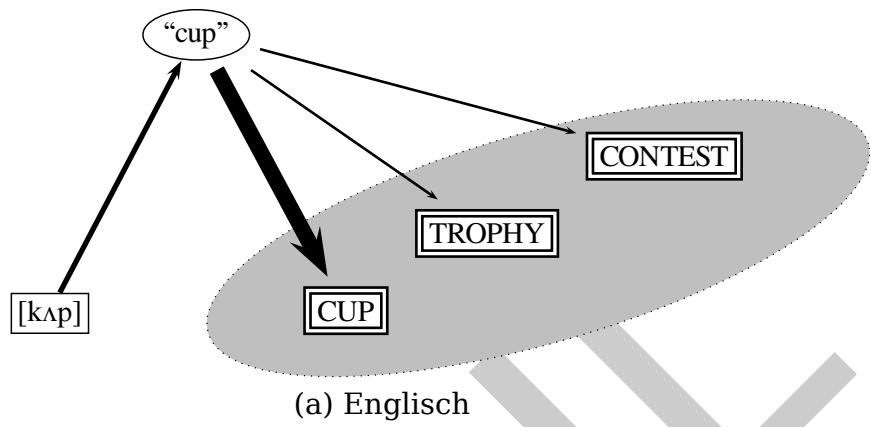
Da die Referenz eines sprachlichen Zeichens jedoch nur dann eindeutig ist, wenn das Zeichen in einem bestimmten Kontext verwendet wird, ist es sinnvoll eine weitere Unterscheidung zwischen *Referenz* und *Referenzpotential* zu machen (Schwarz 1996: 175). Das Referenzpotential eines sprachlichen Zeichens wird dabei verstanden als die Menge aller Möglichen Referenten, die von dem Zeichen denotiert werden können (siehe Grafik oben). Dabei hängt das Referenzpotential eines Zeichens von seiner Bedeutung ab: Je spezifischer diese ist, desto eingeschränkter ist die Anzahl (unterschiedlicher) möglicher Referenten (Löbner 2003: 306). Wenn man bspw. die Wörter Deutsch *Stein* [stain] und Deutsch *Ding* [dɪŋ] vergleicht, dann kann man sagen, dass das Referenzpotential von *Stein* eingeschränkter ist, als das von *Ding*, da das erste ja nur Steine oder steinartige Objekte denotieren kann, während das andere zur Denotation aller möglichen Arten (meist greifbarer) Objekte verwendet werden kann.

Warum ist es so schwer, das Referenzpotential eines sprachlichen Zeichens zu messen?

Bedeutungswandel als Wandel von Referenzpotentialen

Mit Hilfe der Idee vom Referenzpotential kann man das, was beim semantischen Wandel vor sich geht, zuweilen ein wenig erhellen. Wenn wir nur die am häufigsten auftretenden Referenten der Wörter Englisch *cup* „Tasse“, Niederländisch *kop* „Kopf, Tasse“ und Deutsch *Kopf* betrachten, so lässt sich ein bestimmtes Kontinuum bezüglich der häufigsten Referenten der Wörter feststellen. Dieses Kontinuum spiegelt nicht nur frühere Prozesse semantischen Wandels wider, sondern vermag eventuell auch auf zukünftige Prozesse oder generelle Trends zu verweisen. Der Wandel von KOPF zu CHEF, bspw., der wahrscheinlich über eine Zwischenstufe OBERER TEIL vonstatten ging, ist auch im Chinesischen attestiert, wo *shǒu* 首 [sou₂₁₄] „erster, Chef“ ursprünglich „Kopf“ hieß, wie man an der Orakelknochenversion des Zeichens noch gut erkennen kann: 首.

Das Konzept des Referenzpotentials mag vielleicht helfen, semantischen Wandel genauer zu beschreiben. Es hilf jedoch zwangsläufig nicht, ihn regelmäßig zu machen. Angenommen, semantischer Wandel wiese tatsächlich regelmäßige Züge auf, wie würden diese sich äußern, d.h. welcher Art wären diese Regelmäßigkeiten wohl?



Literatur

- Baxter, W. H. (1992). *A handbook of Old Chinese phonology*. Berlin: Mouton de Gruyter.
- Campbell, L. (1999). *Historical linguistics: An introduction*. 2nd ed. Edinburgh: Edinburgh Univ. Press.
- Chen, M. (1972). "The Time Dimension. Contribution toward a Theory of Sound Change". In: *Foundations of Language* 8.4, 457–498. JSTOR: 25000618.
- Máoshī Gǔyīnkǎo* 毛詩古音攷 [Investigation of the old sounds in Mao's compilation of the Book of Odes] (1606). By Chén Dì 陳第 (1541–1617); Critical edition: Chén Dì 陳第 (1606). *Máoshī Gǔyīnkǎo* 毛詩古音攷 [Investigation of the old sounds in Mao's compilation of the Book of Odes]. Sìkù Quánshū 四庫全書 [Complete Library of the Four Treasuries]; Jīngbù 經部; Xiǎoxuélèi 小學類; Míng 明. Internet Archive: 06048676.cn.
- Fox, A. (1995). *Linguistic reconstruction. An introduction to theory and method*. Oxford: Oxford University Press.
- Hock, H. H. and B. D. Joseph (1995 [2009]). *Language history, language change and language relationship. An introduction to historical and comparative linguistics*. 2nd ed. Berlin and New York: Mouton de Gruyter.
- Karlgren, B. (1950). *The book of odes. Chinese text, transcription and translation*. Stockholm: Museum of Far Eastern Antiquities.
- Kiparsky, P. (1988). "Phonological change". In: *Linguistics. The Cambridge survey*. Vol. 1: *Linguistic theory. Foundations*. Ed. by F. J. Newmeyer. Cambridge et al.: Cambridge University Press, 363–415.
- Labov, W. (1981). "Resolving the Neogrammarian Controversy". In: *Language* 57.2, 267–308. JSTOR: 413692.
- (1994). *Principles of Linguistic Change*. Vol. 2: *Internal Factors*. Blackwell.
- Löbner, S. (2003). *Semantik. Eine Einführung*. Berlin: de Gruyter.
- Osthoff, H. and K. Brugmann (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*. Vol. 1. Leipzig: Hirzel.
- Saussure, F. de (1916). *Cours de linguistique générale*. Ed. by C. Bally. Lausanne: Payot; German translation: - (1967). *Grundfragen der allgemeinen Sprachwissenschaft*. Trans. from the French by H. Lommel. 2nd ed. Berlin: Walter de Gruyter & Co.
- Schwarz, M. (1996). *Einführung in die kognitive Linguistik*. Basel and Tübingen: Francke.
- Shījīng* 詩經 [The book of odes] (ca. 1050 BC); Critical edition: *Máoshī* (1922). *Máoshī* 毛詩 [Mao's compilation of the Book of Odes]. Comm. by Máo Hēng 毛亨 (Hàn Dynasty: 207 BC–9 AD). *Sìbù Cóngkān* 四部叢刊 [The collected publications from the Four Categories]. Shanghai: Shāngwù 商务; English translation: Karlgren, B. (1950). *The book of odes. Chinese text, transcription and translation*. Stockholm: Museum of Far Eastern Antiquities.
- Wang, W. S.-Y. (1969). "Competing changes as a cause of residue". In: *Language* 45.1, 9–25. JSTOR: 411748.
- (2006). *Yǔyán, yǔyīn yǔ jìshù* 語言, 語音與技術 [Language, phonology and technology]. Shanghai: Xiānggǎng Chéngshì Dàxué.

„Ist doch alles arbiträr...“

1 Termini

1.1 Biologische Schlüsseltermini

Im Folgenden wollen wir versuchen, uns mit einigen Schlüsseltermini der Biologie vertraut zu machen, welche mitunter helfen können, linguistische Konzepte von Sprachverwandtschaft und Sprachgeschichte besser und klarer zu fassen.

Merkmal

Eines der wichtigsten Konzepte in der Evolutionsbiologie ist das Konzept des *Merkmales* (engl. *character*). Dieses kann allgemein in etwa wie folgt definiert werden:

A biological character can be thought of as a part of an organism that exhibits causal coherence to have a well-defined identity and that plays a (causal) role in some biological processes. (Wagner 2000: 3)

Nenne Beispiele für mögliche biologische Merkmale.

Homologie

Homologie (engl. *homology*) ist der allgemeine Terminus für die zentralste Relation in der Evolutionsbiologie: die Relation der *gemeinsamen Abstammung* (engl. *common descent*) zwischen zwei Merkmalen oder Organismen.

Homology is the relationship of two characters that have descended, usually with divergence, from a common ancestral character. This is important because most of the terminological problems stem from different definitions of homology. Characters can be any genic, structural or behavioral feature of an organism. (Fitch 2000: 227)

Der Terminus "Homologie" beschreibt eine Relation zwischen zwei gleichartigen Objekten. Wie bei den meisten wissenschaftlichen Termini lassen sich auch von diesem Terminus unterschiedliche Ableitungen bilden. So sind die Ableitungen "Homologe" und "homolog" durchaus gebräuchlich. Worauf sollte man bei der Verwendung dieser abgeleiteten Termini allerdings strikt achten, wenn man von deren eigentlicher Bedeutung ausgeht?

Orthologie, Paralogie, Xenologie

Homologie beschreibt allgemein Fälle von gemeinsamer Abstammung zwischen Merkmalen. In der molekularen Biologie, deren grundlegende Vergleichseinheit genetische *Sequenzen* sind, können Untertypen von gemeinsamer Abstammung Unterschieden werden, deren gebräuchlichste die *Orthologie* (engl. *orthology*), die *Paralogie* (engl. *paralogy*) und die *Xenologie* (engl. *xenology*) sind. Diese Untertypen der Homologie wurden in Bezug auf den Genotypen definiert, sie haben keine eindeutige Entsprechung im Phänotypen. Die Bedeutung der Termini kann man wie folgt zusammenfassen:

- **Orthologie** ist die Relation, die für „genes related via speciation“ (Koonin 2005: 311) gilt;
- **Paralogie** bezeichnet die Relation zwischen „genes related via duplication“ (ibid.); und
- **Xenologie** bezieht sich auf die Relation zwischen Genen, „whose history, since their common ancestor, involves an interspecies (horizontal) transfer of the genetic material for at least one of those characters“ (Fitch 2000: 229).

Weiter unten werden wir uns mit der "Entlehnung" als Schlüsselkonzept der historischen Sprachwissenschaft beschäftigen. Mit welchem der drei in diesem Abschnitt eingeführten biologischen Termini wäre diese wohl am besten zu vergleichen, und warum besteht dennoch ein unüberwindlicher Unterschied zwischen dem biologischen und dem linguistischen Konzept?

Analogie

Wir Linguisten werden wahrscheinlich denken, dass wir diesen Terminus kennen, insbesondere, wenn wir in der letzten Sitzung aufgepasst haben. Leider ist dies jedoch nicht ganz richtig, denn Analogie bezeichnet in der Evolutionsbiologie etwas anderes als in der historischen Linguistik und auch im Allgemeinen:

The relationship of any two characters that have descended convergently from unrelated ancestors. (ibid.)

Es mag sein, dass das Zitat von Fitch den Leser keinen Deut klüger macht. Wenn dem so sein sollte, liegt dies vermutlich an dem Wort "convergently", welches auf Deutsch "konvergent" oder einfacher "aufeinander zulaufend" (also "in die selbe Richtung") lautet. Was bezeichnet Analogie in der Evolutionsbiologie also genau?

Isologie

Als wären es nicht genug Logien, denen wir uns bisher widmen mussten! Die Biologen setzen aber trotzdem einen drauf und führen gleich noch das Konzept der *Isologie* (engl. *isology*) ein:

When dealing with molecular data concepts of homology have often been rather confused [...] with the word homology being used to mean several unrelated things, which could perhaps better be given alternative names. In particular, ‘sequence homology’ is often used as a synonym for ‘sequence identity’ (i.e. the number of nucleotides or amino acids that are inferred to be held in common between two sequences). These are not necessarily the same thing [...], since similarity can be the result either of common ancestry or of chance convergence, parallelism or reversal; and ‘isology’ may be a better term to use (Wegnez 1987). (Morrison 2006: 488)

To use the same word, “homology”, for structures with the same ancestry, either in comparative morphology or in molecular biology, is not itself a bad choice. Problems have arisen because the term has been used to express what it was not supposed to express. It is very difficult to quantify to what extent homologous morphological structures are identical. On the other hand, it is very easy to

quantify the degree of identity of two proteins or genes - indeed, this is usually the first interpretative step following sequence analysis. Most of the scientists concerned with sequence comparison have forgotten the morphological, embryological, and evolutionary origins of "homology", and the word has lost in the process its all-or-none concept status. Today the problem is first to restore the true meaning of the word "homology", and then to find an appropriate term for quantitative sequence comparisons. (Wegnez 1987: 516)

Kurz und schmerzlos: Warum brauchen wir (oder besser: die Biologen) einen solchen Terminus, und in welcher Beziehung steht er zu "Homologie" und "Analogie"?

1.2 Linguistische Schlüsseltermini

Sprachverwandtschaft

Sprachverwandtschaft (engl. *genetic [language] relationship*) ist das zentrale Konzept der historischen Sprachwissenschaft und bestimmt die Forschung in all ihren Grundzügen.

The relationship between languages that have a common ancestor; languages that are languages of the same **language family**. (Campbell and Mixco 2007: 68)

Sprachverwandtschaft wurde im Laufe des Seminars bereits mehrfach angesprochen. Welches ist das traditionelle Modell, mit dessen Hilfe Sprachgeschichte -- und damit auch die Entwicklung von Sprachverwandtschaften -- dargestellt wird, und warum ist dieses Modell so problematisch?

Kognazität

Nicht weniger wichtig, aber leider weitaus komplizierter und verwirrender, was den Gebrauch des Terminus betrifft, ist das Konzept der *Kognazität* (engl. *cognacy*). In der historischen Sprachwissenschaft bezeichnet Kognazität primär eine Relation der gemeinsamen Abstammung zwischen zwei Merkmalen. Dieser Terminus selbst ist jedoch weit weniger gebräuchlich als der Terminus *Kognate* (engl. *cognate*), der im Glossar von Campbell und Mixco folgendermaßen erläutert wird:

A word (or morpheme) that is related to a word (morpheme) in sister languages by reason of these words (morphemes) having been inherited by the related languages from a common word (morpheme) of the **proto-language** from which they descend. For example, Italian *cane* /kane/, Portuguese *cão* /kāõ/, French *chien* /šyē/ 'dog' are all cognates, since they descend in these Romance languages from the same original word in Latin (ancestor of the Romance languages): *canis* 'dog'. (ibid.: 33f)

Seien wir ehrlich und schonungslos: Die Beschreibung von Campbell und Mixco ist schrecklich ungenau, fehlerhaft und ungeschickt. Es stellt sich nur die Frage: Warum ist das so? Oder auch: Muss das so sein?

Entlehnung

Wo von Relationen zwischen Sprachen die Rede ist, darf die *Entlehnung* (engl. *borrowing*) natürlich nicht fehlen.

The process in which a language takes linguistic elements from another language and makes them part of its own. The borrowed elements are typically **loanwords**, but borrowing is not restricted just to lexical items taken from one language into another: any linguistic material – sounds, phonological rules, grammatical morphemes, syntactic patterns, semantic associations, discourse strategies – can be borrowed, that is, can be taken over so as to become part of the borrowing language. (Campbell and Mixco 2007: 25f)

Die Definition von Campbell und Mixco weist einen schlimmen Denkfehler auf, der im Prinzip unverzeihlich ist. Welcher Fehler ist das?

2 Historische Relationen

Im Folgenden wollen wir versuchen, uns unabhängig von der Frage, ob es nun um die Linguistik oder die Biologie oder dergleichen geht, klarzumachen, welche historischen Relationen wichtig für die historischen Wissenschaften sind. Die grundlegende Frage, von der wir uns dabei leiten lassen wollen, ist die, ob, und wenn ja welche, „Geschichte“ zwei oder mehr Objekte (seien es nun Wörter oder Gene) miteinander teilen.

Welche Wissenschaften können angeführt werden, um das oben genannte „dergleichen“ mit Inhalt zu füllen?

2.1 Ontologische Relationen

Mit *ontologischen Relationen* sollen im Folgenden diejenigen Relationen bezeichnet werden, die, wenn wir gesunden Menschenverstand annehmen, für Merkmale von historischen Objekten gelten, ob wir diese Relationen nun epistemologisch entdecken können, oder nicht. Da die grundlegende Frage ist, ob zwei oder mehr Merkmale historischer Objekte eine gemeinsame Geschichte miteinander teilen, und wenn dies der Fall ist, welcher Art diese gemeinsame Geschichte ist, können wir für diese Merkmale eine Reihe grundlegender ontologischer Relationen ansetzen.

Gemeinsame Abstammung

Für zwei oder mehr Merkmale *characters* unterschiedlicher Objekte gilt die Relation gemeinsamer Abstammung, wenn diese Merkmale zu einem früheren Zeitpunkt in der Geschichte einmal die Merkmale eines einzigen Objekts waren.

Nenne Beispiele für gemeinsame Abstammung in a) der Linguistik, b) der Biologie, und c) der Theologie.

Untertypen der gemeinsamen Abstammung

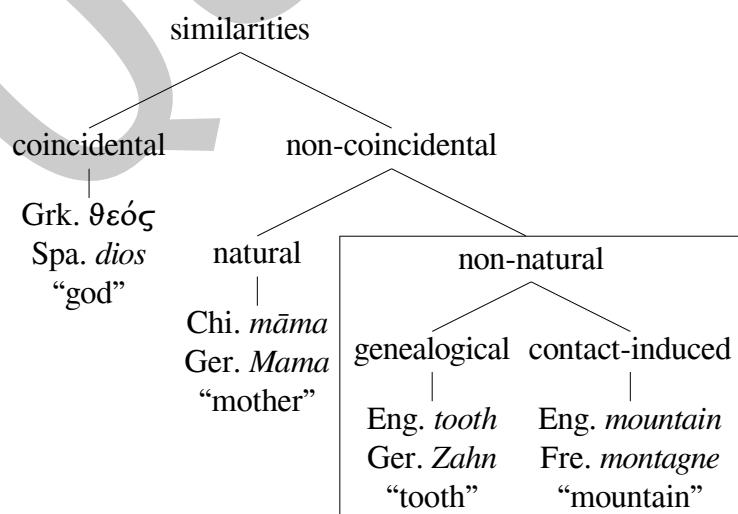
Wir können die grundlegende Relation der gemeinsamen Abstammung weiter ergänzen, indem wir die folgenden Untertypen ansetzen:

- **direkte gemeinsame Abstammung** bezeichnet eine Abstammungsrelation zwischen Merkmalen, die lediglich aus der Auseinanderentwicklung („Spaltung“) der Objekte, welche diese Merkmale aufweisen, resultiert und keine weitere Veränderung der Merkmale einschließt;
- **indirekte gemeinsame Abstammung** bezeichnet eine Abstammungsrelation zwischen Merkmalen, die eine nach der Auseinanderentwicklung der Objekte, welche die Merkmale aufweisen, unabhängige Weiterentwicklung von mindestens einem der Merkmale aufweist; und
- **nicht-vertikale gemeinsame Abstammung** bezeichnet eine Abstammungsrelation zwischen Merkmalen, die durch den (horizontalen) Transfer von mindestens einem der Merkmale zwischen den Objekten, die die Merkmale aufweisen, gekennzeichnet ist.

Es ist sicher klar, dass diese Relationen sich stark an den biologischen Konzepten der "Orthologie", "Paralogie" und "Xenologie" orientieren. Welche Beispiele gibt es nun aber in der historischen Linguistik für diese drei Untertypen der gemeinsamen Abstammung?

2.2 Epistemologische Relationen

Was (ontologisch) der Fall ist, kann uns (epistemologisch) durchaus verborgen bleiben! Um die ontologischen Relationen zwischen Merkmalen zu erschließen, müssen wir immer dann, wenn die Vorfahren der Objekte (wie Sprachen oder Spezies), die wir untersuchen, nicht erhalten sind, auf Indizien zurückgreifen. Das grundlegende Prinzip, mit dem Geschichte in der Biologie, der Linguistik und anderen historischen Wissenschaften dabei rekonstruiert wird, ist der Vergleich von Objekten, welcher realisiert wird als Vergleich von Merkmalen. Die grundlegende Relation, auf die wir uns als Indiz stützen, ist eine *Ähnlichkeitsrelation*, wobei die Frage, wie Ähnlichkeit konkret in der Praxis der jeweiligen Wissenschaften definiert ist, durchaus variieren kann.



Die oben wiedergegebene Abbildung zeigt grundlegende Typen von Ähnlichkeiten zwischen Wörtern verschiedener Sprachen mit Hilfe einer Entscheidungsbaum-Struktur. Bisher wurden unterschiedlichste Konzepte und Termini zur Beschreibung historischer Relationen besprochen. Welche Termini bieten sich an, um die Knoten des Baumes als Relationen zu fassen?

3 Zeichenrelationen

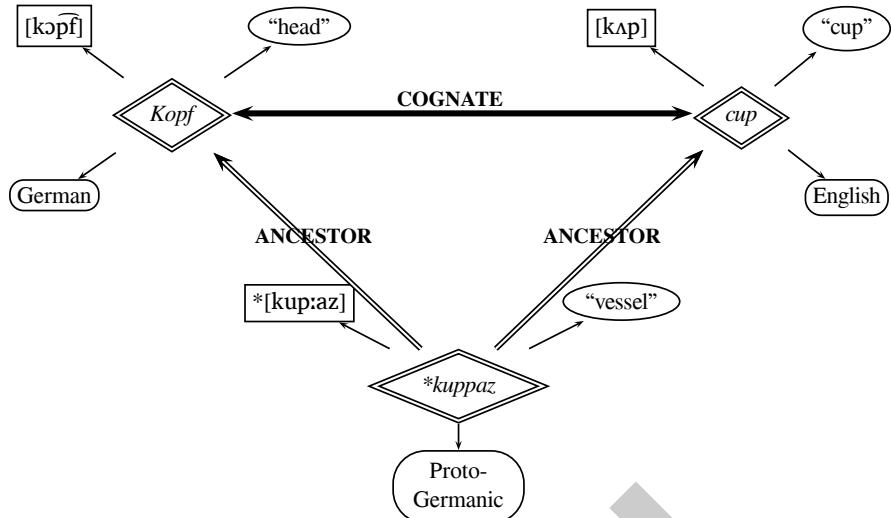
Wenn wir die gängigen, in der Literatur verwendeten Zeichenrelationen in der Linguistik denen in der Biologie gegenüberstellen, und diese wiederum mit den ontologischen historischen Relationen vergleichen, fällt auf, dass die linguistische Terminologie nicht sehr gut geeignet ist, diese ontologischen Relationen auf systematisch ansprechende Weise abzudecken. Während jede der grundlegenden historischen ontologischen Relationen ihren Terminus in der Biologie hat, weist die Linguistik lediglich den Terminus „Kognazität“ auf, der darüber hinaus sowohl direkte als auch indirekte gemeinsame Abstammung bezeichnet. Es ist an der Zeit, diese terminologische Armut zu bekämpfen.

Relations		Terminology	
		Biology	Linguistics
common descent	direct	orthology	?
	indirect	paralogy	cognacy oblique cognacy
	due to lateral transfer	xenology	?

Könnte man beim Betrachten der Tabelle nicht den Eindruck gewinnen, dass die Gegenüberstellung der biologischen und der linguistischen Termini der Linguistik nicht ganz gerecht wird, da die Phänomene, die zu den Relationen führen, Linguisten ja nicht unbekannt sind?

3.1 Kognatenrelationen

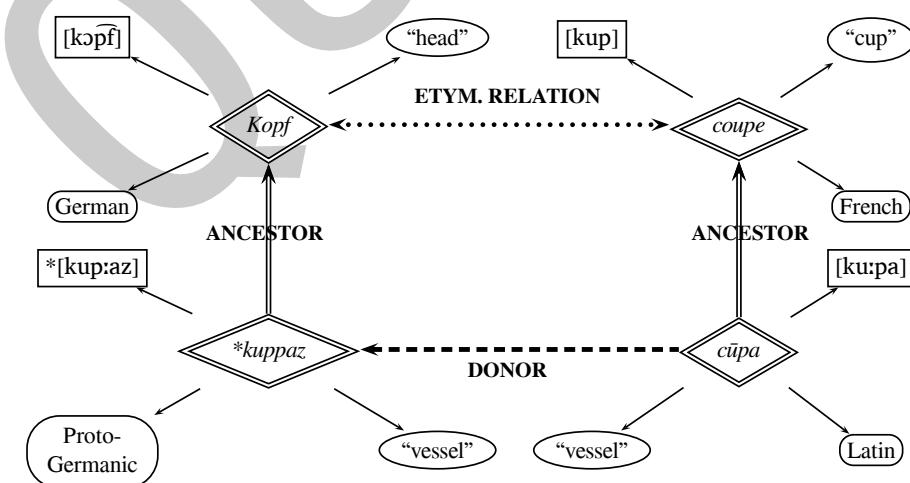
Die Kognazitätsrelation stellt eine wichtige Relation in der historischen Linguistik dar. Im konkreten Fall von sprachlichen Zeichen setzt sie voraus, dass kognate Zeichen direkt von der gemeinsamen Vorgängersprache geerbt wurden, wie das Beispiel auf der folgenden Abbildung zeigt. Voraussetzung für die Kognazität zweier Wörter A und B ist somit, dass für diese Wörter selbst eine Vorgänger-Nachfolger-Relation (engl. *ancestor-descendant relation*) zu einem Wort C einer Vorgängersprache nachweisbar ist.



Worin unterscheidet sich die Vorgänger-Nachfolger-Relation von der Relation der Kognazität? Welche Beispiele kann man für direkte und indirekte (oblique) Kognazität anführen?

3.2 Etymologische Relationen

Wenn zwei Zeichen A und B nicht über ein Zeichen C miteinander verbunden werden können, zu dem sowohl A als auch B in einer Vorgänger-Nachfolger-Relation stehen, dann heißt das noch nicht, dass nicht eine historische Relation zwischen A und B bestehen könnte. Die folgende Abbildung illustriert einen solchen Fall. Die beiden Zeichen Deutsch *Kopf* und Französisch *coupe* sind nicht mehr kognat, ihre Entstehungsgeschichte verlief jedoch nicht vollkommen unabhängig voneinander, da *Kopf* der Nachfolger von Proto-Germanisch **kuppaz* ist, welches wiederum von Latein *cūpa* entlehnt wurde. Um diese Beziehung zu benennen, schlage ich den Terminus „oblique etymologische Relation“ vor.



Der Terminus „oblique etymologische Relation“ wurde relativ unvermittelt eingeführt. Gibt es eine Möglichkeit, aus dem bisher gelernten zu erschließen, warum der Terminus gewählt wurde?

3.3 Zusammenfassung

In der folgenden Tabelle werden die neuen Termini zusammengefasst und den ontologischen Relationen gegenübergestellt. Um deutlich zu machen, dass es sich um Relationen handelt, werden die Termini auch als solche benannt, wobei die traditionelle Relation der Kognazität nicht ersetzt, sondern durch zwei Spezialfälle ergänzt wird. Der „Homologie“ in der Biologie entsprechend wurde der Terminus „etymologische Relation“ gewählt, da in einer dem Gebrauch des Terminus „Homologie“ analogen Weise – nicht nur in der Linguistik – häufig von einer „etymologischen Beziehung zwischen Wörtern“ gesprochen wird.

Relations		Terminology	
common descent	direct	etymological relation	direct cognate relation
	indirect		cognate relation
	due to lateral transfer		oblique etymological relation

In dieser Sitzung wurde versucht, strikt darauf zu achten, die historischen Relationen jeweils strikt für Merkmale (im Gegensatz zu taxonomischen „Objekten“) zu definieren. Ist das gerechtfertigt, oder könnte man die Beziehungen nicht vielleicht gewinnbringend auf historische Beziehungen zwischen Sprachen übertragen?

Literatur

- Campbell, L. and M. Mixco (2007). *A glossary of historical linguistics*. Edinburgh: Edinburgh University Press.
- Fitch, W. M. (2000). “Homology. A personal view on some of the problems”. In: *Trends in Genetics* 16.5, 227–231.
- Koonin, E. V. (2005). “Orthologs, paralogs, and evolutionary genomics”. In: *Annual Review of Genetics* 39, 309–338.
- Morrison, D. A. (2006). “Multiple sequence alignment for phylogenetic purposes”. In: *Australian Systematic Botany* 19, 479–539.
- Wagner, G. P. (2000). “Characters, unit and natural kinds. An introduction”. In: ed. by G. P. Wagner. San Diego et al.: Academic Press, 1–12.
- Wegnez, M. (1987). “Letter to the editor”. In: *Cell* 51, 516.

Literatur

- Campbell, L. and M. Mixco (2007). *A glossary of historical linguistics*. Edinburgh: Edinburgh University Press.
- Fitch, W. M. (2000). “Homology. A personal view on some of the problems”. In: *Trends in Genetics* 16.5, 227–231.

- Koonin, E. V. (2005). "Orthologs, paralogs, and evolutionary genomics". In: *Annual Review of Genetics* 39, 309–338.
- Morrison, D. A. (2006). "Multiple sequence alignment for phylogenetic purposes". In: *Australian Systematic Botany* 19, 479–539.
- Wagner, G. P. (2000). "Characters, unit and natural kinds. An introduction". In: ed. by G. P. Wagner. San Diego et al.: Academic Press, 1–12.
- Wegnez, M. (1987). "Letter to the editor". In: *Cell* 51, 516.

QUAHL

„Konservativ kommt nicht von Konserve!“

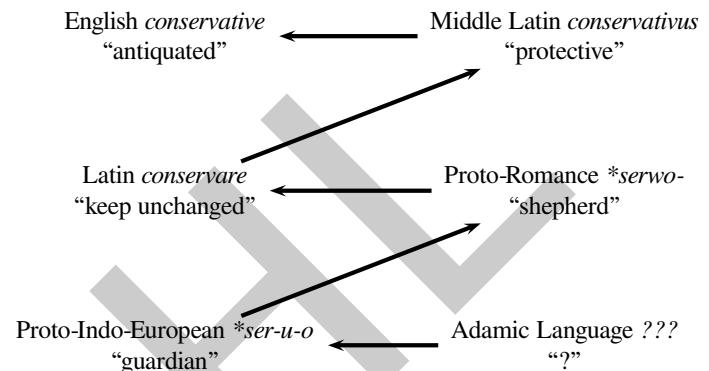
1 Kommt „konservativ“ wirklich nicht von „Konserve“?

„Konservativ kommt nicht von Konserve.“

(Angela Merkel am 10. April 2000)

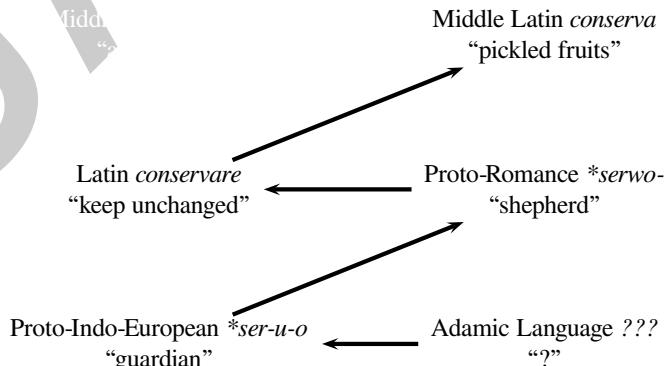
1.1 Woher kommt „konservativ“?

<i>konservativ</i>	
Part of Speech	adjective
frequency	moderate
meaning	“sticking to the traditional”
synonyms	“rückschrittlich”, “antiquiert”, “rückständig”, “unzeitgemäß”
pronunciation	kɔnservati:f
origin	English <i>conservative</i>

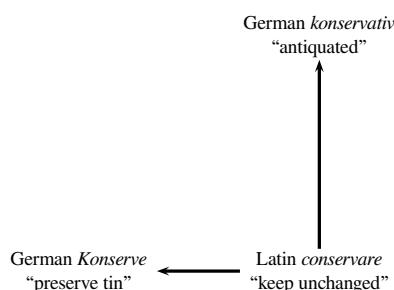
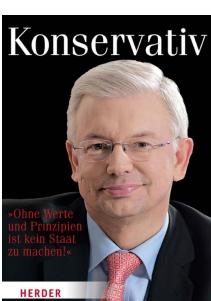


1.2 Woher kommt „Konserve“?

<i>Konserve</i>	
Part of Speech	noun
frequency	moderate
meaning	“preserve tin”
synonyms	“Büchse”, “Konservebüchse”, “Konservendose”
pronunciation	kɔnzərvə
origin	Middle Latin <i>conserva</i>



1.3 Woher kommen „konservativ“ und „Konserve“?



2 Was ist die komparative Methode?

2.1 Die komparative Methode in der Literatur

Wikipedia-Eintrag zur komparativen Methode:

In linguistics, the comparative method is a technique for studying the development of languages by performing a feature-by-feature comparison of two or more languages with common descent from a shared ancestor, as opposed to the method of internal reconstruction, which analyzes the internal development of a single language over time. (*Wikipedia: Comparative Method*)

→ Eine Technik für das Studium der Entwicklung von Sprachen

Eintrag aus dem Oxford Dictionary of Linguistics:

The method of comparing languages to determine whether and how they have developed from a common ancestor. The items compared are lexical and grammatical units, and the aim is to discover correspondences relating sounds in two or more different languages, which are so numerous and so regular, across sets of units with similar meanings, that no other explanation is reasonable. (Matthews 1997)

→ Eine Technik für die Untersuchung von Sprachverwandtschaften

Eintrag im Metzler Lexikon Sprache

Kein Eintrag vorhanden (Glück 2000)

→ Etwas Marginales, das nicht explizit in einem Wörterbuch besprochen werden muss

Eintrag im Lexikon der Sprachwissenschaft:

→ Rekonstruktion (Bußmann 2002)

→ Gleichzusetzen mit linguistischer Rekonstruktion

Fox' Ansichten zur komparativen Methode:

The comparative is both the earliest and the most important of the methods of reconstruction. Most of the major insights into the prehistory of languages have been gained by the applications of this method, and most reconstructions have been based on it. (Fox 1995: 17)

→ Eine bestimmte Rekonstruktionsmethode

Fleischhauers Definition der komparativen Methode

The Comparative Method is the central tool in historical linguistics for historical reconstruction and also classifying languages. A classification done with the Comparative Method is called a genetic classification. The result is that languages are arranged in language family trees. This means that languages are classified according to their genealogical relationships² and are interpreted as being in relation of child- or sisterhood to other languages. Such a way of classifying entities is called phylogenetic classification in biology; a classification by genealogical relationships. (Fleischhauer 2009)

→ Ein Verfahren zur Rekonstruktion und zur genetischen Sprachklassifikation

Klimovs Ansichten zur historisch-vergleichenden Methode

Die Methode der heutigen Komparativistik, welche allgemein unter dem nicht sehr glücklichen Terminus "vergleichend-historische Methode" bekannt ist, stellt eine große Gesamtheit an abstrakten und konkreten Verfahren zur Untersuchung der Geschichte verwandter Sprachen dar, die genetisch auf eine bestimmte einheitliche Tradition der Vergangenheit zurückgehen, welche man üblicherweise als Proto-Sprache oder Grundsprache qualifiziert. Dieses methodische Instrumentarium, auf welches zurückgegriffen wird, um eine große Menge verschiedener Probleme zu lösen, wird verwendet, um ein Erkenntnissystem über die historische Entwicklung von Sprachfamilien aufzubauen, welches seine endgültige Gestalt in From historisch-vergleichender Grammatiken erhält. (Klimov 1990: 6) ¹

→ Oberbegriff für die Verfahren der historisch-vergleichenden Sprachwissenschaft

2.2 Auf der Suche einer einheitlichen Definition

- Verfahren zum Nachweis von Sprachverwandtschaft: Harrison (2003), Anttila (1972)
- Verfahren zum Studium von Sprachentwicklung: Matthews (1997), Klimov (1990)
- Spezifisches Rekonstruktionsverfahren: Fox (1995), Anttila (1972), Rankin (2003), Hoenigswald (1960), Lehmann (1969)
- Gleichzusetzen mit linguistischer Rekonstruktion: Makaev (1977), Bußmann (2002), Glück (2000), Jarceva (1990)
- Verfahren zur Klassifikation von Sprachen: Fleischhauer (2009)
- Oberbegriff für die traditionellen Verfahren der historisch-vergleichenden Sprachwissenschaft: Klimov (1990)

¹Meine Übersetzung, Originaltext: «Методика современной компаративистики, широко известная в лингвистической литературе под довольно неудачным термином "сравнительно-исторический метод", представляет собой большую совокупность методов и конкретных приемов изучения истории родственных языков, генетически восходящих к некоторой единой традиции прошлого, обычно квалифицируемой в качестве праязыка или языка-основой. Этот методический инструментарий, призванный обслуживать решение множества задач, используется для построения системы знаний об историческом развитии языковых семей, формируемой в конечном счете в виде сравнительно-исторических грамматик».

2.3 Komparative Methode: Ein vorläufiger Definitionsversuch

Die komparative Methode ist ein Komplex von Verfahren der historischen Sprachwissenschaft, mit deren Hilfe Sprachen klassifiziert und nicht belegte Sprachstufen rekonstruiert werden, um somit Entwicklungsgeschichte von Sprachen zu schreiben. Die Ergebnisse der komparativen Methoden werden in Form von etymologischen Wörterbüchern, historischen Grammatiken und Entwicklungsschemata (evolutionäre Bäume und Netze) kodiert.

Wenn diese Definition stimmt (was ja nicht unbedingt der Fall sein muss, was ist die „komparative Methode“ dann eigentlich, wenn man sie in den Bereich der historischen Sprachwissenschaft einordnen möchte?

3 Näheres zur komparativen Methode

3.1 Die Beschreibung des Verfahrens bei Ross und Durie

1. Determine on the strength of diagnostic evidence that a set of languages are genetically related, that is, that they constitute a 'family';
2. Collect putative cognate sets for the family (both morphological paradigms and lexical items).
3. Work out the sound correspondences from the cognate sets, putting 'irregular' cognate sets on one side;
4. Reconstruct the protolanguage of the family as follows:
 - a Reconstruct the protophonology from the sound correspondences worked out in (3), using conventional wisdom regarding the directions of sound changes.
 - b Reconstruct protomorphemes (both morphological paradigms and lexical items) from the cognate sets collected in (2), using the protophonology reconstructed in (4a).
5. Establish innovations (phonological, lexical, semantic, morphological, morphosyntactic) shared by groups of languages within the family relative to the reconstructed protolanguage.
6. Tabulate the innovations established in (5) to arrive at an internal classification of the family, a 'family tree'.
7. Construct an etymological dictionary, tracing borrowings, semantic change, and so forth, for the lexicon of the family (or of one language of the family).

Tabelle 1: Komparative Methode bei Ross and Durie (1996)

Was ist in diesem Zusammenhang wohl mit „diagnostic evidence“ gemeint?

3.2 Die wichtigsten Verfahren der komparativen Methode

- **Filiationsverfahren:**² Der Nachweis von Sprachverwandtschaft ist eine Grundvoraussetzung für die Anwendung der linguistischen Rekonstruktion. Im Rahmen der komparativen Methode muss dieser Nachweis als erster Schritt des Verfahrens erbracht werden. Auf welche Art dies geschieht, ist jedoch stark umstritten in der historischen Linguistik. Grundsätzlich gibt es hier widerstreitende Ansichten bezüglich der Aussagekraft lexikalischer gegenüber 'grammatischen' Evidenzen, vgl. die Diskussion in Dybo and Starostin (2008).
- **Rekonstruktionsverfahren:** Rekonstruktionsverfahren sind für sich genommen bereits sehr komplex. Unterschieden werden insbesondere die beiden klassischen Verfahren der externen und der internen Rekonstruktion, sowie die zuätzlichen Verfahren der philologischen Rekonstruktion und der sprachtypologischen Validierung, vgl. bspw. die Beschreibung der Verfahren in Fox (1995).
- **Klassifikationsverfahren:** Zwar wird in der linguistischen Literatur gewöhnlich behauptet, die komparative Methode schließe eindeutige Verfahren der Sprachklassifikation mit ein, jedoch sind viele Fragen hier noch ungeklärt. Während die grobe Klassifizierung von Sprachen keine Probleme bereitet, gibt es bei der Subgruppierung genetisch entfernter verwandter Sprachen, sowie bei der genetischen Klassifizierung von Dialekten oder sehr nah verwandten Sprachen erhebliche Schwierigkeiten, da die traditionellen Kriterien widersprüchliche Ergebnisse liefern. Das grundlegende Prinzip der Klassifikation ist im Rahmen der komparativen Methode jedoch die Identifizierung von Innovationen (Wandelphänomene, die sich für eine bestimmte Gruppe innerhalb einer Sprachfamilie von anderen unterscheiden).³

Dass ein solches Verfahren wie das der komparativen Methode überhaupt möglich ist, setzt bestimmte Grundannahmen voraus. Eine sehr wichtige Grundannahme ist dabei die Annahme Universalität von Sprachwandel, also die Annahme, dass Sprachwandel unabhängig von Ort und Zeit verläuft. Zwei weitere sind die Annahme der Regelmäßigkeit von Lautwandel und die Annahme der Möglichkeit von Sprachspaltung. Warum sind diese Grundannahmen so wichtig für die Verfahren der komparativen Methode?

3.3 Beispiele für die "Erkenntnissysteme" der komparativen Methode

Etymologische Wörterbücher

²Begriff „Filiation“ habe ich geklaut von Gévaudan (2007). Der verwendet ihn zwar nur für einzelne Wörter verschiedener Sprachen, jedoch scheint er mir auch als übergeordneter Terminus für das ansonsten sehr sperrig klingende Problem „Nachweis von Sprachverwandtschaft“ sinnvoll zu sein.

³Dies wird schon bei Brugmann (1886[1897]: 24) deutlich, wenn dieser in Bezug auf die Subgruppierung der indogermanischen Sprachen schreibt: „Das einzige nun, was auf das Verhältnis der einzelnen Sprachzweige zu einander, auf die Art des Hervorgangs der Einzelsprachen aus der idg. Ursprache Licht werfen kann, sind die besonderen Übereinstimmungen zwischen je zwei oder mehreren von ihnen, die Neuerungen, durch die jedesmal gewisse Sprachzweige gegenüber den andern in der Entwicklung vorangeschritten erscheinen.“

- (1) ***sanðaz** ~ ***sanðan** sb.m./n.: ON *sandr* ‘sand’ , OE *sand* id., OFris *sand*, *sond* id., OS *sand* id., OHG *sant* id. From an earlier **samðaz*. Close to Gk ἄμαθος ‘sand’ . Probably derived from **sem-* ‘to pour’ : Gk ἀμάομαι ‘to gather’ , OIr *to-ess* ‘to pour out’ , Lith *semiù*, *sémti* ‘to scoop’ . Pedersen *Kelt. Gr.* II 624; Torp-Falk 430–431; Holthausen *AEEW* 270; Kuiper *Festschr. Kretschmer* 218; Pokorny I 146 (to Gk ψάμμος ‘sand’), 901–902; Fraenkel 774–775; Vries *ANEW* 462; Zalizniak II 171; Frisk I 84; Onions 786; Kluge-Seibold 703.⁴

Sprachklassifikation

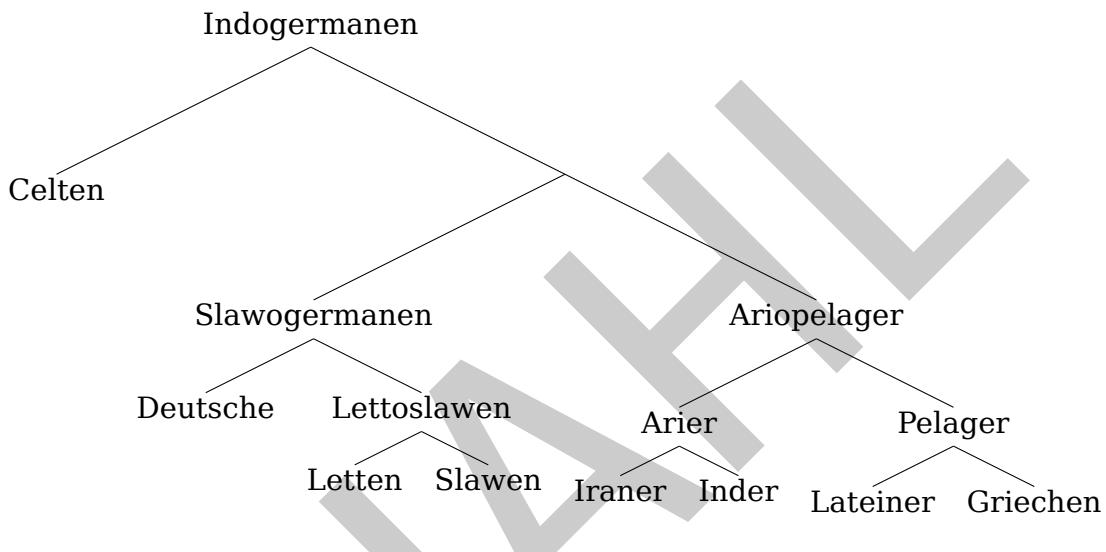


Tabelle 2: Früher Stammbaum der indogermanischen Völker von Schleicher (1853)

Historische Grammatiken

- (2) MANIOS MED VHEVHAKED NUMASIOI
 Manius me fecit Numerio
 Manius mich machen (3.Sg.Perf.) Numerios (Dat.)
 „Manius hat mich für Numerius gemacht“⁵

4 Eine kleine Übungsaufgabe

Die folgende Übungsaufgabe wurde von mir aus dem Russischen übersetzt. Ursprünglich wurde sie von I. B. Itkin verfasst und in Burlak and Starostin (2005: 394f) veröffentlicht.

4.1 Ausgangspunkt

In der zweiten Hälfte des ersten Jahrtausend n. Chr. sprach man im chinesischen Turkestan zwei miteinander nah verwandte (und später ausgestorbene) Sprachen, die zur indogermanischen Sprachfamilie gehörten: Tocharisch A und Tocharisch B. Diese Sprachen waren Nachkommen eines gemeinsamen (gemeintocharischen) Vorgängers.

Ein Linguist studierte die Lautkorrespondenzen zwischen Tocharisch A und Tocharisch B. Folgende Fakten waren ihm bekannt:

⁴Beispiel aus: Orel (OREL: 318)

⁵Beispiel aus: Meiser (1998: 4)

1. Im Gemeintocharischen gab es den Laut o;
2. In Tocharisch B erhielt sich dieser Laut in allen Wörtern ohne jegliche Veränderung;
3. Wie und nach welchen Regeln sich dieser Laut in Tocharisch A veränderte (oder ob er sich überhaupt veränderte), dies ist unter den Forschern umstritten;
4. In tocharisch A gibt es viele Wörter, die aus Tocharisch B entlehnt wurden, was die Arbeit des Linguisten stark erschwert.

Nachdem er darüber nachgedacht hatte, schrieb der Linguist auf einem Blatt Papier einige Wörter nieder, die einander in Tocharisch A und B entsprechen:

Tocharisch A	Tocharisch B	Übersetzung
pokeñ	pokaiñ	Hände
amok	amok	Kunst
lokit	laukito	Gast
kos	kos	wieviel
oñkrac	oñkrocce	unsterblich
mok	moko	alt
potatär	pautotär	er schmeichelt
ka-ts	ka-tso	Bauch
orpank	orpoñk	Podium
oko	oko	Frucht
wärtant	wärtonta	Wald
oktat	oktatse	aus acht Teilen bestehend

Tabelle 3: Die Wortliste des Linguisten

Als er diese Liste durchging, rief der Linguist plötzlich: "Alles klar: das gemeintocharische o hat sich in Tocharisch A folgendermaßen erhalten: ... Aber das bedeutet, dass die Wörter X und Y in Tocharisch A aus Tocharisch B entlehnt wurden!"

4.2 Aufgaben

1. Formulieren Sie die Regel, die der Linguist entdeckt hatte.
2. Welche Wörter im Ausruf des Linguisten wurden durch X und Y ersetzt? Erläutern Sie Ihre Entscheidung.
3. Welche weiteren Wörter, die in der Aufgabe vorgestellt wurden, könnten prinzipiell ebenfalls aus Tocharisch B entlehnt worden sein?
4. Bestimmen Sie, wie die den folgenden Wörtern aus Tocharisch B entsprechen Wörter in Tocharisch A aussehen müssten (keines dieser Wörter ist entlehnt aus Tocharisch A): *ponta* "alle (Frauen)", *okaro* "grasartiges Gewächs", *yok* "Licht", *mokoc* "großer Finger".

4.3 Anmerkungen

y wird wie das deutsche j ausgesprochen (IPA: [j]); c in etwa wie tsch (ein palataler Affrikat, wahrscheinlich IPA [tʃ̩]); ts wie z (IPA: [ts]); ñ wie nj (IPA: [ŋ]); ñ stellt eine Variante von n vor k dar (wahrscheinlich IPA [ŋ̩]); a- wie ein langes a; ä stellt einen besonderen tocharischen Vokal dar.

Literatur

- Anttila, R. (1972). *An introduction to historical and comparative linguistics*. New York: Macmillan.
- Brugmann, K. (1886 [1897]). *Einleitung und Lautlehre*. In: *Vergleichende Laut-, Stammbildungs- und Flexionslehre der Indogermanischen Sprachen*. 2nd ed. Vol. 1. Grundriss der vergleichenden Grammatik der indogermanischen Sprachen 1. Strassburg: Trübner.
- Burlak, S. A. and S. A. Starostin (2005). *Sravnitel'no-istoričeskoe jazykoznanie* [Comparative-historical linguistics]. Moscow: Akademia.
- Bußmann, H., ed. (2002). *Lexikon der Sprachwissenschaft*. 3rd ed. Stuttgart: Kröner.
- Dybo, A. and G. Starostin (2008). "In defense of the comparative method, or the end of the Vovin controversy". In: *Aspekty komparativistiki*. Vol. 3. Ed. by I. S. Smirnov. Orientalia et Classica XI. Moscow: RGGU, 119–258. PDF: <http://starling.rinet.ru/Texts/compmeth.pdf>.
- Fleischhauer, J. (2009). "A Phylogenetic Interpretation of the Comparative Method". In: *Journal of Language Relationship* 2, 115–138.
- Fox, A. (1995). *Linguistic reconstruction. An introduction to theory and method*. Oxford: Oxford University Press.
- Glück, H., ed. (2000). *Metzler-Lexikon Sprache*. 2nd ed. Stuttgart: Metzler.
- Gévaudan, P. (2007). *Typologie des lexikalischen Wandels. Bedeutungswandel, Wortbildung und Entlehnung am Beispiel der romanischen Sprachen*. Stauffenburg-Linguistik ; 45. Tübingen: Stauffenburg.
- Harrison, S. P. (2003). "On the Limits of the Comparative Method". In: *The Handbook of Historical Linguistics*. Ed. by B. D. Joseph and R. D. Janda. Blackwell, 213–243.
- Hoenigswald, H. M. (1960). "Phonetic similarity in internal reconstruction". In: *Language* 36.2, 191–192. JSTOR: 410982.
- Jarceva, V. N., ed. (1990). *Lingvističeskij énciklopedičeskij slovar* (Linguistical encyclopedic dictionary). Moscow: Sovetskaja Enciklopedija.
- Klimov, G. A. (1990). *Osnovy lingvističeskoj komparativistiki* [Foundations of comparative linguistics]. Moscow: Nauka.
- Lehmann, W. P. (1969). *Einführung in die historische Linguistik*. Trans. by R. Freudenberg. Autorisierte, vom Verfasser durchgesehene Übersetzung von Rudolf Freudenberg. Heidelberg: Carl Winter.
- Makaev, E. A. (1977). *Obščaja teorija sravnitel'nogo jazykoznanija* [Common theory of comparative linguistics]. Moscow: Nauka.
- Matthews, P. H., ed. (1997). *Oxford concise dictionary of linguistics*. Oxford: Oxford University Press.
- Meiser, G. (1998). *Historische Laut- und Formenlehre der lateinischen Sprache*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Orel, V., comp. (2003). *A handbook of Germanic etymology*. Leiden: Brill.
- Rankin, R. L. (2003). "The Comparative Method". In: *The handbook of historical linguistics*. Ed. by B. D. Joseph and R. D. Janda. Blackwell handbooks in linguistics. Malden, Mass.: Blackwell, 183–212.
- Ross, M. D. and M. Durie (1996). "Introduction". In: *The comparative method reviewed: Regularity and irregularity in language change*. Ed. by M. Durie. New York: Oxford University Press, 3–38.
- Schleicher, A. (1853). "Die ersten Spaltungen des indogermanischen Urvolkes". In: *Allgemeine Monatsschrift für Wissenschaft und Literatur* 3, 786–787.
- Wikipedia*. Wikipedia. The Free Encyclopedia. URL: <http://en.wikipedia.org>.

Formen der Sprachverwandtschaft

1 Verwandtschaft

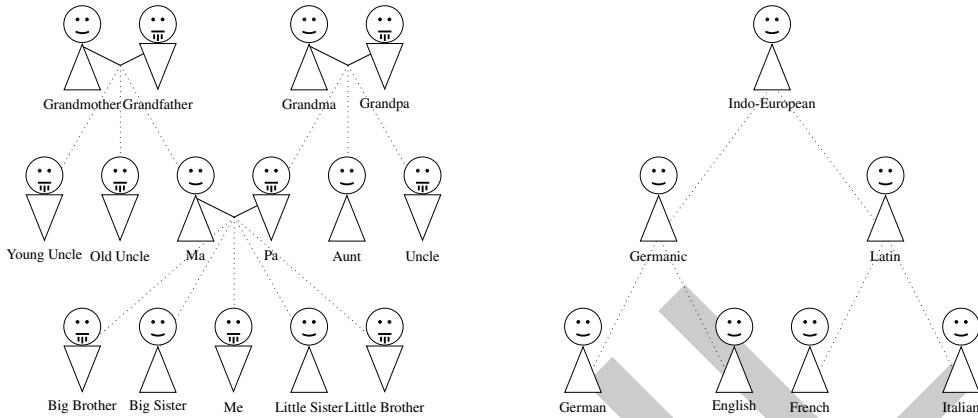


Abbildung 1: Verwandtschaft unter Individuen und Arten

	Individuen	Arten
Vererbung	sporadisch	
Wandel	diskret	
Ursprung	Produktion	

Tabelle 1: Verwandtschaft unter Individuen und Arten

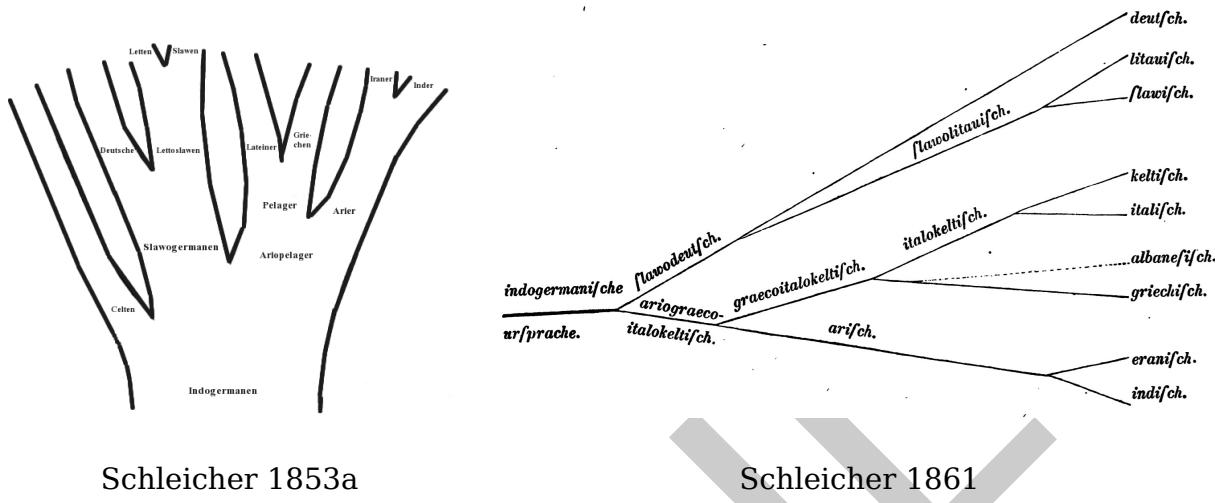
In Abbildung 1 wird der Versuch unternommen, die Unterschiede in Bezug auf die Verwandtschaft von Individuen und Arten zu kontrastieren. Ist die Abbildung, welche die Verwandtschaft von Arten zeigt realistisch genug, oder könnte man sie noch verfeinern, und wenn ja, wie? In der Tabelle wurden bewusst keine Einträge zur Charakterisierung der Verwandtschaft von Arten hinsichtlich der Punkte »Vererbung«, »Wandel« und »Ursprung« gemacht. Wie lassen sich diese characterisieren?

2 Bäume

2.1 Historisches

In den vorherigen Sitzungen wurde ja nun bereits öfter darauf hingewiesen, dass August Schleicher (1821 – 1868) die „Stammbaumtheorie“ in der historischen Linguistik populär machte (Schleicher 1853a, Schleicher 1853b). Die Entwicklung dieser Theorie war keineswegs ein naheliegender, zwangsläufiger Schritt, wie insbesondere dann deutlich wird, wenn man sich die Unterschiede zwischen dem Konzept der Verwandtschaft von Individuen und der Verwandtschaft von Arten noch einmal in Erinnerung ruft, aber auch, wenn man die neue Verwandtschaftsmetapher Schleichers mit der im hebräischen Paradigma gepflegten Verwandtschaftsmetapher vergleicht: „Verwandtschaft der Sprachen“ im Rahmen des hebräischen

Paradigmas war ja dem ursprünglichen Verwandtschaftskonzept viel näher gewesen (sporadische Vererbung, diskreter Wandel, singulärer Ursprung) als die „Verwandtschaft der Sprachen“ in Schleichers Stammbäumen (Geisler and List 2013).



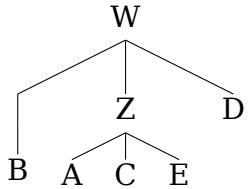
Oben sind zwei Stammbäume August Schleichers abgebildet, ein früher Stammbaum aus dem Jahre 1853 und ein späterer aus dem Jahre 1861. Schleicher selbst kommentierte den zweiten Stammbaum mit den Worten »Die ältesten teilungen des indogermanischen bis zum entstehen der grundsprachen der den sprachstamm bildenden sprachfamilien lassen sich durch folgendes schema anschaulich machen. Die längte der linien deutet die zeitdauer an, die entfernung derselben von einander den verwantschaftsgrad« (ibid.: 6f). Was meint Schleicher eigentlich mit dieser Aussage, und was fällt noch an dem neuen Stammbaum auf?

2.2 Formales

In der Graphentheorie ist ein „Baum“ ein ungerichteter Graph in dem alle Kanten durch einen einfachen Pfad verbunden sind. Ein Baum enthält also keine Zyklen. Der Terminus „Baum“ (engl. „tree“) geht wohl auf zurück, der ihn in einem Artikel von 1857 einführt (Cayley 1857).

Ein Baum stellt eine hierarchische Struktur dar: Ausgehend von einer *Wurzel* (engl. *root*) werden Elemente mit Linien verbunden. Die Elemente (einschließlich der Wurzel) werden dabei *Knoten* (engl. *nodes*) genannt, und die Linien *Zweige* (*branches*). Die Endpunkte der Hierarchie, also die Knoten, die keine *Kinder* haben, werden *Blätter* genannt.

Bäume lassen sich auf unterschiedliche Weise darstellen. Ein gängiges Format in der Evolutionsbiologie, um Bäume in einfacher Textform zu speichern, ist das sogenannte *Newick*-Format. Dieses Format macht sich die Tatsache zunutze, dass Bäume verschachtelten Ausdrücken entsprechen (ibid.). Hier ist ein Beispiel für einen Baum und drei mögliche Arten, den Baum in Newick-Format darzustellen:

Baum**Newick**

- (a) (B,(A,C,E),D);
- (b) (B,(A,C,E)Z,D)W;
- (c) (B:2,(A:1,C:1,E:1):1,D:1);

Was genau ist der Unterschied in den drei Newick-Varianten? Lassen sich Varianten (b) und (c) verbinden?

Wie sehen die Bäume für die folgenden drei Newick-Formeln aus?

- (a)((A,B),(C,D)),E);
- (b)((A,B,C)ABC,((D,E)DE,F)DEF)ABCDEF;
- (c)((A:1,B:1):2,(C:1,D:2):3);

Wie kann man den Baum von Schleicher aus dem Jahre 1861 in Newick-Format schreiben?

2.3 Probleme

Die Probleme der Bäume betreffen selbstverständlich ihre fehlende Aussagekraft in Bezug auf jegliche Art von Kontakt- und Konvergenzphänomenen. Da der grundlegende Mechanismus von baumartiger evolutionärer Entwicklung durch Spaltung und unabhängige Weiterentwicklung gekennzeichnet ist, vereinfachen Bäume die Komplexität der Sprachgeschichte.

Welche Möglichkeit gäbe es, Kontaktphänomene in Sprachstammbäumen Rechnung zu tragen?

3 Wellen

3.1 Historisches

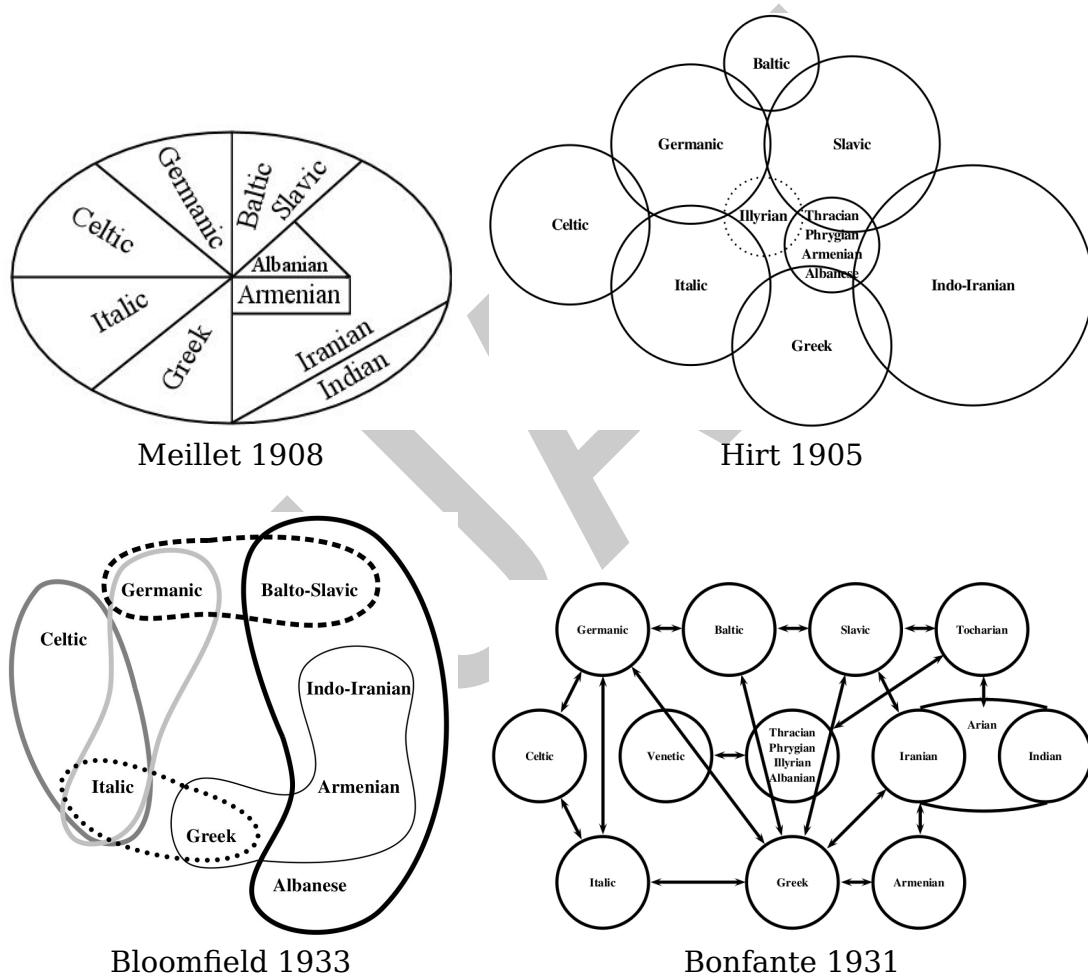
Im Gegensatz zur Biologie, wo die Stammbäume das führende Paradigma wurden, kurz nachdem sie zum ersten Mal postuliert worden waren, verlor sich die Popularität der Bäume in der historischen Linguistik nicht lange, nachdem Schleicher sie begründet hatte. 1872 schrieb Johannes Schmidt (1843 - 1901) das Buch *Die Verwantschaftsverhältnisse der indogermanischen Sprachen*, in welchem er auf verschiedenste Probleme bezüglich der Plausibilität des Stammbaummodells hincwies.

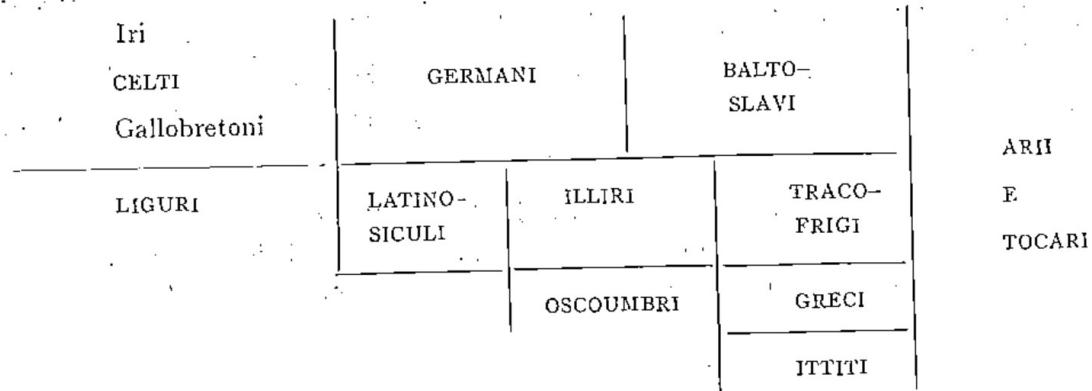
Man mag sich also drehen und wenden wie man will, so lange man an der anschauung fest hält, dass die in historischer zeit erscheinenden sprachen durch merfache gabelungen aus der ursprache hervorgegangen seien,d.h. so lange man einen stammbaum der indogermanischen sprachen annimmt, wird man nie dazu gelangen alle die hier in frage stehenden tatsachen wissenschaftlich zu erklären. (Schmidt 1872: 17)

Er kam zu dem Schluss, dass die indogermanischen Sprachdaten schlicht und einfach *nicht* auf einfache baumartige Differenzierung hinweisen. Anstelle des Baums setzte er die sogenannte *Wellentheorie*, derzufolge bestimmte Wandelphänomene sich wie Wellen in konzentrischen Kreisen um ein Zentrum ausbreiten.

Ich möchte an seine [des Baumes] stelle das bild der welle setzen, welche sich in concentrischen mit der entfernung vom mittelpunkte immer schwächer werdenden ringen ausbreitet. (Schmidt 1872: 27)

Obwohl viele Forscher skeptisch gegenüber dem Stammbaummodell eingestellt waren, taten sie sich schwer mit dem Wellenmodell. Besonders die Formalisierung und Visualisierung stellte eine fast unüberwindliche Schwierigkeit dar. Die Suche nach neuen Metaphern zeit sich in einer Vielzahl unterschiedlichster Visualisierungen der Wellentheorie, die in der Folgezeit veröffentlicht wurde. Diese reichte von einfachen geographischen Karten (Meillet 1922[1908], Pisani 1933, Schmidt 1875) über überlappende Kreise (Hirt 1905), alternierende Grenzen von *Isoglossen* (Bloomfield 1933[1973]), bis hin zu statischen Ähnlichkeitsnetzwerken (Bonfante 1931).





Pisani 1933

Woran mag es wohl liegen, dass sich die Forscher so schwer taten, die Wellentheorie einheitlich zu visualisieren? In welcher linguistischen Teildisziplin spielt die Wellentheorie die bedeutendste Rolle?

3.2 Formales

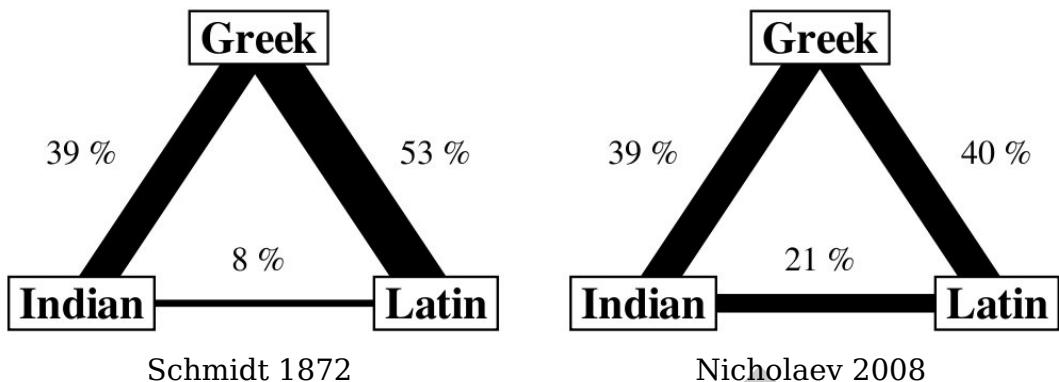
Formal stellt sich tatsächlich die Frage, wie man die Wellentheorie denn nun beschreiben soll. Wenn sie bestimmte Gemeinsamkeiten zwischen Sprachen abbildet, dann scheint der Isoglossenansatz von Bloomfield tatsächlich mit am sinnvollsten zu sein, da er unmittelbar zeigt, worin sich bestimmte Sprachen ähneln. Wenn man der Visualisierung von Hirt folgt, so stellt diese ein sogenanntes „Venn-Diagramm“ dar (benannt nach John Venn, 1834 – 1932), welches meist zur grafischen Veranschaulichung von mengentheoretischen Problemen verwendet wird. Bonfantes Netzwerkansatz stellt selbstverständlich auch eine direkte Lösungsmöglichkeit dar, obwohl hier die Darstellung der Ähnlichkeitsgrade noch nicht visualisiert wurde.

Worin besteht das Problem der Visualisierung der Wellentheorie mit Hilfe von Venn-Diagrammen? Welche Vorteile birgt die Isoglossenvisualisierung?

3.3 Probleme

Ausgangspunkt der Wellentheorie war in der historischen Linguistik die Kritik an den Stammbäumen. Da diese sich nur schwer eindeutig rekonstruieren ließen, wurden sie als unrealistisch kritisiert. Das Problem bei dieser frühen Kritik an dem Stammbaummodell war jedoch, dass sie mathematisch gesehen meist unbegründet war. So beruhte Schmidts ursprüngliche Kritik beispielsweise auf einer Auszählung von indogermanischen Wurzeln, die nur reflexe in verschiedenen Sprachpaaren aufwiesen, wie zum Beispiel geteilte Wurzeln im Lateinischen und Griechischen, im Griechischen und Indischen, und im Indischen und Lateinischen. Schmidt argumentierte nun, dass die komischen Ähnlichkeitsverhältnisse hier keinen Spielraum für die Rekonstruktion von Bäumen ließen. Was er dabei aber missachtete, war das Problem des vorläufigen Wissens in der historischen Sprachwissenschaft. Wenn wir von Schmidts ursprünglichen Zahlen ausgehen, sieht es tatsächlich komisch aus, wie sich die Wurzeln zwischen den drei Sprachpaaren verteilen. Schaut man sich aber moderne Statistiken, beruhend auf Nicholaev (2007)

an, dann sieht das Ganze plötzlich ganz anders aus, wie die folgende Grafik zeigt (Geisler and List 2013):



Auch wenn sich die Mehrheitsverhaltnisse in den geteilten Wurzeln verschoben haben, so sind sie aber nach wie vor ahnlich in ihrer Tendenz. Gleichzeitig ist bereits die ursprngliche Beurteilung Schmidts, dass die Daten zu den drei Sprachen keinen Baum ergben, mathematisch falsch. Wie msste der Baum zwischen dem Lateinischen, dem Griechischen, und dem Indischen aussehen, wenn er genau diese hnlichkeiten zwischen den Sprachen darstellen sollte?

4 Netze

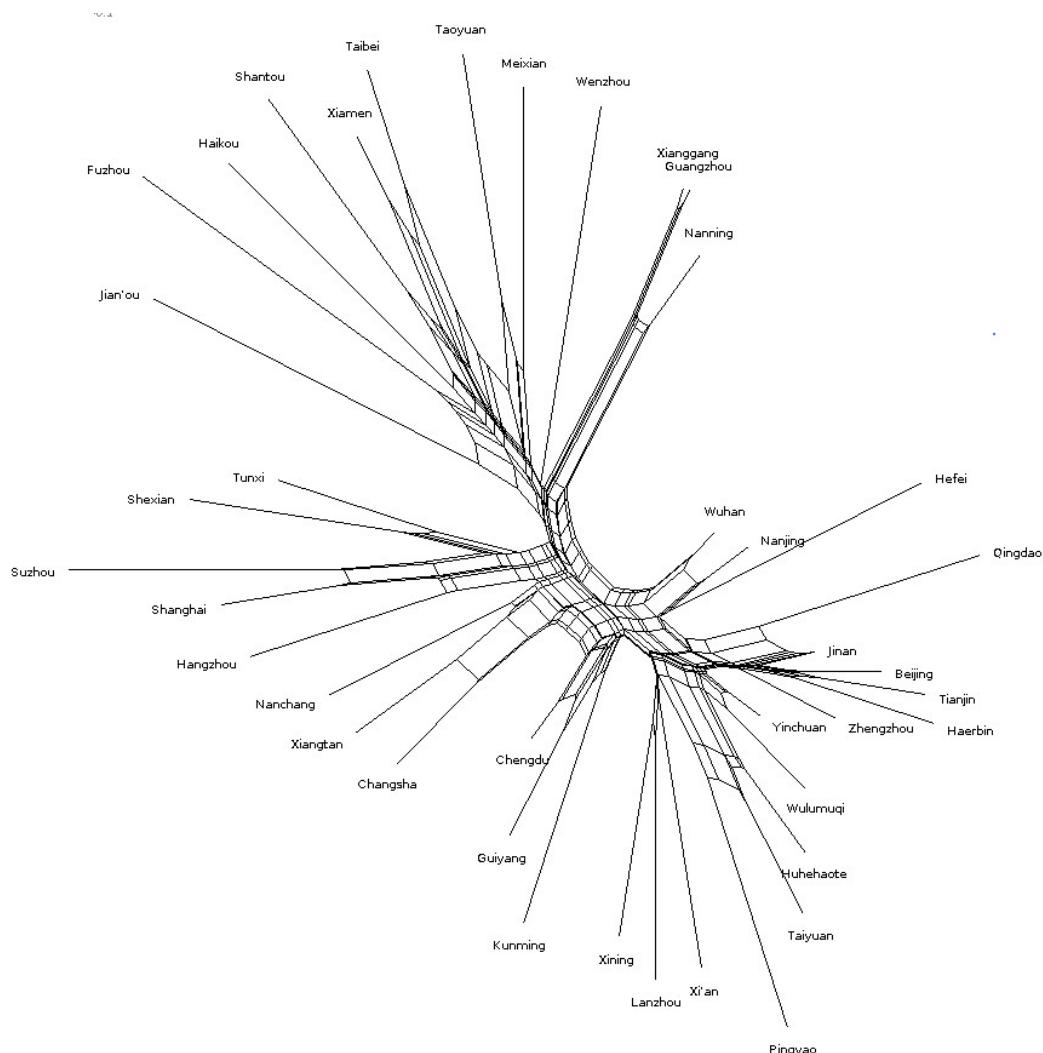
4.1 Historisches

Die wirkliche Idee der Netzwerke ist noch recht jung, sie deutet sich aber schon bei an, der in einer Vorlesung im Jahre 1870 bemerkte:

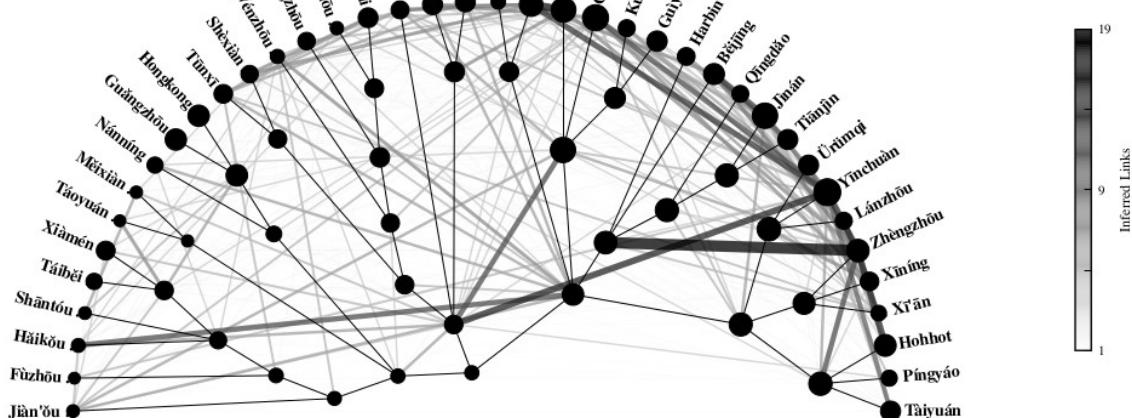
Wir verbinden die Äste und Zweige des Stammbaums durch zahllose horizontale Linien, und er hört auf ein Stammbaum zu sein. (Schuchardt 1870[1900])

Was an dieser Aussage wichtig ist ist ihre explizite Unterscheidung von *vertikalen* und *horizontalen* Beziehungen. Mit horizontalen Linien meint Schuchardt in diesem Zusammenhang ganz klar die Beziehungen zwischen Sprachen, die auftreten, wenn diese in Kontakt miteinander treten und die vom Stammbaum traditionell nicht dargestellt werden können. Er macht also implizit eine Unterscheidung zwischen vertikalen und horizontalen Beziehungen zwischen Sprachen.

Unter Netzwerkansätzen verstehe ich in diesem Zusammenhang sogenannte *evolutionäre Netzwerke*, also Netzwerke, die eine explizite historische Aussage enthalten, eine Wurzel aufweisen, und Vorgängersprachen beinhalten (Morrison 2011). Evolutionäre Netzwerke müssen ganz klar unterschieden werden von *data-display*-Netzwerken, die lediglich der Visualisierung von Konflikten in Daten dienen und bspw. zeigen können, wie „baumartig“ ein bestimmter Datensatz ist (*ibid.*). Die Unterschiede werden deutlich, wenn wir uns Beispiele für *data-display* und *evolutionäre Netzwerke* anschauen.



Data-display Netzwerk von 40 Chinesischen Dialekten



(Mehr oder weniger) evolutionäres Netzwerk derselben 40 Dialekte

Worin liegen die prägnantesten Unterschiede zwischen den beiden Netzwerken?

4.2 Formales

Formal handelt es sich bei evolutionären Netzwerken um Bäume, die um zusätzliche und spezifisch gekennzeichnete Zweige erweitert werden. Basierend auf einem Baum, der die vertikale Entwicklung einer Sprachfamilie anzeigt, werden also die „horizontalen“ Beziehungen der Sprachen durch eine weitere Reihe von Kanten dargestellt.

Der Baum ignoriert horizontale Sprachbeziehungen, evolutionäre fügen sie explizit wieder hinzu. Was aber machen die Wellenmodelle? Sind evolutionäre Netzwerke eine Kombination von Welle und Baum? Und wenn nicht, worin besteht dann der mögliche Mehrwert wellentheoretischer Überlegungen?

4.3 Probleme

Das größte Problem evolutionärer Netzwerke ist die Tatsache, dass der Faktor der Unsicherheit nicht ausgemerzt werden kann. In vielen Fällen ist eben leider nicht vollständig klar, ob Beziehungen zwischen Sprachen die Folge von Entlehnung oder gemeinsamem Ursprung sind.

Gibt es vielleicht weitere Alternativen zu den hier besprochenen Formen der Sprachverwandtschaft? Wie verhält sich der Text von Anttila (1972) zu diesen Fragestellungen?

Literatur

- Anttila, R. (1972). *An introduction to historical and comparative linguistics*. New York: Macmillan.
- Bloomfield, L. (1933 [1973]). *Language*. London: Allen & Unwin.
- Bonfante, G. (1931). "I dialetti indo-europei". In: *Annali del R. Istituto Orientale di Napoli* 4, 69–185.
- Cayley, A. (1857). "On the theory of the analytical forms called trees". In: *Philosophical Magazine* 4 (13), 172–176.
- Geisler, H. and J.-M. List (2013). "Do languages grow on trees? The tree metaphor in the history of linguistics". In: *Classification and evolution in biology, linguistics and the history of science. Concepts – methods – visualization*. Ed. by H. Fänger, H. Geisler, T. Halling, and W. Martin. Stuttgart: Franz Steiner Verlag, 111–124.
- Hirt, H. (1905). *Die Indogermanen. Ihre Verbreitung, ihre Urheimat und ihre Kultur*. Vol. 1. Strassburg: Trübner. Internet Archive: dieindogermanen01hirtuoft.
- Meillet, A. (1922 [1908]). *Les dialectes Indo-Européens*. Paris: Librairie Ancienne Honoré Champion. Internet Archive: lesdialectesindo00meil.
- Morrison, D. A. (2011). *An introduction to phylogenetic networks*. Uppsala: RJR Productions.
- Nicholaev, S. (2007). *Indo-European etymology*. Available online: <http://starling.rinet.ru/>.
- Pisani, V. (1933). *Studi sulla preispreis delle lingue indo-europee*. Rome: Giovanni Bardi.
- Schleicher, A. (1853a). "Die ersten Spaltungen des indogermanischen Urvolkes". In: *Allgemeine Monatsschrift für Wissenschaft und Literatur* 3, 786–787.
- (1853b). "O jazyku litevském, zvláště na slovanský. Čteno v posezení sekcí filologické královské společnosti Nauk dne 6. června 1853". In: *Časopis Českého Muzeum* 27, 320–334.
- (1861). *Compendium der vergleichenden Grammatik der indogermanischen Sprache*. Vol. 1: *Kurzer Abriss einer Lautlehre der indogermanischen Ursprache, des Altindischen (Sanskrit), Alteriranischen (Altbaktrischen), Altgriechischen, Altitalischen (Lateinischen, Umbrischen, Oskischen), Altkeltischen (Altirischen), Altslawischen (Altbulgarischen), Litauischen und Altdeutschen (Gotischen)*. Weimar: Böhlau.
- Schmidt, J. (1872). *Die Verwantschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.
- (1875). *Zur Geschichte des indogermanischen Vocalismus*. Weimar: Hermann Böhlau.
- Schuchardt, H. (1870 [1900]). *Über die Klassifikation der romanischen Mundarten. Probe-Vorlesung, gehalten zu Leipzig am 30. April 1870*. Graz. URL: <http://schuchardt.uni-graz.at/cgi-bin/print.cgi?action=show&type=pdf&id=724>.

„Traue keiner Statistik...“

1 Kleinigkeiten vorweg...

1.1 Innere und äußere Sprachgeschichte

Georg von der Gabelentz (1840 – 1893) war ein genialer Sprachwissenschaftler, dessen Werk in der modernen Sprachforschung leider allzu oft ignoriert wird. Da 2013 nicht nur Grimm-Jahr, sondern auch Gabelentz-Jahr ist, rufen wir uns noch einmal einige seiner berühmtesten Sätze in Erinnerung.

Der Zweig der Sprachforschung, der uns hier beschäftigt, hat es zunächst mit den trockensten Einzelthatsachen zu thun: Sind die Sprachen *A* und *B* miteinander verwandt, und in welchem Grade? Giebt es dieses Wort oder jene Form in der und der Sprache oder in der und der Zeit der Sprachgeschichte? wie lautet es da? Welche Gesetzmässigkeit herrscht in den lautlichen Abweichungen? Besteht im einzelnen Falle Urgemeinschaft oder Entlehnung? Was ist alles Gemeingut, was neu hinzu erworben? u. s. w. Alles das klingt und is auch wirklich sehr trocken. Was die menschliche Rede im Innersten bewegt, was sonst die Wissenschaft von den Sprachen der Völker zu einer der lebensvollsten macht, das tritt hier zunächst zurück: nur einige ihrer Ausläufer ranken in das Seelen- und Sittenleben der Völker hinüber. Der einzelsprachliche Forscher kann gar nicht schnell genug die fremde Sprache in's eigene Ich aufnehmen: der Sprachhistoriker steht draussen vor seinem Gegenstande: hier der Anatom, da der Cadaver. (Gabelentz 1891: 145)

Wir werden, um Missverständnisse zu vermeiden, gut thun, zwischen äusserer und innerer Sprachgeschichte zu unterscheiden. Die äussere Geschichte einer Sprache ist die Geschichte ihrer räumlichen und zeitlichen Verbreitung, ihrer Verzweigungen und etwaigen Mischungen (Genealogie). Die innere Sprachgeschichte erzählt und sucht zu erklären, wie sich die Sprache in Rücksicht auf Stoff und Form allmählich verändert hat. (ibid.: 146)

In der Biologie unterscheidet man zwischen Phylogenie und Ontogenese, wobei Phylogenie die (ausschnittsweise) Geschichte aller Arten bezeichnet, und Ontogenese die Geschichte einzelner Organismen. Lassen sich diese Termini mit der inneren und der äusseren Sprachgeschichte Gabelentz' vergleichen, und wenn nicht, welche grundlegenden Probleme und Unterschiede lassen sich festhalten?

1.2 Darstellung

Wir müssen unterscheiden zwischen Schemata (seien es Wellen oder Bäume) zur Darstellung von **Daten**, und Schemata zur Darstellung von **Geschichte** (vgl. die Unterscheidung zwischen *data-display* und *evolutionary* in Morrison 2011: 42f). Schleichers Baum von 1861 ist dabei ein klares Beispiel für ein Schema, das den Anspruch hat, ein geschichtliches Schema zu sein, während die Beispiele für die Visualisierung der Wellentheorie wohl eher als Datendarstellungsschemata bezeichnet werden sollten, da sie nicht für sich in Anspruch nehmen, äussere Sprachgeschichte zu modellieren. Schemata zur Darstellung von Daten können unter Umständen in Schemata zur Darstellung von Geschichte überführt werden, jedoch hängt die Überführbarkeit davon ab, ob die Daten die Rekonstruktion von Geschichte auch erlauben. Die Frage ist also, wenn man Sprachen miteinander vergleicht, welche Unterschiede zwischen Sprachen tatsächlich eine Rekonstruktion

der äusseren Sprachgeschichte erlauben, wie schon Karl Brugmann (1849 – 1919) in seinen Ausführungen zur Problematik der genealogischen Klassifikation der acht großen indogermanischen Sprachgruppen deutlich machte:

Im ganzen ist also nur wenig, was aus den spezielleren Übereinstimmungen zwischen einzelnen von den acht Hauptgruppen für die Beziehungen der Völker zu einander in sogen. voreinzelsprachlicher Zeit mit grösserer Wahrscheinlichkeit entnommen werden kann. Und jedenfalls treten, so viel wir heute wissen, nirgends speziellere Gemeinsamkeiten, die als gemeinsame Neuerungen erscheinen, in *so grosser Anzahl* entgegen, dass man auf Grund derselben die betreffenden Sprachzweige in derselben Art zu Einheiten zusammenschliessen dürfte [...]. Dies gilt selbst für den Fall, dass man keine von diesen Übereinstimmungen als nur zufällig und keine als auf Entlehnung beruhend betrachten wollte. (Brugmann 1904[1970]: 21f)

Brugmann äußert sich vordergründig über das Problem der genealogischen Klassifikation der indogermanischen Sprachen. Dabei nennt er jedoch auch einige sehr wichtige Prinzipien, die bei der phylogenetischen Rekonstruktion (= genealogische Klassifikation) beachtet werden müssen. Welche sind dies?

2 Lexikostatistik

2.1 Hintergrund

Die Lexikostatistik stellt ein statistisch basiertes Verfahren zur Ermittlung von Verwandtschaftsbeziehungen zwischen Sprachen (und damit zur phylogenetischen Rekonstruktion) dar. Sie wurde von Morris Swadesh (1909 – 1967) in einer Reihe von Artikeln zu Beginn der 50er Jahre des 20. Jahrhunderts vorgestellt und weiterentwickelt (Swadesh 1950, Swadesh 1952, Swadesh 1955). In der Folgezeit mehrte sich jedoch die Kritik an der Methode (Bergsland and Vogt 1962, Hoijer 1956, Rea 1973) und kam am Ende aus der Mode. Grundsätzlich werden im Rahmen der Lexikostatistik historisch relevante Gemeinsamkeiten zwischen Sprachen ausgezählt. Die zugrunde liegenden Zahlen können dann weiterverwendet werden, um genealogische Bäume automatisch zu rekonstruieren, oder um (unter der Annahme konstanten Wandels) Sprachspaltungszeitpunkte zu ermitteln. Die Methode erlebte zu Beginn des 21. Jahrhunderts eine Wiedergeburt im Rahmen neuer quantitativer biologischer Ansätze, mit deren Hilfe genealogische Bäume automatisch aus spezifischen Sprachdaten gewonnen werden können (Atkinson and Gray 2006, Gray and Atkinson 2003).

Um Spekulation wird gebeten: Aus welchen Gründen wurde die Lexikostatistik so schnell wieder verworfen? Aus welchen Gründen wurde sie so euphorisch wiedereingeführt?

2.2 Grundannahmen

Die Grundannahmen der Lexikostatistik wurden in einer Vielzahl von Arbeiten besprochen (Gudschinsky 1956, Sankoff 1969). Basierend auf diesen Arbeiten lassen sie sich in etwa wie folgt zusammenfassen (vgl. Geisler and List im Druck):

1. The lexicon of every human language contains words which are relatively resistant to borrowing and relatively stable over time due to the meaning they express: these words constitute the *basic vocabulary* of languages.

2. *Shared retentions* in the basic vocabulary of different languages reflect their *degree of genetic relationship*, i.e. they are representative for the reconstruction of language phylogenies.

Zwei kurze Fragen: (1) Welche Wörter gehören wohl zum Basisvokabular? (2) Worin besteht der Unterschied zwischen Punkt 2 und den Punkten, die Brugmann (siehe oben) angesprochen hatte?

2.3 Praktische Umsetzung

Was theoretisch sehr abstrakt und wohlüberlegt klingen mag, zeigt sich in der Praxis als ein recht einfaches Verfahren, das wohl am besten in fünf Schritte unterteilt werden kann (vgl. Geisler and List im Druck):

1. Compilation: Compile a list of basic vocabulary items (a Swadesh-list).
2. Translation: Translate the items into the languages that shall be investigated.
3. Cognate Judgments: Search the language entries for cognates.
4. Coding: Convert the cognate information into a numerical format.
5. Computation: Perform a computational analysis (cluster analysis, tree calculation) of the numerical data, which allows to make conclusions regarding the phylogeny of the languages under investigation.

Es gibt unterschiedliche Swadesh-Listen, auf denen lexikostatistische Analysen beruhen. Die folgende Tabelle zeigt eine sehr kleine Liste von 35 Items, die von Sergej Jachontov vorgeschlagen wurde (Burlak and Starostin 2005: 13):

1	BLOOD	2	BONE	3	DIE	4	DOG	5	EAR
6	EGG	7	EYE	8	FIRE	9	FISH	10	FULL
11	GIVE	12	HAND	13	HORN	14	I	15	KNOW
16	LOUSE	17	MOON	18	NAME	19	NEW	20	NOSE
21	ONE	22	STONE	23	SUN	24	TAIL	25	THIS
26	THOU	27	TONGUE	28	TOOTH	29	TOOTH	30	WATER
31	WHAT	32	WHO	33	SALT	34	WIND	35	YEAR

Kurzes Hirnstürmen: In welchen Arbeitsschritten kann man wohl bei einem solchen Verfahren die größten Probleme erwarten? Welche Probleme können aus Konzeptlisten wie der von Jachontov entstehen?

2.4 Kritik

Auf grundlegende Probleme wurde bereits kurz nach dem Erscheinen von Swadeshs ersten Veröffentlichungen zur Lexikostatistik hingewiesen. Diese betrafen Punkte, die auch aus den traditionellen Methoden zur genealogischen Klassifikation von Sprachen gut bekannt sind. Die wichtigsten Punkte sind dabei:

- **Entlehnung:** Unentdeckte Entlehnungen können die Ergebnisse verfälschen.
- **Aussagekraft:** Lexikalische Ersetzung ist als Prozess nicht aussagekräftig für Sprachgeschichte.

- **Fehleranfälligkeit:** Die Methode ist fehleranfällig, da die Daten auf eine problematische Weise erstellt werden.

Während die ersten beiden Punkte in der Literatur von oben bis untern durchgekaut wurden, ohne dass sie den Befürwortern der Lexikostatistik tatsächlich argumentatorisch entgegentreten konnten, wurde der letzte Punkt nur wenig beachtet, obwohl er das größte Problem der Lexikostatistik und ihrer Nachfolgermethoden anspricht: Das Problem der Subjektivität der Daten, die eine lexikostatistische Analyse produziert. Neuere Vergleiche haben dabei zeigen können, dass die Daten, die von Forscherteams unabhängig produziert werden, derartig große Unterschiede aufweisen, dass dies zu Unterschieden von über 30% in von den Daten automatisch berechneten Baumtopologien führt (**Geisler2012** im Druck). Die größten Probleme liegen dabei weniger im Bereich der Kognatenzuweisungen (Schritt 3, obwohl auch dieser problematisch ist), sondern bereits im Bereich der Übersetzung (Schritt 2). Denn bereits hier zeigen sich große Unterschiede zwischen unabhängig erstellten Datensätzen, die zeigen, dass mangelnde Kompetenz der Forscher in den Einzelsprachen, aber auch mangelnde Beschreibung der Konzepte in den Konzeptlisten, zu einer Vielzahl von Unterschieden bereits in den Ausgangsdaten führen können.

Wie lassen sich die vielen Unterschiede in den Übersetzungen (Schritt 2) erklären?

3 Computergestützte phylogenetische Rekonstruktion

Ziel der Rekonstruktion evolutionärer Entwicklungsszenarios (Stammbäume, Netze) ist es, aus einer bestimmten Anzahl von Daten evolutionäre Klassifikationssysteme abzuleiten. Unter evolutionären Klassifikationssystemen verstehe ich grundsätzlich genealogische Klassifikationssysteme, also Klassifikationssysteme, die nicht nach oberflächlicher Ähnlichkeit klassifizieren, sondern den Anspruch haben, darzustellen, wie die Daten im Laufe der Geschichte entstanden sind, wie sie sich entwickelt haben. Da dabei meist mit großen Datenmengen gearbeitet wird, kommt computerbasierten Verfahren in der Stammbaum- und Netzwerkrekonstruktion eine große Bedeutung zu.

3.1 Ermittlung der Komparanda

Für die automatische Rekonstruktion von Stammbäumen und Netzwerken ist es zunächst entscheidend, zu bestimmen, was verglichen werden soll. Die Komparanda müssen dabei in ein für die Computeranalyse adäquates Format überführt werden, das in den meisten Fällen numerisch ist. Grundsätzlich unterscheidet man dabei zwei verschiedene Ansätze:

Distanzbasierte Ansätze Aus den Daten werden Distanzwerte zwischen den zu vergleichenden Taxa (in unserem Falle Sprachen) abgeleitet. Distanzwerte sind dabei beliebige Zahlen zwischen 0 (Identität der Taxa) und ∞ (Nichtidentität der Taxa), wobei Taxa, die weiter voneinander entfernt sind, einen größeren Distanzwert zugewiesen bekommen, als Taxa, die einander näher liegen.

Charakterbasierte Ansätze Die Daten werden nicht in Distanzmatrizen transformiert. Stattdessen wird jedes Taxon in Bezug auf eine Reihe von Eigenschaften charakterisiert. Für jede dieser Eigenschaften kann ein Taxon dabei verschiedene Zustände

aufweisen. Die einfachste Zustandsbeschreibung ist dabei die Anwesenheit oder Abwesenheit der jeweiligen Eigenschaft (dargestellt in einer Binärmatrix). Komplexere Zustände können den Taxa jedoch ebenfalls zugewiesen werden

	L_1	L_2	L_3
L_1	0	10	20
L_2	10	0	30
L_3	20	30	0

(a) Distanzmatrix

Character	C_1	C2	C_3	C_4	C_5
L_1	0	1	1	1	0
L_2	1	1	0	1	0
L_3	0	0	0	1	1

(b) Merkmalsmatrix

Was haben diese Matrizen mit evolutionären Bäumen zu tun?

3.2 Computerbasierte Stammbaumrekonstruktion

Um aus den in Distanzform oder Charakterform kodierten Daten mit Hilfe des Computers einen Stammbaum abzuleiten, müssen diese geclustert werden. Hierzu sind in der Biologie verschiedene Verfahren entwickelt worden. Es gibt eine Vielzahl verschiedenster Softwareprogramme, welche diese Aufgabe erfüllen. Den einzelnen Anwendungen innerhalb dieser Software liegen dabei verschiedene Algorithmen zugrunde, welche aus den Daten sukzessive einen Baum erzeugen, oder aus einer Vielzahl möglicher Bäume den wahrscheinlichsten Baum ermitteln. Zu nennen sind in diesem Zusammenhang:

Neighbor-Joining Clusterverfahren für Distanzdaten, entwickelt von Saitou and Nei (1987), welches sich insbesondere deshalb großer Beliebtheit erfreut, weil es eine sehr geringe Laufzeit hat. Um einen schnellen Überblick über die Daten zu bekommen, ist es daher sehr gut geeignet.

Bayesianische Analysen Verfahren für Charakterdaten, das auf bayesianischer Heuristik beruht und innerhalb der möglichen Bäume nach den wahrscheinlichsten sucht. Dieses Verfahren wird insbesondere von dem Team um Russel Gray (vgl. Gray and Atkinson 2003) verwendet.

Maximale Parsimonie Charakterbasiertes Verfahren, welches den evolutionären Baum errechnet, der am wenigsten evolutionären Wandel benötigt, um die Daten zu erklären.

Parsimonie-basierte Verfahren werden oft als »inhärent linguistisch« bezeichnet, weil sie die Anzahl von Wandelereignissen reduzieren. Welches Prinzip steckt hinter der Grundannahme, dass reduzieren des Wandels zu einer korrekten Rekonstruktion der historischen Ereignisse führt?

3.3 Daten

Als Eingabedaten werden in den phylogenetischen Analysen im Rahmen der historischen Linguistik zumeist Swadesh-Listen (vgl. Swadesh 1950, 1952, 1955) verwendet. Jedes Item ("Bedeutungsslot") wird dabei zunächst als Charakter angesehen. Diesem wird für jede Sprache ein Zustand zugewiesen, wobei der Zustand als identisch kodiert wird, wenn die jeweiligen Spracheinträge kognat sind.

Um diese Daten in ein Distanzformat zu überführen, genügt es, die Anzahl der kognaten Einträge zwischen zwei Sprachen auszuzählen, ihren prozentualen

Anteil zu berechnen und anschließend durch die Formel $1 - \text{Prozentzahl} / 100$ umzuwandeln. Für das Beispiel in Tabelle ?? ergibt sich somit: $1 - 40 / 100 = 0.6$, da von 10 Items 4 Spracheinträge (40%) kognat sind.

Number	Word	German	State	English	State
1	all	all	1	all	1
2	ashes	Asche	1	ashes	1
3	bark	Rinde	1	bark	2
4	belly	Bauch	1	belly	2
5	big	groß	1	big	2
6	bird	Vogel	1	bird	2
7	bite	beissen	1	bite	1
8	black	schwarz	1	black	2
9	blood	blut	1	blood	1
10	bone	Knochen	1	bone	2

Bei der Überführung in ein Charakterformat wird meist eine Binarisierung vorgenommen, weil die meisten Programme nur binäre Zustände akzeptieren. Hierzu wird jeder Sprachzustand, der pro Item vorkommt, als gesonderter Charakter (Kognatenset) gewertet und anschließend die Anwesenheit bzw. Abwesenheit durch Einsetzen von 1 bzw. 0 im jeweiligen Charakterfeld ausgedrückt.

Number	Charakter	German	English
1	*ala-	1	1
2	*askō-	1	1
3	*randō	1	0
3	*birkō	0	1
4	*puk-	1	0
4	*balgi-	0	1
5	*griuna-	1	0
5	*puk-	0	1
6	*fugla-	1	0
6	*biran-	0	1
...

Mit der Binarisierung von Kognatensets ist ein ganz spezifisches Problem verbunden, das in der Lexikostatistik in ihrer Reinform nicht auftritt, welches?

3.4 Formate

Drei gängige Formate sollen in diesem Zusammenhang genannt werden:

Nexus-Format Ein sehr allgemeines Format, das als Eingabeformat in einer Vielzahl von Softwarepaketen Verwendung findet. Möglich sind Charakterdaten, Distanzdaten, Netzwerkdarstellungen und Baumdaten.

Phylip Distance Format Ein sehr gängiges Distanzformat, das in der Phylip-Software (Felsenstein 2005) Verwendung findet, aber auch von anderen Softwarepaketen eingelesen werden kann.

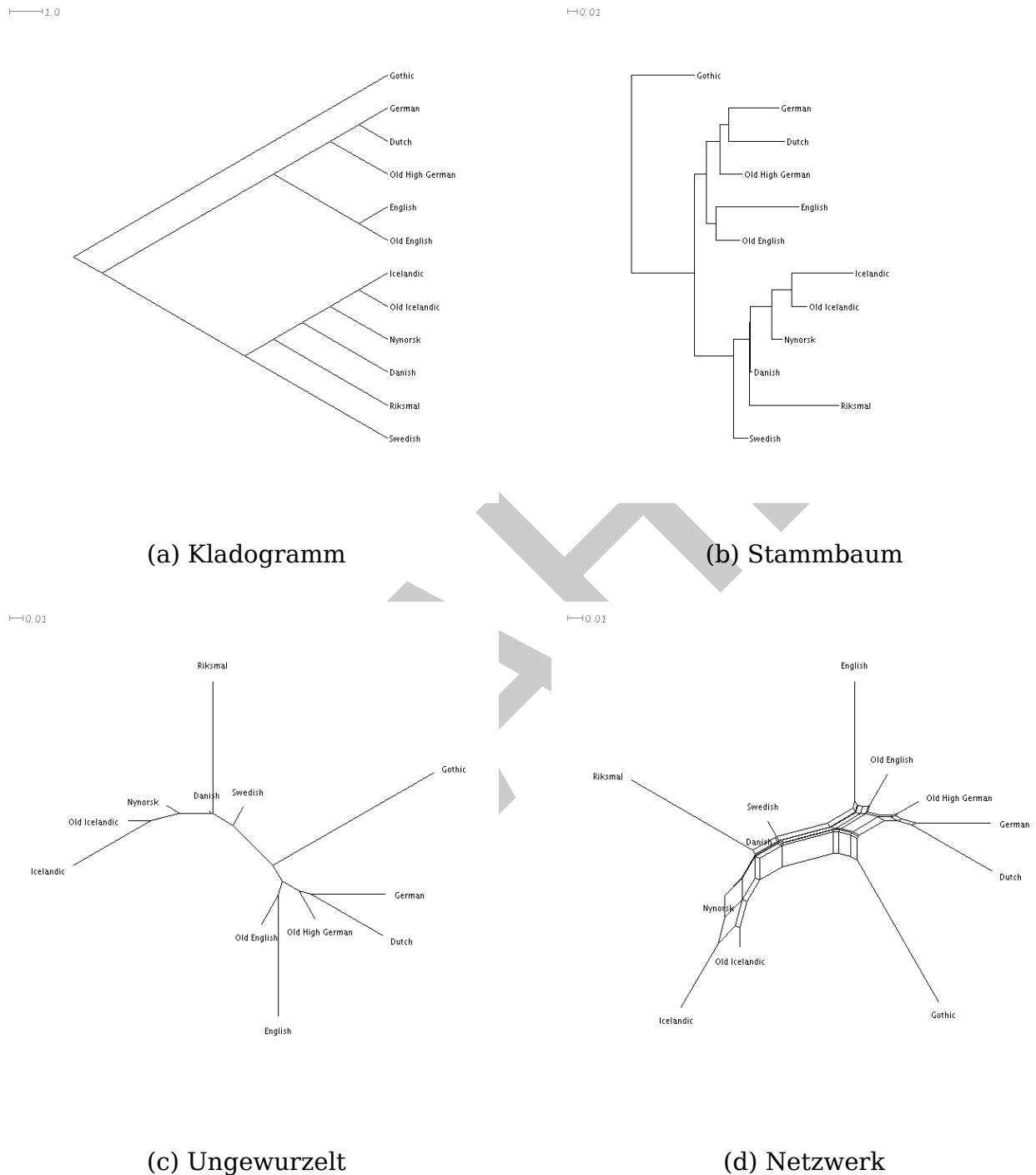
Beispiel für das Nexusformat

```
#nexus
BEGIN Taxa;
DIMENSIONS ntax=12;
TAXLABELS 'Germa' 'Engli' 'Dutch' 'Icela' 'Nynor' 'Riksm' 'Swedi' 'Danis'
```

Beispiel für das Phytip-Format

Wenn man sich das Nexus-Format anschaut, kann man daraus irgendwelche grundlegenden Eigenschaften von Computerformaten ableiten?

3.5 Graphische Darstellung



Worin bestehen die jeweiligen grundlegenden Unterschiede zwischen den Darstellungsformaten? Kann man die Formater hinsichtlich ihres »Realismus« bewerten?

4 Übungsaufgabe: Mikronesisch (V. I. Belikov)

Die folgende Aufgabe wurde entnommen und übersetzt aus Burlak and Starostin (2005: 404f).

4.1 Ausgangspunkt

Die karolinischen Inseln gehören zu Mikronesien. Zu diesen gehören, u. a., die Inseln Voleai, Puluvat, Satava, Sonsorol, Truk und Uliti. Diese Gruppe von Inseln bewohnen die Nachfolger eines ehemals einheitlichen Volkes, welche einander nah verwandte Sprachen sprechen. In diesen Sprachen gibt es selbstverständlich viele gemeinsame Wörter. Unten werden die Prozente der gemeinsamen Wörter für einige Sprachen aufgezeigt, die von den Bewohnern der sechs genannten Inseln gesprochen werden.

	Uliti	Truk	Sonsorol	Sataval	Puluvat
Voleai	70-80	60-70	50-60	70-80	60-70
Puluvat	60-70	70-80	50-60	70-80	
Sataval	60-70	70-80	50-60		
Sonsorol	60-70	50-60			
Truk	50-60				

Gegeben ist ferner ein Fragment einer Karte dieses Gebietes. Auf ihm sind die Namen der Inseln Voleai, Puluvat, Sonsorol und Truk durch Ziffern von 1 bis 4 ersetzt worden. Außerdem sind die Namen dreier weiterer Inseln (darunter auch die Insel Sataval) mit Lateinischen Buchstaben A,B und C ersetzt worden.



4.2 Aufgaben

- Finden Sie heraus, welche Inseln sich hinter den Ziffern von 1 bis 4 verbergen.
- Finden sie heraus, mit welchem lateinischen Buchstaben die Insel Sataval kodiert wurde.

4.3 Zusatzaufgabe

- Versuchen Sie, die Entwicklung der Sprachen mit Hilfe eines Baumes darzustellen, der möglichst optimal die attestierten Ähnlichkeiten erklärt.

Literatur

Atkinson, Q. D. and R. D. Gray (2006). "How old is the Indo-European language family? Illumination or more moths to the flame?" In: *Phylogenetic methods and the prehistory of languages*. Ed. by P. Forster and C. Renfrew. McDonald Institute monographs. Cambridge, Oxford, and Oakville: McDonald Institute for Archaeological Research, 91–109.

- Bergsland, K. and H. Vogt (1962). "On the Validity of Glottochronology". In: *Current Anthropology* 3.2, 115–153. JSTOR: 2739527.
- Brugmann, K. (1904[1970]). *Kurze vergleichende Grammatik der indogermanischen Sprachen. Auf Grund des fünfbandigen 'Grundrisses der vergleichenden Grammatik der indogermanischen Sprachen von K. Brugmann und B. Delbrück' verfasst*. Photomechanischer Nachdruck 1970. Walter de Gruyter & Co., Berlin. Strassburg: Karl J. Trübner.
- Burlak, S. A. and S. A. Starostin (2005). *Sravnitel'no-istoričeskoe jazykoznanie* [Comparative-historical linguistics]. Moscow: Akademia.
- Felsenstein, J. (2005). *PHYLIP (Phylogeny Inference Package) version 3.6*. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- Gabelentz, H. G. C. (1891). *Die Sprachwissenschaft. Ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: T. O. Weigel. Internet Archive: diessprachwissen00gabegoog.
- Geisler, H. and J.-M. List. "Beautiful trees on unstable ground. Notes on the data problem in lexicostatistics". In: *Die Ausbreitung des Indogermanischen. Thesen aus Sprachwissenschaft, Archäologie und Genetik*. Ed. by H. Hettrich. Wiesbaden: Reichert.
- Gray, R. D. and Q. D. Atkinson (2003). "Language-tree divergence times support the Anatolian theory of Indo-European origin". In: *Nature* 426.6965, 435–439.
- Gudschinsky, S. C. (1956). "Three Disturbing Questions concerning Lexicostatistics". In: *International Journal of American Linguistics* 22.3, 212–213. JSTOR: 1264017.
- Hoijer, H. (1956). "Lexicostatistics. A critique". In: *Language* 32.1, 49–60. JSTOR: 410652.
- Morrison, D. A. (2011). *An introduction to phylogenetic networks*. Uppsala: RJR Productions.
- Rea, J. A. (1973). "The Romance data of pilot studies for glottochronology". In: *Diachronic, areal and typological linguistics*. Ed. by H. M. Hoenigswald and R. H. Langacre. Vol. 11. Current Trends in Linguistics. The Hague; Paris: Mouton, 355–367.
- Saitou, N. and M. Nei (1987). "The neighbor-joining method: A new method for reconstructing phylogenetic trees". In: *Molecular Biology and Evolution* 4.4, 406–425.
- Sankoff, D. (1969). "Historical linguistics as stochastic process". (A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy). PhD thesis. Montreal: McGill University.
- Swadesh, M. (1950). "Salish internal relationships". In: *International Journal of American Linguistics* 16.4, 157–167. JSTOR: 1262898.
- (1952). "Lexico-statistic dating of prehistoric ethnic contacts. With special reference to North American Indians and Eskimos". In: *Proceedings of the American Philosophical Society* 96.4, 452–463. JSTOR: 3143802.
- (1955). "Towards greater accuracy in lexicostatistic dating". In: *International Journal of American Linguistics* 21.2, 121–137. JSTOR: 1263939.

„Vaterschaftstest für Sprachen...“

1 Allgemeines

1.1 Das Grundproblem

Historische Linguisten würden am liebsten den ganzen Tag Sprachen vergleichen und Ursprachen rekonstruieren. Das Problem dabei ist aber, dass man als historischer Linguist nicht wild drauf los vergleichen darf, sondern nur solche Sprachen vergleichen sollte, von denen man weiß, dass sie miteinander verwandt sind. Daher stellt sich für jeden historischen Linguisten das Problem, nachweisen zu müssen, dass zwei oder mehr Sprachen tatsächlich miteinander verwandt sind.

Dieses Grundproblem des Nachweises von Sprachverwandtschaft wird in der historischen Linguistik auf sehr unterschiedliche Weise angegangen. Nach wie vor herrscht keine Einigkeit darüber, wie der Verwandtschaftsbeweis tatsächlich erbracht werden sollte. Die Diskussion über die Unterschiedlichen Indizien, die als Beweismittel angeführt werden, nimmt zuweilen Formen von Grabenkämpfen an.

Warum kann ein historischer Linguist nicht einfach drauf los vergleichen? Warum muss überhaupt erst nachgewiesen worden sein, dass die Sprachen, die verglichen werden, auf einen gemeinsamen Vorgänger zurückgehen?

1.2 Historisches

Der im Zusammenhang mit Sprachverwandtschaftsnachweisen meistzitierte Forscher ist wohl Sir William Jones (1746 – 1794), der in Indien tätig war, wo er mit dem Sanskrit in Berührung kam und auf dessen überraschende Ähnlichkeit mit dem Griechischen und Lateinischen aufmerksam wurde. In dem *Third anniversary discourse, on the Hindus* aus dem Jahre 1786 liest sich dies wie folgt:

The Sanscrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists. (Jones 1798: 422f)

Obwohl Jones oft als „Entdecker des Indogermanischen“ gepriesen wird, war er keineswegs der erste, dem die Ähnlichkeiten zwischen Sanskrit und den europäischen Sprachen auffiel. Der Italiener Filippo Sassetti (1540 – 1588) zum Beispiel hatte bereits 200 Jahre vor Jones darauf hingewiesen, dass es auffällige Ähnlichkeiten zwischen dem Sanskrit und dem Italienischen gebe:

Alles Wissenschaftliche ist in einer Sprache verfasst, die sie „Sanscrita“ nennen, was soviel heißt wie „gut artikuliert“. Unsere Sprache hat viele Gemeinsamkeiten mit dieser, darunter viele unserer Wörter, insbesondere die Zahlen 6, 7, 8, und 9, „Gott“, „Schlange“, und viele andere. (Sassetti 1855: 415)¹

¹Meine Übersetzung, Originaltext: „Sono scritte le loro scienze tutte in una lingua, che dimandano Sanscrita, che vuol dire bene articolata. [...] et ha la lingua d' oggi molte cose comuni con quella, nella quale sono molti de' nostri nomi, e particolarmente de' numeri il 6, 7, 8 e 9, Dio, serpe, et altri assai.“

Woran Sassetti dabei dachte, waren Ähnlichkeiten zwischen Wörtern, wie sie in der folgenden Tabelle wiedergegeben sind:

Sanscruta Italienisch	<i>sarpá-</i> <i>serpe</i>	s s	a ɛ	r r	p p	a ə
Sanscruta Italienisch	<i>devá-</i> <i>Dio</i>	d d	e i	v -	a o	
Sanscruta Italienisch	<i>saptá-</i> <i>sette</i>	s s	a ɛ	p -	t t̪	a ə

Was genau fällt auf beim Betrachten der Wortpaare, die Sassetti als Beispiele anführte, und den Ausführungen von Jones? Worauf genau fußt das, was als "Beweis" der Verwandtschaft angeführt wird?

2 Probleme

2.1 Zirkularität

Ein grundlegendes und vieldiskutiertes Problem des Verwandtschaftsbeweises ist das Problem der **Zirkularität**. Georg von der Gabelentz (1840 – 1893) fasst dieses wie folgt zusammen:

Es leuchtet ein, dass man, solange man nur sprachgeschichtliche Zwecke verfolgt, nur genetisch verwandte Sprachen miteinander vergleichen darf. Und umgekehrt ist es einleuchtend, dass der Beweis der Verwandtschaft, wo er nötig ist, nur im Wege der Vergleichung geführt werden kann. So scheint es, als drehten wir uns im Kreise. In der That ist aber die vergleichende Arbeit, die nur die Familienzugehörigkeit erweisen will, summarisch im Gegensatze zu jenen minutiösen Untersuchungen, die die innere Sprachgeschichte erheischt. Zudem ist jene Arbeit die vorbereitende, und schon darum muss sie zuerst betrachtet werden. (Gabelentz 1891: 150)

Moment mal: Zirkularität ist uns doch auch schon früher begegnet, aber wo nur? Und was genau schlägt Gabelentz jetzt eigentlich vor, um dieses Problem zu umgehen?

2.2 Comparanda

Neben dem Problem der Zirkularität stellt insbesondere auch das Problem der Wahl der Comparanda eines der Hauptprobleme beim Nachweis von Sprachverwandtschaft dar. Gabelentz lässt uns diesbezüglich wissen:

Dass der Wolf zum Hundegeschlechte gehört, lehrt uns ein einziger Blick. Dass aber die Blindschleiche nicht eine Schlange, sondern eine Eidechsenart ist, erfahren wir erst, wenn wir dem Thiere die Haut abstreifen und es anatomisch untersuchen. Beiderlei kommt auch in der Sprachenwelt vor, nur dass hier noch viel öfter die Verwandtschaftsmerkmale unter der Haut zu suchen sind. (ibid.: 151)

Alles schön und gut, Sprachverwandtschaft liegt unter der Haut, wie Gabelentz sagt. Aber was sind den jetzt bitteschön vergleichbare Comparanda in der Linguistik, und wie können wir Sprachen aufschneiden, um ihnen unter die Haut zu schauen?

2.3 Monogenese oder Polygenese

Ein großer Streitpunkt war im Laufe der Geschichte schon immer die Frage nach dem Ursprung der Sprache, die, um hipper zu klingen, heute zuweilen auch als die Frage nach der „Evolution von Sprache“ (*evolution of language*) gestellt wird. Viele Arbeiten, die die Wörter „Evolution“ und „Sprache“ im Titel haben, beschäftigen sich mit ebendieser Frage. Zuweilen gibt es Probleme aufgrund dieser Terminologie, da auch etwas hippere historische Linguisten anfangen, von „Sprachevolution“ zu reden, darunter jedoch eigentlich die innere oder äußere Sprachgeschichte, also die Frage der genealogischen Sprachklassifikation oder die Frage des Sprachwandels, verstehen. Die grundsätzliche Frage bezüglich des Sprachursprungs ist die, ob Sprache nur einmal, oder an mehrmals und unabhängig voneinander entstanden ist:

Woher noch immer die Menge der Sprachstämme? und woher die grosse Menge der Sprachen, die noch keinem bekannten Stämme zugeordnet sind? War es wirklich so, wie Manche glauben, dass an mehreren Orden der Erde, unabhängig von einander sich sprachlose Anthropoiden zu sprachbegabten Menschen entwickelt haben? [...] Wie aber, wenn jene Anderen Recht hätten, die da annehmen, die sprechende Menschheit, also auch die menschliche Sprache habe sich aus einer ursprünglichen Einheit differenziert? (Gabelentz 1891: 151)

Wie immer dem auch sein mag, ob also eine Monogenese oder eine Polygenese für den Sprachursprung zutreffen mag, Linguisten haben die Frage strenggenommen schon sehr früh dergestalt entschieden, dass sie fortan nicht wieder gestellt werden darf. Und zwar bereits im Jahre 1866, in dem die Statuten der Zeitschrift der linguistischen Gesellschaft von Paris genau dies zum Ausdruck brachten:

Die Gesellschaft lässt keine Art von Mitteilungen – weder bezüglich des Ursprungs von Sprache, noch bezüglich der Schaffung einer Universalsprache – zu. (Société de Linguistique de Paris 1871: „Statuts“, III)²

Schön und gut, aber was soll das Problem des Sprachursprungs jetzt eigentlich mit dem Problem des Nachweises von Sprachverwandtschaft zu tun haben? Sind das nicht zwei verschiedene Tassen Tee?

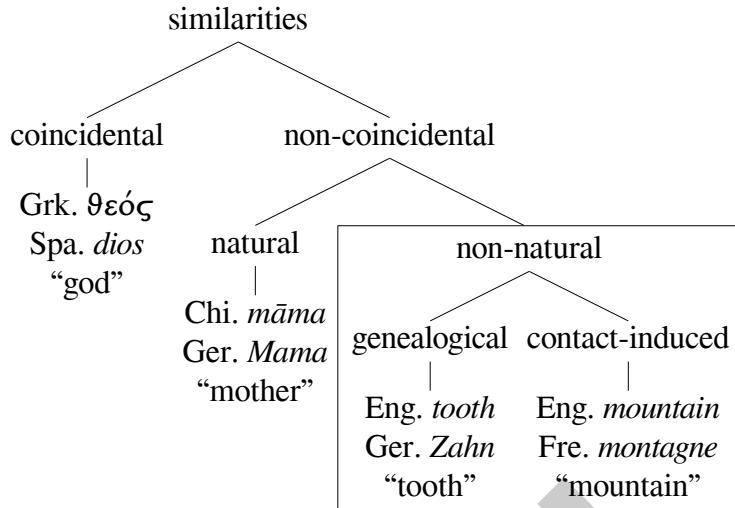
3 Lösungsansätze

3.1 Ähnlichkeiten

Genealogische und kontaktinduzierte Ähnlichkeiten

Das erste Teilproblem, das im Rahmen des Nachweises von Sprachverwandtschaft gelöst werden muss, ist das der Indizien. Grundlegend gilt hier, dass nur Ähnlichkeiten zwischen Sprachen als Indizien in Frage kommen können. Ferner gilt auch, dass nicht alle Ähnlichkeiten zwischen Sprachen aufgrund von Sprachverwandtschaft entstanden sind, wie wir rasch feststellen können, wenn wir uns noch mal den Entscheidungsbaum aus Sitzung 5 in Erinnerung rufen:

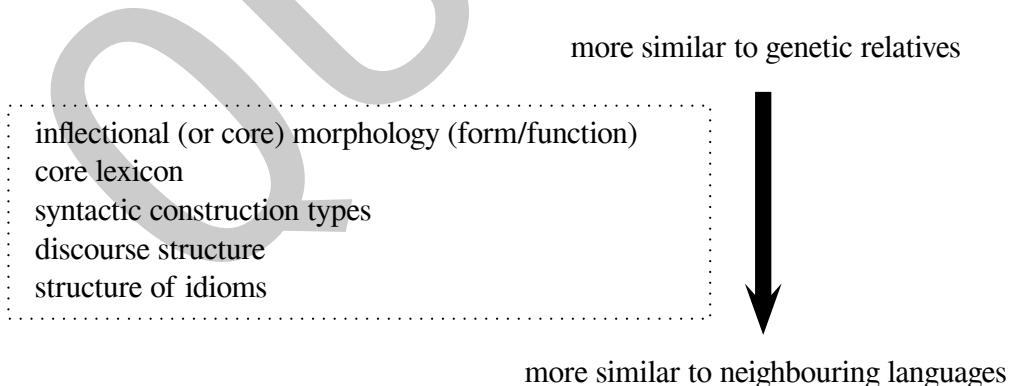
²Meine Übersetzung, Originaltext: „La Société n'admet aucune communication concernant, soit l'origine du langage, soit la création d'une langue universelle.“



Warum ist es so viel schwieriger, nicht-natürliche Ähnlichkeiten auseinanderzuhalten, als nicht-zufällige, oder zufällige?

Skala der Entlehnbarkeit

Das Problem des Nachweises von Sprachverwandtschaft ist, dass, während natürliche und zufällige Ähnlichkeiten relativ leicht von den nicht-natürlichen zu unterscheiden sind, die nicht-natürlichen Ähnlichkeiten selbst nicht eindeutig voneinander getrennt werden können. Sprachverwandtschaft liegt nur dann vor, wenn die Ähnlichkeiten aus der Verwandtschaft resultieren, nicht hingegen, wenn sie die Folge von Sprachkontakt sind. Leider gibt es diesbezüglich keine absoluten Universalien, es lassen sich jedoch gewisse grundlegende Tendenzen aufzeigen, die in den meisten Fällen verlässlich sind. So schlägt Aikhenvald (2006: 5) beispielsweise die folgende Skala vor:



Moment mal, was war noch mal das "core lexicon"? Irgendwie klingt das komisch vertraut...

3.2 Grabenkämpfe

Die Frage nach den „richtigen“ Ähnlichkeiten für den Nachweis von Sprachverwandtschaft führte zu einem regelrechten Grabenkampf in der historischen Linguistik, der sich um die Frage zündete, welche Teile des sprachlichen Systems als Grundlage für den Nachweis von Sprachverwandtschaft eigentlich in Frage kämen.

In diesem Zusammenhang stehen sich nachwievor zwei Lager in hasserfüllter Rivalität gegenüber: das Lager der „Grammatikalisten“, die der Ansicht sind, dass Sprachverwandtschaft nur durch „Grammatik“ (oder vielmehr paradigmatische Morphologie als Spezialfall von Grammatik) nachgewiesen werden kann, und das Lager der „Lexikalisten“, welche der Ansicht sind, Grammatik sei für den Nachweis von Sprachverwandtschaft weniger bedeutsam als durch Lautkorrespondenzen abgesicherte Entsprechungen (nachgewiesene Kognaten) in stabilen Teilen des Lexikons. In den traditionell stark vertretenen Disziplinen der historischen Sprachwissenschaft (insbesondere der Indogermanistik) überwiegt dabei die Zahl derer, welche nur, oder vorwiegend, grammatische Evidenz als Nachweis von Sprachverwandtschaft einfordern und gestatten.

Die beiden unten wiedergegebenen Zitate aus einem Aufsatz von Dybo und Starostin nennen zwei offensichtliche Gründe, die für die Grammatik als Grundlage des Nachweises von Sprachverwandtschaft zu sprechen scheinen. Welche sind das?

One major reason for this is historical: it is no big secret that the Indo-European family was recognized primarily on the basis of the amazing similarity between the paradigmatic systems of Old Indian and classic European languages like Greek or Latin, and, since the general methodology of comparative linguistics grew out of working with Indo-European languages, morphological comparison, by the very force of tradition, is still held in high esteem and frequently suggested as a universal means for establishing relationship. (Dybo and Starostin 2008: 124f)

Another reason lies in the intuitive sphere. Morphology (and grammar in general) is traditionally seen as the “skeleton” of the language, its main constituent which, in comparison with lexics that “comes and goes”, is relatively stable and thus far more valid for the first stage of comparison. Thus, if the languages compared do not seem to share much common morphology, but are nevertheless quite close lexically, for many linguists the obvious explanation will be that the languages are not related, but show traces of extensive contacts (“convergence”). (ibid.: 125)

Die Grammatikalisten

Gewöhnlich gilt in der Indogermanistik Franz Bopp (1791 – 1867) mit seinem 1816 erschienenen Buch „Über das Conjugationssystem der Sanskritsprache, in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen (insbesondere gothischen) Sprache“ als Begründer des Verfahrens zum grammatischen Nachweis von Sprachverwandtschaft. „Während vorher Vermutungen nur durch einzelne Wortvergleiche gestützt waren, erfolgte Bopps Nachweis der Verwandtschaft über den Vergleich der Grammatik“ (Meier-Brügger 2002: 12).

Der prägende Einfluss auf die Indogermanistik und auch Teile der allgemeinen historischen Sprachwissenschaft stammt jedoch von Antoine Meillet (1866 – 1936), der sich in seinem Buch „La méthode comparative en linguistique historique“ aus dem Jahre 1925 betont, dass der Nachweis von Sprachverwandtschaft vorwiegend auf grammatischen Ähnlichkeiten zwischen Sprachen beruhen sollte.

In einem neueren Artikel, der bewusst auf den Aussagen Meillets aufbaut, prägte Nichols (1996) als moderne Vertreterin der Grammatikalisten den Begriff der „individuell-identifizierenden Evidenz“ (*individual-identifying evidence*) als oberstem Kriterium für den Nachweis von Sprachverwandtschaft. Gemeint ist damit

„evidence that identifies a unique individual proto-language rather than [...] evidence that identifies a set of languages or a type of language“ (Nichols 1996: 48), also Evidenz, die nicht typologischer sondern sehr spezifischer Natur ist, die also auf eine Ursprache mit individuellen Strukturen verweist und nicht auf eine Sprachen mit bestimmten typologischen Eigenschaften. Diese Evidenz ist laut Nichols „primarily grammatical and includes morphological material with complex paradigmatic and syntagmatic organization“ (*ibid.*: 41).

Unten ist ein Zitat von Meillet wiedergegeben, in dem er erläutert, worin die Beweiskraft grammatischer Indizien besteht. Worin genau besteht das, was Meillet als "grammatische Korrespondenzen" bezeichnet? Kann man die Wortgleichungen Deutsch *gut, besser, (am) best(en)* [gu:t, besər, best] = Englisch *good, better, best* [gud, betər, best] als Beispiel für diese Art von Evidenz ansehen? Welche Schwachstellen hat die grammatischistische Position?

Grammatical correspondences are proof, and only they are rigorous proof, provided one makes use of the material detail of the forms and that it is established that particular grammatical forms used in the languages under consideration go back to a common source. (Meillet 1925, Übersetzung übernommen von Nichols 1996: 47)

Die Lexikalisten

Im Gegensatz zum offiziellen Mantra der Indogermanistik, dass Sprachverwandtschaft nur auf Basis von Grammatik (also paradigmatischer Morphologie) nachgewiesen werden könnte, ist es interessant, sich die Argumentationen einiger prominenter Vertreter der Disziplin ein wenig genauer anzuschauen. In vielen Fällen, in denen von Sprachverwandtschaft die Rede ist, wird nämlich gerade *nicht* paradigmatische Morphologie als Indiz für diese angeführt, sondern lexikalische Entsprechungen, die durch reguläre Lautkorrespondenzen abgesichert sind.

Obwohl in der Indogermanistik allgemein die Annahme vorherrscht, dass nur paradigmatische Morphologie tatsächlich den Nachweis von Sprachverwandtschaft erbringen könne, findet man doch in den Argumentationen verschiedenster Grammatikalisten immer wieder Verweise auf lexikalische Übereinstimmungen, die als (zusätzliche) Indizien für Sprachverwandtschaft angeführt werden. Das gesamte Verfahren der linguistischen Rekonstruktion schließlich, beruht zum großen Teil auf lexikalischen Gleichungen. Man kann davon ausgehen, dass - wenn zwei Sprachen erwiesenermaßen miteinander verwandt sind - immer auch lexikalische Entsprechungen gefunden werden können.

Der Grund, warum lexikalische Evidenzen von den Grammatikalisten verworfen werden, besteht darin, dass das Lexikon üblicherweise als der instabilste Teil von Sprachen angesehen wird. Wörter lassen sich leicht entleihen, werden durch andere ersetzt, oder verschwinden völlig aus dem Wortschatz. Dem Misstrauen in das Lexikon als tragfähigen Indizienlieferanten für Sprachverwandtschaftsnachweise stellen Vertreter der Lexikalisten das Konzept des **Basisvokabulars** gegenüber. Grundannahme ist dabei, dass es „einige Bereiche der Lexik [gibt], in denen Entlehnung fast unmöglich ist, bspw. Pronomen, Bezeichnungen für Körperteile, Verwandtschaftsbeziehungen, wichtige Naturerscheinungen, einige häufig verwendete Verben und Adjektive“ (Jachontov 1965: 14).³

³Meine Übersetzung, Originaltext: „[Но есть] некоторые области лексики, где заимствования

Das Konzept des Basisvokabulars wird traditionell mit dem Namen von Morris Swadesh (1909 -- 1967) in Verbindung gebracht (vgl. die vorherige Sitzung). Welche der Konzepte HAND, FRIDGE, FOOT, NOSE, SUN, SNOW, STONE, SUNDAY und EXCREMENT gehören wohl eher zum Basisvokabular, welche nicht, und warum?

Wer hat Recht?

Am Ende haben wahrscheinlich beide Lager ein bisschen Recht: Nichols (1996) liegt sicher richtig, wenn sie betont, dass Suppletiv-Paradigmen mitunter mehr Überzeugungskraft besitzen als einfache Paare von Lexemen. Dybo and Starostin (2008) haben aber sicherlich auch Recht, wenn sie betonen, dass es oftmals die lexikalischen Entsprechungen sind, die zum Nachweis von Sprachverwandtschaft angeführt werden, und dass ganz ohne diese wohl auch schwer nachzuweisen wäre, dass zwei Sprachen wirklich verwandt sind.

Was von den meisten Diskutanten in diesem Zusammenhang leider übersehen wird, ist, dass die Trennung in das grammatische und das lexikalische Lager an sich gar nicht vollzogen wird, denn auch der „grammatische Nachweis“ der Sprachverwandtschaft setzt ja voraus, dass Lautkorrespondenzen ermittelt werden können. Allein die Tatsache, dass eine Sprache ein suppletives Komparationsparadigma für ein Wort mit der Bedeutung „gut“ aufweist, genügt nicht. Nur im Zusammenhang mit den nachweisbaren Lautkorrespondenzen erlangt es unmittelbare Beweiskraft.

Anstelle der leidigen Unterscheidung von grammatischen und lexikalischen Indizien kann daher meines Erachtens der Nachweis von „systematisch-funktionalen“ Ähnlichkeiten als Grundlage für den Nachweis von Sprachverwandtschaft angesehen werden. Unter „systematisch-funktional“ wird dabei verstanden, dass die Ähnlichkeiten einerseits auf einem Vergleich ganzer (Teil-)Systeme von Sprachen beruhen, und andererseits als funktionale Entsprechungen (also ähnlich in Bezug auf das „Funktionieren“ sprachlicher Elemente) aufgezeigt werden können. Konkret sind damit vorwiegend regelmäßige Lautkorrespondenzen in kognaten Wörtern und Morphemen gemeint. Das lässt sich in etwa wie folgt zusammenfassen:

Systematische Ähnlichkeiten: Als Folge regelmäßigen Lautwandels lassen sich bis zu einer gewissen Zeittiefe systematische Ähnlichkeiten im Lexikon miteinander verwandter Sprachen auffinden. *Systematisch* impliziert hierbei, dass die Ähnlichkeiten *sprachspezifisch* sind.

Reguläre Lautkorrespondenzen: Die grundlegenden Sprachspezifischen Ähnlichkeiten, welche nachweisbar sind, sind sich *funktional entsprechende Laute* im lexikalischen System miteinander verwandter Sprachen. *Funktional* meint hierbei, dass die einander entsprechenden Laute in den jeweiligen Systemen die gleiche *distinktive Funktion* innehaben. Sie dienen also auf ähnliche Weise der Bedeutungsunterscheidung, insofern als sie in ähnlichen *Positionen* in kognaten Wörtern auftauchen. Eine *substantielle* Ähnlichkeit der Lautsegmente ist dabei nicht zwingend erforderlich, auch wenn sich korrespondierende Lautsegmente in nah miteinander verwandten Sprachen meist auch substantiell ähneln.

почти невозможны, например: местоимения, названия частей тела, родственных отношений, важнейших явлений природы, некоторые наиболее употребительные глаголы и прилагательные и т. п.”

Kognaten: Können reguläre Lautkorrespondenzen für zwei oder mehr Sprachen nachgewiesen werden, lassen sich zwangsläufig auch *Kognaten* identifizieren. Kognaten sind Wörter, die von den Sprachen aus einer gemeinsamen Vorgängersprache geerbt wurden. Kognaten lassen sich nur mit Rückgriff auf reguläre Lautkorrespondenzen nachweisen, jedoch sind die Lautkorrespondenzen keine zwingende Voraussetzung dafür, dass zwei Wörter kognat *sind*, denn die Sprachen könnten sich auch so weit auseinanderentwickelt haben, dass keine Kognazität mehr nachgewiesen werden kann. In diesem Falle kann eine Sprachverwandtschaft jedoch auch nicht mehr nachgewiesen werden.

In der Indogermanistik wird oft betont, dass Sprachverwandtschaft nicht die beste, sondern die einzige Erklärung für die gegebenen Indizien sei. Inwiefern ist dies problematisch und woran mag es liegen, dass gerade dies in der historischen Linguistik so oft betont wird?

4 Lösung zur Aufgabe von letzter Woche

- **Aufgabe 1:** 1 - Sonsorol, 2 - Voleai, 3 - Puluvat, 4 - Truk
- **Aufgabe 2:** Sataval - A

Literatur

- Aikhenvald, A. Y. (2006). "Grammars in contact. A cross-linguistic perspective". In: *Grammars in contact*. Ed. by A. Y. Aikhenvald and R. M. W. Dixon. Oxford: Oxford University Press, 1–66.
- Dybo, A. and G. Starostin (2008). "In defense of the comparative method, or the end of the Vovin controversy". In: *Aspekty komparativistiki*. Vol. 3. Ed. by I. S. Smirnov. Orientalia et Classica XI. Moscow: RGGU, 119–258. PDF: <http://starling.rinet.ru/Texts/compmeth.pdf>.
- Gabelentz, H. G. C. (1891). *Die Sprachwissenschaft. Ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: T. O. Weigel. Internet Archive: [diesprachwissen00gabegoog](http://www.archive.org/details/diesprachwissen00gabegoog).
- Jachontov, S. E. (1965). *Drevnekitajskij jazyk* [Old Chinese]. Moscow: Nauka.
- Jones, W. (1798). "The third anniversary discourse, delivered 2 February, 1786, by the president. On the Hindus". In: *Asiatick Researches: or, Transactions of the Society Instituted in Bengal for Inquiring into the History and Antiquities, the Arts, Sciences, and Literature, of Asia* 1, 415–431. Google Books: [iSSoAAAAYAAJ](http://books.google.com/books?id=iSSoAAAAYAAJ).
- Meier-Brügger, M. (2002). *Indogermanische Sprachwissenschaft*. In collab. with M. Fritz and M. Mayrhofer. 8th ed. Berlin and New York: de Gruyter.
- Meillet, A. (1925). *La méthode comparative en linguistique historique*. Oslo and Leipzig: Aschehoug.
- Nichols, J. (1996). "The comparative method as heuristic". In: *The comparative method reviewed. Regularity and irregularity in language change*. Ed. by M. Durie. With an intro. by M. D. Ross and M. Durie. New York: Oxford University Press, 39–71.
- Sassetti, P. (1855). "Lettere edite e inedite di Filippo Sassetti: Raccolte e annotate da Ettore Marcucci". In: *Google Book Search*.
- "Statuts" (1871). "Statuts. Approuvés par décision ministérielle du 8 Mars 1866". In: *Bulletin de la Société de Linguistique de Paris* 1, III–IV.

Sequenzalinierung

1 Allgemeines vorweg

1.1 Sequenzen

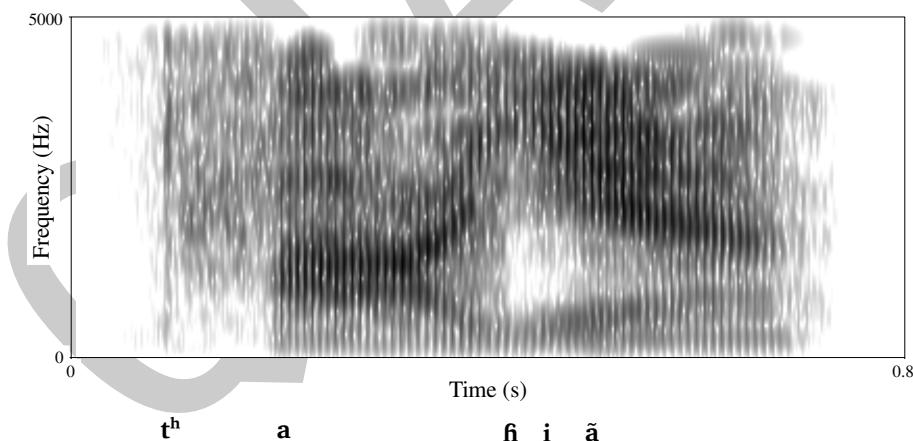
Viele Strukturen, mit denen wir es im Alltag und in der Wissenschaft zu tun haben, können als *Sequenzen* dargestellt werden. Der Vogelgesang, der uns morgens weckt, ist eine Sequenz von Schallwellen, die Filme, die wir anschauen, sind Sequenzen von Bildern, und die Gerichte, die wir kochen, basieren auf den Sequenzen von Instruktionen, die wir einem Rezept entnommen haben.

Was haben Rezepte, Vogelgesang und Filme gemein?

Diskrete und kontinuierliche Entitäten

Oft sind die Objekte, die wir als Sequenzen modellieren, nicht *diskret*, sondern das Ergebnis kontinuierlicher Variablen (Raum, Zeit, etc., vgl. Kruskal 1983: 130).

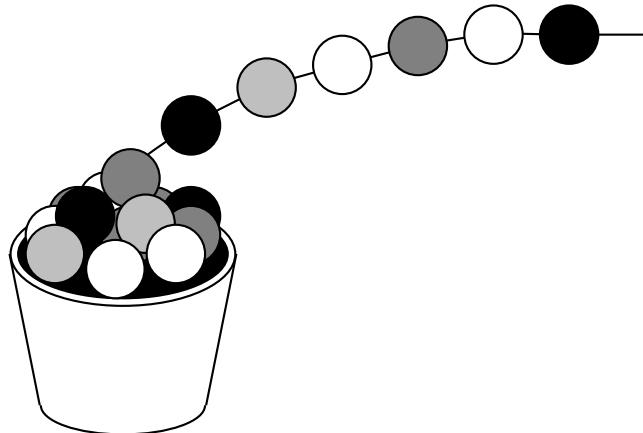
In der Linguistik hat das „Diskretmachen“ des Kontinuierlichen eine lange Tradition. Da unser Blick auf Sprache normalerweise geblendet wird vom alphabetischen Denken, wird häufig ignoriert, dass die natürliche Erscheinungsform des Sprechens ein Kontinuum ist und dass die Segmentierung der Rede in Lauteinheiten das Ergebnis einer expliziten Analyse ist: „Neither the movements of the speech organs nor the acoustic signal offers a clear division of speech into successive phonetic units“ (IPA Handbook 1999: 5).



Die Graphik zeigt das Spektrogramm der Aussprache von Shanghaisch *tàiyáng* 太阳 [t̪a₃₃fiā₄₄] ``Sonne''. Was fällt auf, wenn man die IPA-Transkription des Wortes mit dessen Widerspiegelung im Spektrogramm vergleicht?

Formale Definition von Sequenzen

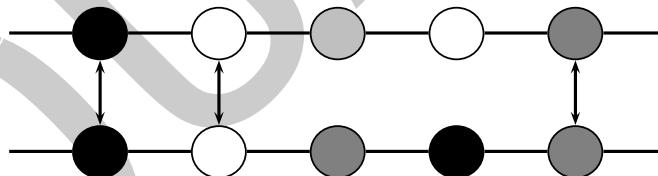
Definition: Ein *Alphabet* ist eine nicht-leere endliche Menge deren Elemente *Buchstaben* genannt werden. Eine *Sequenz* ist eine geordnete Liste von Buchstaben, die aus dem Alphabet gezogen werden. Die Elemente von Sequenzen werden *Segmente* genannt, die *Mächtigkeit* einer Sequenz ist die Anzahl ihrer unterschiedlichen Buchstaben, und die *Länge* einer Sequenz ist die Anzahl ihrer Segmente. (vgl. Böckenbauer and Bongartz 2003: 30f)



Was hat die Abbildung mit den Perlen auf der Schnur mit Sequenzen und Mengen zu tun?

1.2 Sequenzvergleiche

Sequenzvergleiche können mitunter recht kompliziert werden. Es genügt nicht, Sequenzen als einfache Menge aufzufassen und bspw. festzustellen, aus welchen Segmenten sie bestehen. Da Sequenzen sowohl eine *Struktur* als auch einen *Gehalt* haben, muss ein Vergleich von Sequenzen auf eben diese Tatsache Rücksicht nehmen. Bevor man also einzelne Segmente miteinander vergleicht, muss immer auch festgestellt werden, ob sich diese Segmente überhaupt *entsprechen*. Das ist leicht, wenn beide Sequenzen die gleiche Struktur haben. Dann muss man nämlich lediglich vergleichen, an welchen Positionen die Sequenzen sich unterscheiden. Dies lässt sich sogar relativ leicht quantifizieren, indem man einfach die unterschiedlichen Positionen zählt.

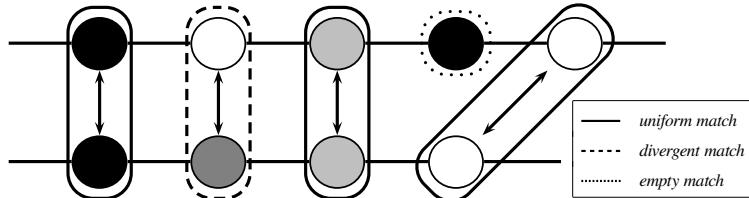


Die Anzahl unterschiedlicher Positionen zwischen zwei Sequenzen gleicher Struktur ist auch bekannt als die sogenannte *Hamming-Distanz*, benannt nach R. W. Hamming (1915 -- 1998), der sie erstmals in einem Aufsatz aus dem Jahre 1950 einföhrte (Hamming 1950). Wie groß ist die Hamming-Distanz zwischen den beiden Perlenketten auf der Abbildung oben?

Die Korrespondenzperspektive

Um Sequenzen mit unterschiedlicher Struktur vergleichen zu können, reicht die Hamming-Distanz allein nicht aus. Will man zwei oder mehr Sequenzen unterschiedlicher Struktur vergleichen, ist es zunächst notwendig, zu ermitteln, wie die Segmente miteinander *korrespondieren*. Dabei ist es zunächst wichtig, festzustellen, ob die Segmente überhaupt korrespondieren. Wenn zwei Segmente miteinander korrespondieren, so wird dies als *Match* bezeichnet. Wenn ein Element zu keinem anderen Segment korrespondiert, wird dies als *leeres Match* (*empty*

match) bezeichnet. Die Matches selbst können wiederum unterteilt werden in unterschiedliche Typen, und zwar in solche, bei denen die korrespondierenden Segmente identisch sind (*einheitliches Match, uniform match*) und solche, bei denen die korrespondierenden Segmente unterschiedlich sind (*divergierendes Match, divergent match*).



Die obige Abbildung fasst die grundlegenden Typen von Segment-Korrespondenzen zusammen. Mit diesen kann man prinzipiell alle Unterschiede zwischen Sequenzen beschreiben. Man könnte die grundlegenden Typen von Sequenzkorrespondenzen jedoch auch um mindestens zwei weitere Typen erweitern. Welche sind dies wohl?

Die Editierperspektive

Die zuvor besprochene Korrespondenzperspektive stellt nur eine mögliche Perspektive dar, um Unterschiede zwischen Sequenzen zu modellieren. Eine in der Literatur weit verbreitete Perspektive ist die sogenannte *Editierperspektive*. Die grundlegende Idee dieser Perspektive ist es, Unterschiede zwischen Sequenzen mit Hilfe von *Editieroperationen* darzustellen, also den Basisoperationen, die man benötigt, um die eine Sequenz in die andere zu transformieren. Die grundlegenden Editieroperationen sind hierbei die *Substitution*, die *Insertion*, und die *Deletion* (zusammengefasst auch als *Indel* bezeichnet). Sie wurden zuerst von dem russischen Mathematiker V. I. Levenshtein (? - ?) eingeführt (Levenshtein 1965).

Perspective	
Correspondence	Edit
uniform match	
divergent match	
empty match	insertion deletion
	transposition
complex match	expansion

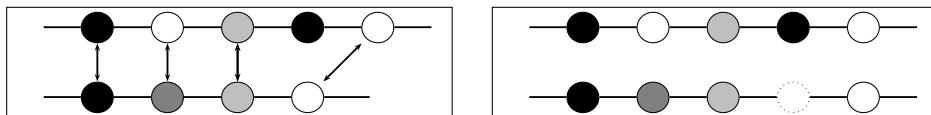
Oh weh! Flix der Felertäufel hat wieder zugeschlagen und in der obigen Tabelle eine Reihe von Zellen geweißt. Schaffst Du es, die ursprünglichen Zelleninhalte wiederherzustellen?

1.3 Alinierung

Formale Definition

Die *Alinierung* stellt die gebräuchlichste Form dar, um Unterschiede zwischen Sequenzen aufzuzeigen. Formal kann man eine Alinierung in etwa wie folgt definieren:

Eine **Alinierung** von n ($n > 1$) Sequenzen ist eine n -reihige Matrix, in der alle Sequenzen dergestalt angeordnet werden, dass alle matchenden Segmente in derselben Spalte erscheinen, während nicht-matchende Segmente, die aus leeren Matches resultieren, durch Gap-Symbole angezeigt werden. (vgl. Gusfield 1997: 216)



Die Levenshtein-Distanz zwischen zwei Sequenzen S_1 und S_2 ist definiert als die Anzahl von Editieroperationen, die notwendig ist, um S_1 in S_2 zu transformieren. Mit Hilfe des Konzepts der Alinierung lässt sich dieses Distanzmaß leicht auf die Hamming-Distanz zurückführen. Wie genau?

Alinierungsmodi

Wir haben bei der Beschreibung der Alinierung bisher stets angenommen, dass diese sich auf die Sequenzen in ihrer ganzen Länge beziehen muss. Eine lediglich teilweise Alinierung ist jedoch genau so gut möglich und zuweilen sogar durchaus sinnvoll. Je nachdem, welche Teile von Sequenzen bei der Alinierung genau berücksichtigt werden, lassen sich unterschiedliche **Alinierungsmodi** unterscheiden:

- **semi-globale Alinierung** beruht nicht notwendigerweise auf dem vollständigen Vergleich zweier Sequenzen. Stattdessen ist es möglich, Präfixe und Suffixe zu ignorieren, wenn diese die Kosten einer optimalen Alinierung zu stark erhöhen würden,
- **lokale Alinierung** berücksichtigt nur die Alinierung von Subsequenzen, wobei der Rest der Sequenzen ignoriert wird,
- **diagonale Alinierung** setzt eine globale Alinierung aus lokalen Teilalinierungen von *Diagonalen*, also Alinierungen, die keine Lücken enthalten, zusammen.

Modus	Alinierung														
global	G R E E N C A T F I S H H U N T E R A F A T C A T - - - H U N T E R														
semi-global	G R E E N - C A T F I S H H U N T E R - - - - A F A T C A T H U N T E R														
lokal	GREEN CATFISH H U N T E R A F A T CAT H U N T E R														
diagonal	- - - - G R E E N C A T F I S H H U N T E R A F A T - - - - C A T - - - H U N T E R														

Die obige Tabelle zeigt, wie im Rahmen der vier Alinierungsmodi zwei Beispieldaten aliniert werden. Versuche, die Unterschiede darzustellen.

2 Phonetische Alinierung

Obwohl Alinierungsanalysen eine der allgemeinsten Möglichkeiten darstellen, Sequenzen zu vergleichen, steckt ihre Verwendung in der historischen Linguistik

noch in den Kinderschuhen. Natürlich alinieren historische Linguisten eigentlich *immer* Wörter und haben dies auch schon immer getan, da ohne Alinierung überhaupt keine regulären Lautkorrespondenzen ermittelt werden könnten. Der Sprachvergleich basierte lange Zeit jedoch eher auf einer impliziten Alinierung, die selten visualisiert wurde, und wenn doch, dann nur aus illustrativen Zwecken.

Language	Alignment						
Russian	s	-	ɔ	n	ts	ə	-
Polish	s	w	ɔ	n ^j	ts	ɛ	-
French	s	-	ɔ	l	-	ɛ	j
Italian	s	-	o	l	-	e	-
German	s	-	ɔ	n	-	ə	-
Swedish	s	-	u:	l	-	-	-

(a) Globale Alinierung

Language	Alignment						
Russian	s	ɔ	-	-	n	ts	ə
Polish	s	-	w	ɔ	n ^j	ts	ɛ
French	s	ɔ	l	-	-	-	-
Italian	s	o	l	-	-	-	e
German	s	ɔ	-	-	-	-	nə
Swedish	s	u:	l	-	-	-	-

(b) Lokale Alinierung

Die obige Tabelle stellt zwei verschiedene Alinierungen von einsprachlichen Reflexen von Urindogermanisch *séh₂uel- dar, eine scheinbar naheliegende globale Alinierung, und eine realistische lokale Alinierung. Was fällt beim Vergleich der beiden Alinierungen auf? Welche Gründe mag es geben, dass eine korrekte Alinierung in der historischen Linguistik so verdammt schwierig ist? Warum sind Alinierungen dennoch so unheimlich wichtig für die historische Linguistik?

2.1 Die zwei Arten von Ähnlichkeit

Die deutschen Wörter *schlafen* und *Flaschen* sind recht ähnlich, da sie – phonetisch gesehen – aus sechs verschiedenen Lauten bestehen, die einander recht ähnlich sind. Ähnliche Ähnlichkeit besteht zwischen den Wörtern *Obst* und *Post*. Eine andere Form von Ähnlichkeit besteht zwischen den Wörtern *Kerker* und *Tanten*. Die Ähnlichkeit besteht hier nicht in der Substanz der Lautsequenzen, sondern in der Struktur: beide Wörter bestehen aus einer identischen Kette unterschiedlicher Zeichen, wie man leicht sehen kann, wenn man die Wörter aliniert: Jedes distinkte Element korrespondiert direkt mit einem distinkten Element des anderen Wortes, und das Umwandeln der einen Lautsequenz in die andere kann durch eine einfache Ersetzungstabelle erreicht werden, was nicht möglich ist für *schlafen* und *Flaschen*, da die beiden Lautsequenzen nicht strukturell äquivalent sind. Basierend auf diesen Überlegungen können wir zwei grundlegende Arten von Ähnlichkeiten zwischen Wörtern unterscheiden: *substantielle* und *strukturelle* Ähnlichkeit.

Lass (1997: 130) verwendet die Termini *phänotypische* und *genotypische* Ähnlichkeit um zufällige oberflächliche Ähnlichkeiten zwischen Wörtern von Ähnlichkeiten zu unterscheiden, die aus der Kognazität von Wörtern resultieren. In welchem Zusammenhang steht dieses Ähnlichkeitskonzept zu dem von *substieller* und *struktureller* Ähnlichkeit?

2.2 Typen des Lautwandels

Die lange Forschungstradition in der historischen Linguistik hat dazu geführt, dass eine beträchtliche Reihe von *Lautwandeltypen* von verschiedenen Forschern postuliert wurde. Leider liegt diesen Lautwandeltypen eine unstete Terminologie zugrunde, die von sehr konkreten Termini, die sehr konkrete Arten von Lautwandel beschreiben, bis hin zu sehr generellen Termini, die sehr abstrakte Klassen

von Lauten abdecken, reicht. Was in der Literatur als „Lautwandeltyp“ bezeichnet wird, reicht von Phänomenen wie dem „Rhotazismus“ (Trask 2000: 288), der, einfach gesagt, auf den Wandel von [s] zu [r] referiert, bis hin zur *Lenisierung* (engl. *lenition*), die einen Wandeltyp bezeichnet, „in which a segment becomes less consonant-like than previously“ (ibid.: 190). Manche Termini sind ferner eher „erklärend“ denn „beschreibend“, da sie gleichzeitig einen Grund dafür nennen, warum sich ein Laut gemäß einem bestimmten Typ wandelt. So wird *Assimilierung* in vielen Textbüchern nicht nur beschreiben als „[a] change in which one sound becomes more similar to another“, sondern es wird gleichzeitig betont, dass dies „through the influence of a neighboring, usually adjacent, sound“ geschieht (Campbell and Mixco 2007: 16). Dies ist natürlich problematisch, da eine Beschreibung einer Erklärung immer vorangehen sollte.

In der folgenden Tabelle werden fünf mehr oder weniger triviale Lautwandeltypen abgeleitet, indem, basierend auf der Gleichsetzung von Lautwandel mit einer Funktion, lediglich das Verhältnis zwischen Eingabe und Ausgabe als Klassifikationskriterium zugrunde gelegt wird. Da Lautwandel kontextabhängig verlaufen kann, ist es gut möglich, dass eine solche Funktion zusätzliche Parameter benötigt (wie die Silbenumgebung, den vorangehenden oder nachfolgenden Laut, usw.).

Typ	Beschreibung	Darstellung	Beispiel
Kontinuation		$x > x$	Althochdeutsch <i>hant</i> > Deutsch <i>Hand</i>
Substitution	Ersetzung eines Lauts		Althochdeutsch <i>snēo</i> > Deutsch <i>Schnee</i>
Insertion	Gewinn eines Lauts	$\emptyset > y$	Althochdeutsch <i>ioman</i> > Deutsch
	Verlust eines Lauts	$x > \emptyset$	Althochdeutsch <i>angust</i> > Deutsch <i>Angst</i>
Metathesis	Umsiedlung	$xy > yx$	Urslavisch <i>*žbltъ</i> > Tschechisch <i>žlutý</i> „gelb“

Flix der Felertäufel treibt es aber heute wirklich mal wieder zu bunt! Da hat er schon wieder einige Zellen geweißt. Kannst Du sie fachgerecht rekonstruieren? Woran erinnern Dich diese Typen des Lautwandels?

2.3 Lautklassen

Beim Alinieren in der historischen Linguistik ist es wichtig, der Tatsache Rechnung zu tragen, dass substantielle Ähnlichkeit zwischen Lauten nicht notwendigerweise auch auf deren Kognazität hinweist. Nur, wenn zwei Segmente auch systematisch (also in den Sprachsystemen) korrespondieren, sollten sie tatsächlich als ähnlich angesehen werden. In den Schritten des Sprachvergleichs kann diese systematische Ähnlichkeit jedoch schwer ermittelt werden, denn zu Beginn eines Sprachvergleichs sind weder die kognaten Wörter bekannt, noch die regulär korrespondierenden Laute. Für automatische Ansätze zur phonetischen Alinierung und zur automatischen Kognatenerkennung ist es daher wichtig, einen heuristischen Ansatz zu entwickeln, der nicht nach *absolut* korrespondierenden Segmenten sucht, sondern nach *wahrscheinlich* korrespondierenden. Theoretische Grundlage eines solchen Ansatzes ist die von vielen Autoren geteilte Überzeugung, dass die unterschiedlichen Typen von Sprachwandel mit unterschiedlicher Häufigkeit und Wahr-

scheinlichkeit auftreten. Daraus folgt weiter, dass sich in den Sprachen der Welt bestimmte Muster von Lautkorrespondenzen häufiger finden lassen, als andere.

Um dieser „Korrespondenzwahrscheinlichkeit“ im automatischen Sprachvergleich Rechnung zu tragen, bietet sich das Konzept der „Lautklassen“ an, das erstmals von Dolgopolsky (1964) vorgeschlagen wurde. Grundlegende Idee dabei ist, dass Laute, die häufig in Korrespondenzbeziehung in genetisch verwandten Sprachen auftreten, in Klassen (Typen) eingeteilt werden können. Es wird dabei angenommen, dass „phonetic correspondences inside a ‚type‘ are more regular than those between different ‚types‘“ (ibid.: 35).

No.	Cl.	Description	Examples
1	"P"	labial obstruents	p, b, f
2	"T"	dental obstruents	d, t, θ, ð
3	"S"	sibilants	s, z, ſ, ʒ
4	"K"	velar obstruents, dental and alveolar affricates	k, g, ts, tʃ
5	"M"	labial nasal	m
6	"N"	remaining nasals	n, ŋ, ɳ
7	"R"	liquids	r, l
8	"W"	voiced labial fricative and initial rounded vowels	v, u
9	"J"	palatal approximant	j
10	"Ø"	laryngeals and initial velar nasal	h, fi, ɳ

Die Tabelle oben zeigt Dolgopolskys ursprüngliches Lautklassenschema. Was fällt auf, wenn man die oben wiedergegebenen Reflexe des Urindogermanischen Worts für Sonne entsprechend diesem Schema in Lautklassensequenzen umwandelt?

2.4 Sekundärstrukturen

Abgesehen von einer *primären Struktur* können Sequenzen auch eine *sekundäre Struktur* haben. Primäre Struktur meint hier die Ordnung der Segmente. Unter sekundärer Struktur wird die Anordnung *sekundärer Segmente*, d. h. von Segmenten, die aus dem Gruppieren von primären Segmenten zu höheren Einheiten resultieren, verstanden. Worauf genau die sekundäre Struktur basiert, ist dabei nebenschließlich. Im Zusammenhang mit phonetischen Sequenzen spielen jedoch sekundäre Segmente wie *Silben*, *Morpheme*, *Wörter* und *Sätze*, eine mitunter wichtige Rolle. In der phonetischen Alinierungen kommt die wichtigste Rolle dabei der Morphemstruktur von Wörtern zu, da diese bei der manuellen Alinierung ohnehin meist implizit angewendet wird, und erst hilft eine realistische Alinierung von Lautsequenzen vorzunehmen.

Die Tabelle unten gibt ein Beispiel für die Unterschiede zwischen primärer und sekundärer Alinierung. Während die primäre Alinierung eine falsche Korrespondenz zwischen dem finalen [t] und dem initialen [tʰ] ansetzt, setzt die sekundäre im Falle von Yinchuan richtig zwei Morpheme zj⁵¹ "Sonne" und tʰou¹ "Kopf (Suffix)" an, denen ein Morphem zit³ "Sonne" in Haikou gegenübersteht.

Primäre Alinierung						
Haikou	z	i	-	t	-	3
Beijing	ẓ	ḷ	51	ṭ ^h	ou	1

Sekundäre Alinierung						
Haikou	z	i	t	3	-	-
Beijing	ẓ	ḷ	-	51	ṭ ^h	ou

Da die Morphemstruktur chinesischer Wörter meist eindeutig mit deren Silbenstruktur gleichzusetzen ist (eine Silbe = ein Morphem), ist die sekundäre Alinierung in diesen Fällen einfach. Wie sieht dies bei anderen Sprachen aus?

Literatur

- Böckenbauer, H.-J. and D. Bongartz (2003). *Algorithmische Grundlagen der Bioinformatik*. German. Stuttgart, Leipzig, and Wiesbaden: Teubner.
- Campbell, L. and M. Mixco (2007). *A glossary of historical linguistics*. Edinburgh: Edinburgh University Press.
- Dolgopolsky, A. B. (1964). "Gipoteza drevnejšego rodstva jazykovych semej Severnoj Evrazii s verojatnostej točky zrenija [A probabilistic hypothesis concerning the oldest relationships among the language families of Northern Eurasia]". In: *Voprosy Jazykoznanija* 2, 53–63; English translation: Dolgopolsky, A. B. (1986). "A probabilistic hypothesis concerning the oldest relationships among the language families of northern Eurasia". In: *Typology, relationship and time. A collection of papers on language change and relationship by Soviet linguists*. Ed. and trans. from the Russian by V. V. Shevoroshkin. Ann Arbor: Karoma Publisher, 27–50.
- Gusfield, D. (1997). *Algorithms on strings, trees and sequences*. Cambridge: Cambridge University Press.
- Hamming, R. W. (1950). "Error detection and error detection codes". In: *Bell System Technical Journal* 29.2, 147–160.
- IPA Handbook (1999). *Handbook of the International Phonetic Association. A guide to the use of the international phonetic alphabet*. Cambridge: Cambridge University Press.
- Kruskal, J. B. (1983). "An overview of sequence comparison. Time warps, string edits, and macromolecules". In: *SIAM Review* 25.2, 201–237. JSTOR: 2030214.
- Lass, R. (1997). *Historical linguistics and language change*. Cambridge: Cambridge University Press.
- Levenshtein, V. I. (1965). "Dvoičnye kody s ispravleniem vypadenij, vstavok i za-meščenij simvolov [Binary codes with correction of deletions, insertions and replacements]". In: *Doklady Akademij Nauk SSSR* 163.4, 845–848; English translation: – (1966). "Binary codes capable of correcting deletions, insertions, and reversals". In: *Soviet Physics Doklady* 10.8, 707–710.
- Trask, R. L., comp. (2000). *The dictionary of historical and comparative linguistics*. Edinburgh: Edinburgh University Press.

Automatische Kognatenerkennung

1 Die komparative Methode

Die Methode der heutigen Komparativistik, welche allgemein unter dem nicht sehr glücklichen Terminus „vergleichend-historische Methode“ bekannt ist, stellt eine große Gesamtheit an abstrakten und konkreten Verfahren zur Untersuchung der Geschichte verwandter Sprachen dar, die genetisch auf eine bestimmte einheitliche Tradition der Vergangenheit zurückgehen, welche man üblicherweise als Proto-Sprache oder Grundsprache qualifiziert. Dieses methodische Instrumentarium, auf welches zurückgegriffen wird, um eine große Menge verschiedener Probleme zu lösen, wird verwendet, um ein Erkenntnissystem über die historische Entwicklung von Sprachfamilien aufzubauen, welches seine endgültige Gestalt in From historisch-vergleichender Grammatiken erhält. (Klimov 1990: 6)¹

Ja, nee, is klar! Eine Methode, die aus einem Sammelsurium unterschiedlichster Verfahren besteht? Manchmal haben die historischen Linguisten wirklich ein paar Knoten locker. Wenn das ultimative Ziel dieser Methode darin bestehen soll, die historische Entwicklung von Sprachfamilien aufzuzeigen, dann kann man ja sicher eine ganze Menge Unterverfahren ansetzen, aus denen diese "Methode" besteht. Und die wären?

1.1 Grundannahmen

Zu den unumstößlichen Grundannahmen der komparativen Methode zählen die Annahme der *Regelmäßigkeit von Lautwandel* (*Regularitätspostulat*), der *Universalität von Sprachwandel* (*Universalitätspostulat*), und der *Möglichkeit von Sprachspaltung* (*Postulat der baumartigen Sprachentwicklung*). Entsprechend der ersten Annahme verläuft Lautwandel demnach dergestalt, dass er große Teile des Lexikons einer Sprache erfasst und lediglich vom phonetischen Kontext abhängt, nicht jedoch von anderen Faktoren, wie bspw. der Bedeutung der Wörter. Die zweite Annahme besagt, dass Sprachwandel unabhängig vom Ort, der Zeit und der Sprache eintritt, und die dritte besagt, dass unabhängiger Sprachwandel bei geographischer Trennung der Sprecher einer Sprache im Laufe der Zeit zu Sprachspaltung führen kann.

Welcher linguistischen Tradition lässt sich das Regularitätspostulat der komparativen Methode zuschreiben und worin bestehen die Probleme dieser Tradition? Ist das dritte Postulat wirklich wichtig genug, dass es in diesem Zusammenhang extra genannt werden muss?

¹Meine Übersetzung, Originaltext: «Методика современной компаративистики, широко известная в лингвистической литературе под довольно неудачным термином „сравнительно-исторический метод”, представляет собой большую совокупность методов и конкретных приемов изучения истории родственных языков, генетически восходящих к некоторой единой традиции прошлого, обычно квалифицируемой в качестве праязыка или языка-основой. Этот методический инструментарий, призванный обслуживать решение множества задач, используется для построения системы знаний об историческом развитии языковых семей, формируемой в конечном счете в виде сравнительно-исторических грамматик».

1.2 Arbeitsweise

Allgemeine Arbeitsweise

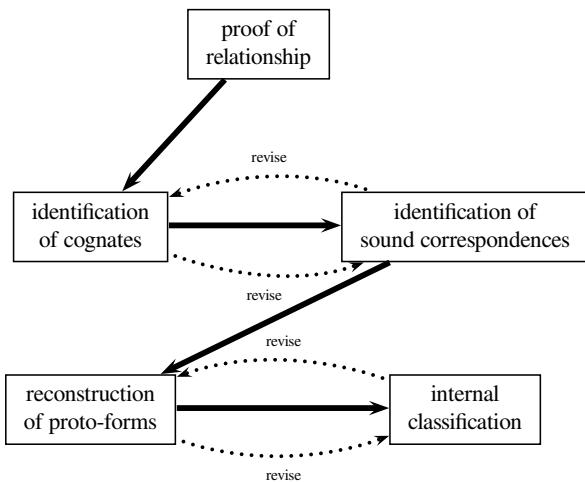
Hinsichtlich der allgemeinen Arbeitsweise der komparativen Methode sind sich die Forscher nicht immer einig. Eine der explizitesten Zusammenfassungen, die ich bisher finden konnte, stammt von Ross and Durie (1996: 6f):

1. Determine on the strength of diagnostic evidence that a set of languages are genetically related, that is, that they constitute a 'family';
2. Collect putative cognate sets for the family (both morphological paradigms and lexical items).
3. Work out the sound correspondences from the cognate sets, putting 'irregular' cognate sets on one side;
4. Reconstruct the protolanguage of the family as follows:
 - a Reconstruct the protophonology from the sound correspondences worked out in (3), using conventional wisdom regarding the directions of sound changes.
 - b Reconstruct protomorphemes (both morphological paradigms and lexical items) from the cognate sets collected in (2), using the protophonology reconstructed in (4a).
5. Establish innovations (phonological, lexical, semantic, morphological, morphosyntactic) shared by groups of languages within the family relative to the reconstructed protolanguage.
6. Tabulate the innovations established in (5) to arrive at an internal classification of the family, a 'family tree'.
7. Construct an etymological dictionary, tracing borrowings, semantic change, and so forth, for the lexicon of the family (or of one language of the family).

Nicht, dass das die einzige Frage wäre, die hier offen bleiben kann, aber was verstehen die Autoren wohl unter "diagnostischer Evidenz", und warum ist es wichtig, Sprachverwandtschaft zu Beginn der komparativen Methode bereits nachgewiesen zu haben?

1.2.1 Der Versuch eines Arbeitsschemas

Jeder historische Linguist hat wohl eine etwas andere Auffassung von dem, was die komparative Methode denn eigentlich sei. Daher ist es schwierig, einen verbindlichen Konsens zu finden. Basierend auf den bekanntesten Beschreibungen in der Literatur lässt sich jedoch ein fünfstufiges Arbeitsschema festlegen, wie es auf der Graphik zur Rechten zu sehen ist. Die gestrichelten Linien zwischen den unterschiedlichen Arbeitsschritten deuten dabei den iterativen Charakter der Methode an.



Die komparative Methode wird oft als iteratives Verfahren beschrieben, wobei der iterative Charakter als eine große Stärke der Methode hervorgehoben wird. Was bedeutet "iterativ" überhaupt, und warum sollte das eine Stärke sein?

Kognatenerkennung

Wenn wir nur das Verfahren zur Kognatenerkennung betrachten, das in der komparativen Methode zur Anwendung kommt, so lässt sich dieses in etwa wie folgt beschreiben:

- Erstelle eine erste Liste möglicher *Kognatensets*.
- *Aliniere* die Wörter in der Kognatenliste.
- Extrahiere eine erste Liste möglicher *Lautkorrespondenzen* aus den Alinierungen.
- Verbessere die Kognatenliste und die Korrespondenzliste durch
 - Hinzufügen von mit der Korrespondenzliste kompatiblen und Entfernen von mit der Korrespondenzliste nicht kompatiblen Kognatensets,
 - Hinzufügen von mit der Kognatenlisteliste kompatiblen und Entfernen von mit der Kognatenliste nicht kompatiblen Korrespondenzsets.
- Hör auf, wenn die Resultate publikationsreif sind.

Der iterative Charakter der komparativen Methode erstreckt sich streng genommen durch alle ihre Arbeitsschritte. Somit hängt auch das Verfahren zur Kognatenerkennung von den ihm im Schema vorangehenden (Nachweis von Sprachverwandtschaft) und den ihm nachfolgenden Schritten (Rekonstruktion und interne Klassifikation) ab. Versuche, zu erläutern, wie man sich diese gegenseitige Abhängigkeit vorstellen kann.

Zusammenfassung

Die komparative Methode ist ein iteratives Verfahren zur Kognatenerkennung, sowie zur linguistischen und zur phylogenetischen Rekonstruktion. Grundlage des Verfahrens zur Kognatenerkennung sind Verfahren zur phonetischen Alinierung und zur Endekung regulärer Lautkorrespondenzen.

Die Vorteile der komparativen Methode bestehen in ihrer *Flexibilität* und ihrer *Zuverlässigkeit*. Die Methode ist flexibel, insofern als sie gleichzeitig auf viele Sprachen (*multilingualer Aspekt*) und auf unterschiedlichste Sprachdaten angewendet werden kann (*multimodaler (?) Aspekt*). Ihre Verlässlichkeit begründet sich dadurch, dass die Wirksamkeit der Methode im Verlaufe der Geschichte inzwischen an vielen Beispielen aufgezeigt werden konnte. Derzeit gibt es noch keine wirkliche Alternative zu diesem Verfahren.

Bisher wurden nur die Vorteile der komparativen Methode aufgezeigt. Wie sieht es aber mit den Nachteilen aus?

2 Automatische Kognatenerkennung

2.1 Kognatenerkennung als Problem

In der Mathematik und Algorithmik ist es üblich, bestimmte Aufgaben als „Probleme“ zu klassifizieren. Der Gedanke, der wahrscheinlich dahintersteht ist, dass es leichter ist, etwas zu lösen, wenn man sich vorher klar darüber geworden ist, was eigentlich gelöst werden soll. In diesem Sinne können wir auch im Zusammenhang mit der automatischen Kognatenerkennung von einem „Problem“ sprechen, wobei wir zwischen unterschiedlichen Problemstufen unterscheiden können.

Bevor wir zu den richtigen Problemen kommen, sollten wir uns noch einmal kurz in Erinnerung rufen, was Kognaten eigentlich sind.

Generelles Problem der Kognatenidentifikation

Gegeben sind Wortlisten verschiedener Sprachen. Finde alle Wörter in diesen Listen, die kognat sind.

Charakteristik	Ja	Nein
Unterscheidung von Lehn- und Erbwörtern		
Semantische Ähnlichkeit wird vorausgesetzt		
Auf unbegrenzte Anzahl von Sprachen anwendbar		

Die Tabelle enthält drei Charakteristika des Problems der Kognatenidentifikation. Ob diese Charakteristika jedoch vorliegen, oder nicht, wurde offengelassen. Kreuze die Ausprägungen der Charakteristika an, die sich aus der Beschreibung des generellen Problems der Kognatenidentifikation ableiten lassen.

Generelles Problem der Identifikation etymologisch verwandter Wörter

Gegeben sind Wortlisten verschiedener Sprachen. Finde alle Wörter in diesen Listen, die etymologisch verwandt sind.

Charakteristik	Ja	Nein
Unterscheidung von Lehn- und Erbwörtern		
Semantische Ähnlichkeit wird vorausgesetzt		
Auf unbegrenzte Anzahl von Sprachen anwendbar		

Kreuze die entsprechenden Ausprägungen der Charakteristika des generellen Problems der Identifikation etymologisch verwandter Wörter an. Wie könnte man dieses Problem alternativ nennen?

Spezielles Problem der Identifikation etymologisch verwandter Wörter

Gegeben ist eine mehrsprachige Wortliste, also eine Wortliste, die semantisch alinierte Wörter verschiedener Sprachen enthält. Finde alle Wörter in dieser Liste, die etymologisch verwandt sind.

Charakteristik	Ja	Nein
Unterscheidung von Lehn- und Erbwörtern		
Semantische Ähnlichkeit wird vorausgesetzt		
Auf unbegrenzte Anzahl von Sprachen anwendbar		

Kreuze die entsprechenden Ausprägungen der Charakteristika des speziellen Problems der Identifikation etymologisch verwandter Wörter an. Woran mag es liegen, dass von den bisher postulierten Algorithmen zur automatischen Kognatenerkennung vorwiegend dieses Problem im Gegensatz zu den beiden zuvor genannten allgemeineren Problemen adressiert wurde?

2.2 Die Entdeckung von Lautkorrespondenzen

Signifikanz und Gradualität

Die Entdeckung von Lautkorrespondenzen ist ein stochastisches Problem. Es geht dabei weniger um die absolute Anzahl an Korrespondenzen, die im Rahmen eines Sprachvergleichs aufgefunden werden können, als vielmehr darum, ob diese Anzahl tatsächlich signifikant ist. Dies wiederum führt automatisch zu einer Abkehr vom üblicherweise bemühten „Absolutheitspostulat“: Traditionell werden Lautkorrespondenzen nämlich meist als absolut angesehen: Entweder treten sie auf, oder nicht. Tatsächlich aber sollten Lautkorrespondenzen angesichts der für die historische Linguistik typischen sich ständig ändernden Datenlage als graduell angesehen werden, wobei manche Lautkorrespondenzen wahrscheinlicher und manche weniger wahrscheinlich sind.

Manche Wissenschaften haben sich im Laufe ihrer Geschichte eine absolutierende Attitüde zugelegt. Die historische Linguistik gehört zweifellos auch dazu. Womit mag dies zusammenhängen? Lassen sich gute Gründe leichter erkennen, wenn man die historische Linguistik mit ähnlichen Wissenschaften vergleicht?

Quantifizieren von Lautkorrespondenzen

In der Bioinformatik ist es üblich, die Wahrscheinlichkeit von Korrespondenzen zu ermitteln, indem man attestierte mit erwarteten (zufälligen) Verteilungen vergleicht. Auf die Linguistik übertragen bedeutet das, dass wir eine attestierte Verteilung korrespondierender Laute vergleichen mit einer Verteilung, die wir erwarten würden, wenn zwei oder mehr Sprachen nicht miteinander verwandt sind. Normalerweise werden in der historischen Linguistik dabei lange Listen möglicher Kognaten angeführt, wie in der folgenden Liste, die Beispielwörter aus dem Italienischen und dem Französischen aufführt:

Meaning	Italian	French
“square”	pjats:a	plas
“feather”	pjuma	plym
“flat”	pjano	plā

Meaning	Italian	French
“tear”	lakrima	larm
“tongue”	lingwa	läg
“moon”	luna	lyn

Tatsächlich sind es aber nicht nur mögliche Kognaten, auf denen die Identifizierung erster möglicher Lautkorrespondenzen beruht, sondern mögliche *alinierte* Kognaten, wie sie die folgende Liste zeigt:

"square"	p j a ts: a
"feather"	p j u m a
"flat"	p j a n o

"tear"	l a k r i m a
"tongue"	l i ñ w a
"moon"	l u n a

Das Quantifizieren von Lautkorrespondenzen könnte man nun als einfachen Prozess des Zählens darstellen, für den man eine einfache Matrix anlegt, und in dieser alle Kookkurrenzen aller Lautkombinationen auflistet (siehe linke Tabelle unten). Tatsächlich wird man damit dem oben erwähnten Regularitätspostulat des Lautwandels nur unzulänglich gerecht, da Lautwandel ja nur selten unabhängig vom phonetischen Kontext stattfindet. Daher ist es sinnvoller, auch den Kontext in die Quantifizierung miteinzubeziehen, indem man Laute nicht nur durch ihre Substanz, sondern auch durch ihren sonorischen Kontext (initial, steigende, fallende Sonorität, final, etc.) charakterisiert (Tabelle links). Basierend auf derartigen Überlegungen lässt sich bereits ein der manuellen komparativen Methode recht nahe kommendes automatisches Modell der Kognatenerkennung erstellen.

	p	j	a	l	...
p	3	0	0	0	...
l	0	3	0	3	...
a	0	0	1	0	...
...

	p / #	j / C	a / C	l / C	...
p / #	3	0	0	0	...
l / #	0	0	0	3	...
l / C	0	3	0	0	...
a / V	0	0	1	0	...
...

Ist die Einbeziehung des phonetischen Kontexts für die Entdeckung regulärer Lautkorrespondenzen tatsächlich so entscheidend? Begründe Deine Meinung mit Rückgriff auf die Beispielwörter aus dem Italienischen und Französischen.

2.3 Clusterung

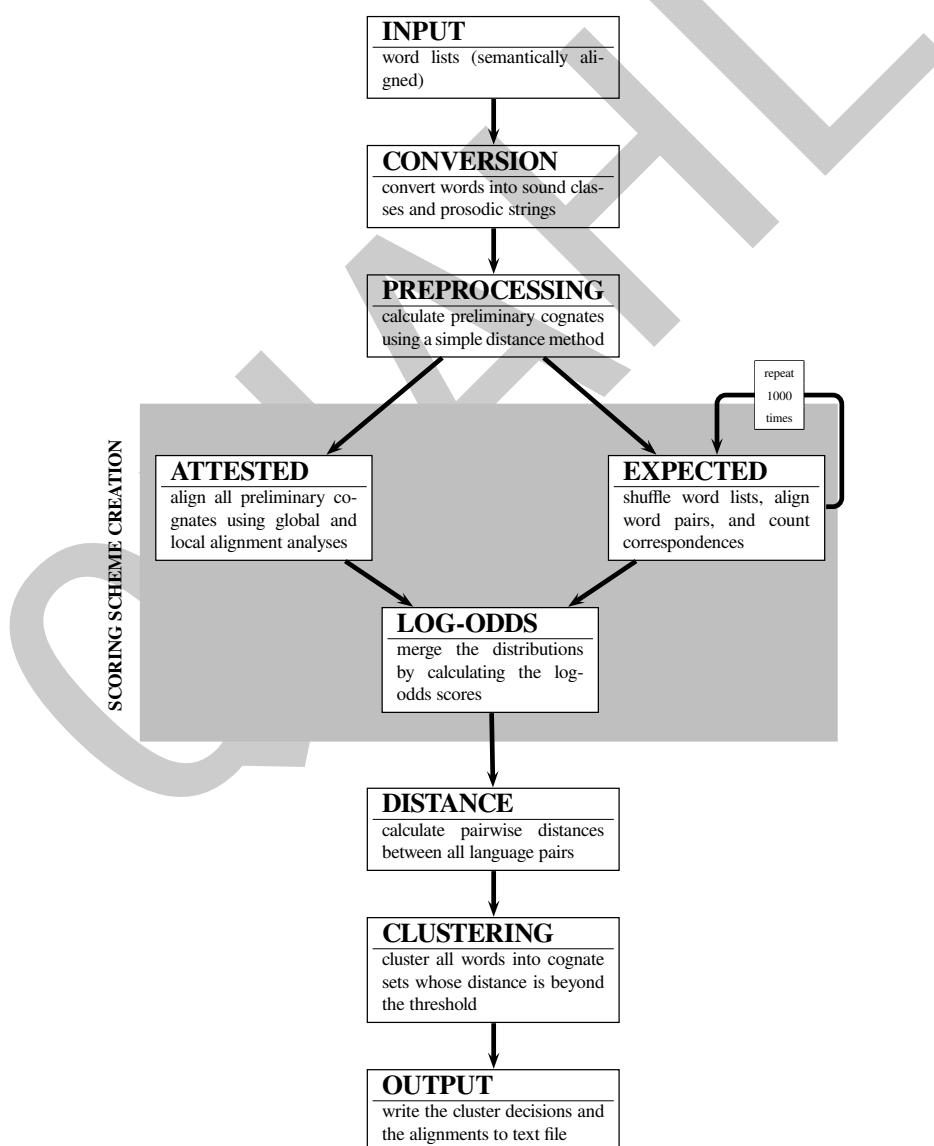
Unter *Clusterung* versteht man normalerweise die Einteilung von Objekten in Gruppen. Somit stellt die Unterteilung von Deutsch, Englisch, Italienisch und Französisch in germanische und romanische Sprachen eine Clusterung dar. Wenn wir uns die Probleme der Kognatenerkennung genauer anschauen, fällt auf, dass es sich auch hierbei primär um Clusteraufgaben handelt. Es geht nämlich darum, Wörter in Gruppen einzuteilen, wobei Wörter, die der gleichen Gruppe zugeordnet werden, jeweils als miteinander „kognat“ oder „etymologisch verwandt“ klassifiziert werden. Die Wörter Deutsch *Zahn* [tsa:n], Italienisch *dente* [dente], Niederländisch *tand* [tand], Russisch *зуб* [zup], und Englisch *tooth* [tu:θ] würde man dabei beispielsweise zwei verschiedenen Clustern zuschreiben, einem, dem alle Wörter angehören, die ein Reflex von Urindogermanisch *dehnt- „Zahn“ sind (*Zahn*, *dente*, *tand* und *tooth*), und einem, dem die Wörter angehören, die ein Reflex von Urindogermanisch *ǵombʰ-o- „Nagel“ sind (*zub*) (DERKSEN: 549).

	tsa:n	dente	tand	zup	tu:θ
tsa:n	0.00	0.53	0.35	0.57	0.57
dente	0.53	0.00	0.10	0.97	0.52
tand	0.35	0.10	0.00	0.86	0.39
zub	0.57	0.97	0.86	0.00	0.70
tu:θ	0.57	0.52	0.39	0.70	0.00

In automatischen Analysen haben Clusterverfahren den großen Vorteil, dass die Evidenz, die beim Vergleich von lediglich einem Sprachpaar fehlen mag, durch allgemeine Evidenz ausgeglichen werden kann. Die Verwendung multipler Evidenz (Sturtevant 1920: 11) ist ein grundlegendes Charakteristikum der komparativen Methode, dem auf diese Weise Rechnung getragen werden kann.

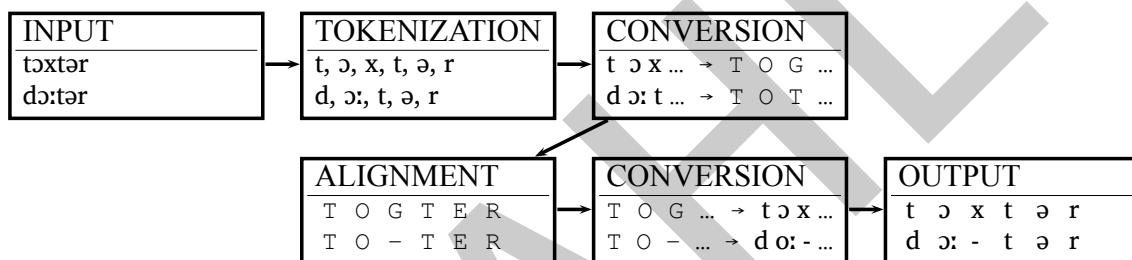
Die Tabelle zeigt die paarweisen Distanzen, die mit Hilfe der SCA-Alinierungsmethode (List 2012b) errechnet wurden, zwischen den oben erwähnten fünf unterschiedlichen Wörtern für "Zahn" auf. Inwiefern lässt sich anhand dieser Tabelle zeigen, wie wichtig es ist, sich beim Sprachvergleich auf multiple Evidenz zu stützen? Beachte dabei die in hellgrau und grau unterlegten Zellen.

2.4 LexStat



Die Graphik zeigt das Schema der LexStat-Methode zur automatischen Kognatenerkennung (List 2012a). Diese Methode nimmt die meisten der bisher erwähnten

Ideen zur automatischen Kognatenerkennung auf und vereint sie in einem Rahmenwerk, das recht nah den grundlegenden Ideen der traditionellen komparativen Methode folgt. LexStat basiert auf einer alinierungsbasierten Strategie zur automatischen Identifikation von Lautkorrespondenzen und bietet einen Lösungsansatz für das *spezielle Problem der Identifikation etymologisch verwandter Wörter*. Phonetische Alinierung kommt in zwei Schritten des Programms zum Tragen: als anfängliche Heuristik, die helfen soll, auch nur „verheißungsvolle Wortpaare“ in die engere Auswahl möglicher Kognaten zu nehmen, und als abschließendes Evaluierungsverfahren, mit dessen Hilfe die Distanzen zwischen den Wörtern der Eingabe berechnet werden, die dann mit Hilfe des Clusterverfahrens in eindeutige „Kognazitätspostulate“ überführt werden. Die phonetische Alinierung selbst basiert in der anfänglichen Heuristik auf dem Lautklassen-Verfahren. Es werden also nicht direkt phonetische Sequenzen aliniert, sondern durch Umwandlung der IPA-Zeichen in ein erweitertes Lautklassenschema erzeugte Lautklassensequenzen. Die folgende Abbildung verdeutlicht das zugrundeliegende Schema.



Obwohl die LexStat-Methode auf der automatischen Identifikation regulärer Lautkorrespondenzen basiert, ist sie prinzipiell nicht in der Lage, zwischen Xenologen und Kognaten zu unterscheiden. Woran liegt das?

Literatur

- DerkSEN, R., comp. (2008). *Etymological dictionary of the Slavic inherited lexicon*. Leiden Indo-European Etymological Dictionary Series 4. Leiden and Boston: Brill.
- Klimov, G. A. (1990). *Osnovy lingvisticheskoy komparativistiki* [Foundations of comparative linguistics]. Moscow: Nauka.
- List, J.-M. (2012a). "LexStat. Automatic detection of cognates in multilingual wordlists". In: *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*. (Avignon, France, Apr. 23-24, 2012). Association for Computational Linguistics, 117-125.
- (2012b). "SCA. Phonetic alignment based on sound classes". In: *New directions in logic, language, and computation*. Ed. by M. Slavkovik and D. Lassiter. LNCS 7415. Berlin and Heidelberg: Springer, 32-51.
- Ross, M. D. and M. Durie (1996). "Introduction". In: *The comparative method reviewed: Regularity and irregularity in language change*. Ed. by M. Durie. New York: Oxford University Press, 3-38.
- Sturtevant, E. H. (1920). *The pronunciation of Greek and Latin*. Chicago: University of Chicago Press. Internet Archive: pronunciationgr00unknngog.

Entlehnungserkennung

1 Allgemeines zur Entlehnung

Entlehnung (engl. *borrowing*) bezeichnet in einem engen Sinne, die Übertragung von Wörtern einer Sprache in eine andere Sprache.

Narrowly, the transfer of a word from one language into a second language, as a result of some kind of contact [...] between speakers of the two. (Trask 2000: 44)

In einem weiteren Sinne bezeichnet Entlehnung auch die allgemeine Übertragung linguistischer Merkmale von einer Sprache in eine andere.

Broadly, the transfer of linguistic features of any kind from one language to another as a result of contact. (ibid.)

Nenne je ein Beispiel für Entlehnung im engeren und im weiteren Sinne des Wortes.

1.1 Terminologisches

Der Terminus *Entlehnung* bezeichnet meist einen konkreten Prozess. Alternativ kann auch das etwas allgemeinere Wort *Übertragung (transfer)* verwendet werden. Entlehnung im engeren Sinne würde dann als *lexikalische Übertragung (lexical transfer)* bezeichnet. Soll auf das generelle Phänomen verwiesen werden, bietet sich der von Weinreich (1953[1974]) verwendete Terminus *lexikalische Interferenz (lexical interference)* an.

Im Folgenden betrachten wir Entlehnung vereinfacht als Übertragung von linguistischem Material von einer in eine andere Sprache. Die Sprache, die Material liefert, nennen wir die *Donorsprache (donor language)*, die Sprache, die Material aufnimmt, nennen wir die *Empfängersprache (recipient language)*. Das linguistische Material, das übertragen wird, nennen wir *Ursprungsmaterial (source material)*, wenn auf die Form, in der das Material in der Donorsprache vorliegt, verwiesen wird, und *Zielmaterial (target material)*, wenn auf die Form des Materials in der Empfängersprache verwiesen wird.

Schön und gut, aber was ist, wenn wir nun in einem bestimmten Fall nicht wissen, welche Sprache die Donorsprache und welche die Empfängersprache ist? Wie sollen wir sie dann nennen? Und wie sollen wir das Ursprungs- und das Zielmaterial dann nennen? Moment mal, hatten wir darüber nicht schon früher geredet? Das war doch so...

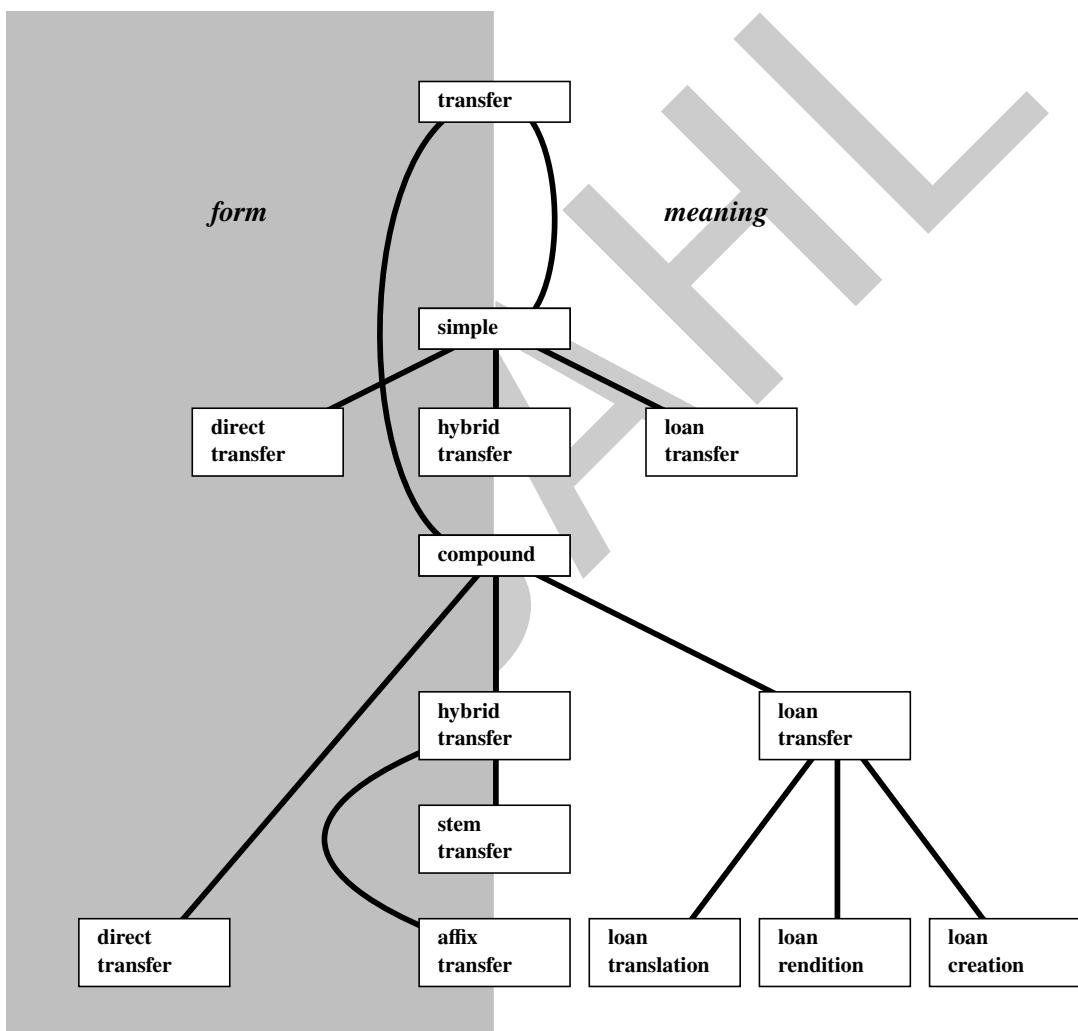
1.2 Mechanismen

Wenn wir im Sinne von Saussure von einem bilateralen Zeichenmodell ausgehen, so können grob zwei Fälle lexikalischer Übertragung unterschieden werden, nämlich die Fälle der *direkten Übertragung (direct transfer)*, in denen ein Wort als Ganzes übertragen wird, und die Fälle der *indirekten Übertragung* (auch *semantische Entlehnung, semantic borrowing*), in denen die Bedeutung eines in der Empfängersprache bereits vorhandenen Wortes erweitert oder umgestaltet wird, um

die Form-Bedeutungs-Einheit des Ursprungswortes in der Donorsprache nachzubilden. In Bezug auf einfache Wörter sind diese Fälle mehr oder weniger eindeutig auseinanderzuhalten, im Falle von Komposita können sich jedoch Mischformen ergeben.

The ways in which one vocabulary can interfere with another are various. Given two languages, *A* and *B*, morphemes may be transferred from *A* into *B*, or *B*-morphemes may be used in new designative functions on the model of *A*-morphemes with whose content they are identified; finally, in the case of compound lexical elements, both processes may be combined. (Weinreich 1953[1974]: 47)

Basierend auf Weinreich (ibid.: 47-62) können diese Prozesse in dem folgenden Schema dargestellt werden:



Die wichtigsten Prozesse mit ihren entsprechenden deutschen Termini sind dabei:

- **Übertragung einfacher Wörter**

- **Übertragung** (*transfer*): Direkte Übertragung von Form und Bedeutung eines sprachlichen Zeichens von der Donor- in die Empfängersprache.
- **Reproduktion** (*reproduction*): Erweiterung des Denotationsbereiches eines sprachlichen Zeichens der Empfängersprache in Anlehnung an die Bedeutung eines sprachlichen Zeichens in der Donorsprache.

- **Hybride Übertragung** (*hybrid transfer*): Erweiterung des Denotationsbereiches eines einem linguistischen Zeichen der Donorsprache homophonen sprachlichen Zeichens der Empfängersprache.

- **Übertragung komplexer Wörter**

- **Übertragung** (*transfer*): Direkte Übertragung von Form und Bedeutung eines komplexen sprachlichen Zeichens von der Donor- in die Empfängersprache.
- **Reproduktion** (*reproduction*): Nachbildung komplexer sprachlicher Zeichen der Donorsprache durch sprachliche Zeichen der Akzeptorsprache.
 - * **Lehnübersetzung** (*loan translation*): Exakte Nachbildung des komplexen sprachlichen Zeichens der Donorsprache in der Empfänger sprache.
 - * **Lehnübertragung** (*loan rendition*): Nachbildung des komplexen sprachlichen Zeichens in der Donorsprache, die das Zeichen nicht vollständig als Modell annimmt.
 - * **Lehnschöpfung** (*loan creation*): Bildung eines komplexen sprachlichen Zeichens in der Empfängersprache, dessen Denotationsbereich einem komplexen sprachlichen Zeichen der Donorsprache entspricht.
- **Hybride Übertragung** (*hybrid transfer*): Bildung eines komplexen sprachlichen Zeichens in der Akzeptorsprache, die teils auf direkter Übertragung, teils auf Reproduktion sprachlicher Zeichen der Donorsprache beruht.

Unten sind Beispiele für mögliche Resultate von lexikalischer Interferenz gegeben. Versuche, diese Fälle mit Hilfe des oben wiedergegebenen Klassifikationsschemas genauer zu charakterisieren.

No.	Wort	Hinweise	Mechan.
1	Russisch <i>futbol</i> „Fußball“	Russisch <i>mjač</i> „Ball“, <i>noga</i> „Fuß, Bein“	
2	Russisch <i>gol</i> „Tor“	Russisch <i>vorota</i> „Tor“	
3	Chinesisch <i>kěkǒukělè</i> „Coca Cola“	wörtl. „mundet - genießbar“	
4	Deutsch <i>Handy</i>	Englisch <i>cellphone</i> „Handy“	
5	Deutsch <i>Job</i>		
6	Russisch <i>bruderschaft</i> „Bruderschaft“	Russisch <i>brat</i> „Bruder“	
7	Deutsch <i>Rechner</i>	Englisch <i>Computer</i>	
8	Deutsch <i>Maus</i> (für Computer)	Englisch <i>mouse</i>	
9	Deutsch <i>Festplatte</i>	Englisch <i>hard disk</i>	
10	Deutsch <i>Wolkenkratzer</i>	Englisch <i>sky scraper</i>	
11	Chinesisch <i>xīngqīyī</i> „Montag“	wörtl. „Woche-eins“, <i>xīngqīèr</i> „Dienstag“, wörtl. „Woche-zwei“	
12	Deutsch <i>Fenster</i>	Latein <i>fenestra</i> „Fenster“	
13	Deutsch <i>Kopf</i>	Englisch <i>cup</i> , Latain <i>cūpa</i> „Gefäß“	
14	Chinesisch <i>kù</i> „bitter, cool“		
15	Chinesisch <i>xǐnǎo</i> „einer Gehirnwäsche unterziehen“	wörtl. „waschen Gehirn“	

1.3 Prozesse

Mit der Übertragung allein enden die Prozesse der Entlehnung zumeist nicht. In der Empfängersprache schließen sich meist Folgeprozesse an. Aus semantischer Perspektive treten diese Folgeprozesse insbesondere dann auf, wenn die Bedeutung des Zielwortes bereits von einem indigenen Wort in der Empfängersprache ausgedrückt wird und somit zwei Wörter miteinander in „Konkurrenz“ treten. Weinreich (1953[1974]: 54) unterscheidet hier zwischen *Unschärfe* (*confusion*), *Schwund* (*disappearance*) und *Spezialisierung* (*specialisation*) als Folge von Entlehnung.

Except for loanwords with entirely new content, the transfer or reproduction of foreign words must affect the existing vocabulary in one of three ways: (1) confusion between the content of the new and old word; (2) disappearance of the old words; (3) survival of both the new and old word, with a specialisation in content. (ibid.)

Aus Perspektive der übertragenen Form kann man als Folge von lexikalischer Interferenz zwei weitere Prozesse feststellen, nämlich die *Nativisierung* (*nativization*) und die ihr entgegengesetzte *Hyperverfremdung* (*hyper-foreignization*, vgl. Hock and Joseph 1995[2009]: 257f). Nativisierung bezeichnet dabei die Anpassung der Ursprungsform an die phonotaktischen Gegebenheiten der Empfängersprache (ibid.: 247-257). Hyperverfremdung ist ein umgekehrter Prozess, der dazu führt, dass der fremde „Klang“ der Ursprungsform in der Zielform in einem derartigen Maße von den Sprechern zu bewahren versucht wird, dass er von ihr abweicht.

Nenne Beispiele für Nativisierung und Hyperverfremdung im Deutschen.

1.4 Zahlen

Wie groß ist der Einfluss von lexikalischer Interferenz auf die Sprachgeschichte? Wie häufig sind Fälle der Entlehnung? Da die Häufigkeit von Entlehnungsprozessen vorwiegend von soziokulturellen Faktoren abhängt, zu denen die soziokulturelle Situation, in der Sprachen gebraucht werden, die geographische Distanz zwischen interferierenden Sprachen und das Prestige spezifischer Sprachvarietäten innerhalb einer Sprachgemeinschaft gehören, können Entlehnungsprozesse unterschiedlich große Teile des Lexikons einer Sprache erfassen. Diese reichen vom einfachen Austausch bestimmter kultureller Begriffe bis hin zu einer massiven Ersetzung des ursprünglichen Vokabulars. In der World Loanword Database (Haspelmath and Tadmor 2009) wurde die Häufigkeit direkter Entlehnung in einer Probe von 1460 Glossen („Konzepte“), die in 41 verschiedene Sprachen übersetzt wurden, untersucht. Laut der Datenbank gibt es große Unterschiede zwischen den Entlehnungshäufigkeiten der einzelnen Sprachen. Diese reichen von 1.2% (Mandarin-Chinesisch) bis hin zu 67% (Selice Romani). Auf der englischen Wortliste wurden dabei 41% der Wörter als Entlehnungen identifiziert. Die Gründe für die unterschiedlichen Entlehnungshäufigkeiten sind sehr unterschiedlich. Im Falle des Chinesischen ist die Entlehnungsresistenz zum Beispiel auf die sehr restriktive phonologische und phonotaktische Struktur der Sprache zurückzuführen, die es den Sprechern stark erschwert, fremdes Formmaterial in ihre Sprache aufzunehmen. Aus diesem Grund entlehnen chinesische Sprecher neue Wörter vorwiegend indirekt, also semantisch. Da die Datenbank nur Fälle direkter Entlehnung auflistet, bleiben diese Fälle verborgen. Ein Beispiel ist das Wort *fēiqùláiqì* 飞去来器 „Bumerang“, das wohl kaum ein indigenes chinesisches Wort sein dürfte, obwohl der Bumerang selbst auch unabhängig von den australischen Ureinwohnern erfunden wurde. Dieses beruht auf einer Lehnprägung, wie die wörtliche Übersetzung des Wortes als „hin-zurück-flieg-Gerät“ deutlich macht.

Die niedrige Zahl an Entlehnungen im Chinesischen wurde durch dessen restriktive phonologische und phonotaktische Struktur erklärt. Gibt es weitere Gründe, die zu niedrigen (oder umgekehrt hohen) Entlehnungsraten führen können, und wenn ja, welche sind dies?

2 Traditionelle Methoden der Lehnworterkennung

Traditionell werden Lehnwörter im Rahmen der historischen Linguistik durch Sequenzvergleiche ermittelt. Entscheidend sind für das Gelingen dieses Unterfangens dabei zwei Punkte:

1. Lehnwörter müssen den Ursprungswörtern phonetisch ähneln, da ansonsten keine Ursprungswörter identifiziert werden könnten.
2. Wenn die genetische Verwandtschaft der interferierenden Sprachen erwiesen ist, muss die Ähnlichkeit zwischen Lehn- und Ursprungswort von den korrespondenzbasierten Ähnlichkeiten zwischen den kognaten Wörtern der interferierenden Sprachen abweichen, da es ansonsten ja plausibler wäre, die möglichen Lehnwörter und Ursprungswörter als Kognaten zu identifizieren.

In der folgenden Tabelle werden etymologisch verwandte Wörter im Deutschen und Englischen einander gegenübergestellt. Zusätzlich zu den neuhighdeutschen Wörtern werden auch die jeweils etymologisch verwandten Mittelhochdeutschen Formen angegeben. Was fällt auf, wenn man die Wortpaare miteinander vergleicht? Bei welchen Wortpaaren könnte es sich um Entlehnungen handeln?

No.	Deutsch	Englisch	Mittelhochd.	Notiz
1	Dach	thatch	dah	
2	Daumen	thumb	dūm	
3	Degen	thane	degan	
4	Dill	dill	tilli	
5	Ding	thing	ding	
6	drei	three	drī	
7	Dunst	dunst	tunst	
8	Durst	thirst	durst	
9	denken	think	denken	
10	Damm	dam	tam	
11	dumm	dumb	tumb	
12	Dieb	thief	diob	
13	Dorn	thorn	dorn	
14	dreschen	thresh	dreskan	
15	Drossel	throat	drozze	

3 Quantitative Methoden der Lehnworterkennung

Obwohl es einige neuere Ansätze in der historischen Linguistik gibt, in denen versucht wird, von den Stammbaum-basierten Analysen wegzukommen, sind tatsächliche Methoden zur Lehnworterkennung bisher sehr rar. Prinzipiell wäre es dabei möglich, wie im oben gezeigten Beispiel auch automatisch vorzugehen, also Lautkorrespondenzen zu stratifizieren und Unterschiede festzustellen. Diese Ansätze wurden bisher aber noch nicht genauer verfolgt, da bisher noch keine ausreichend guten Methoden zur Entdeckung regulärer Lautkorrespondenzen entwickelt wurden. Ein alternativer Ansatz, der bisher in einer veröffentlichten Studie verwendet wurde, kommt aus der Biologie und besteht im Vergleich von sogenannten phyletischen Mustern vor dem Hintergrund von Referenzbäumen. Dieser wird im Folgenden genauer vorgestellt (Nelson-Sathi et al. 2011).

Ist die Methode von McMahon et al. (2005) eine Methode zur Entdeckung von Entlehnungen?

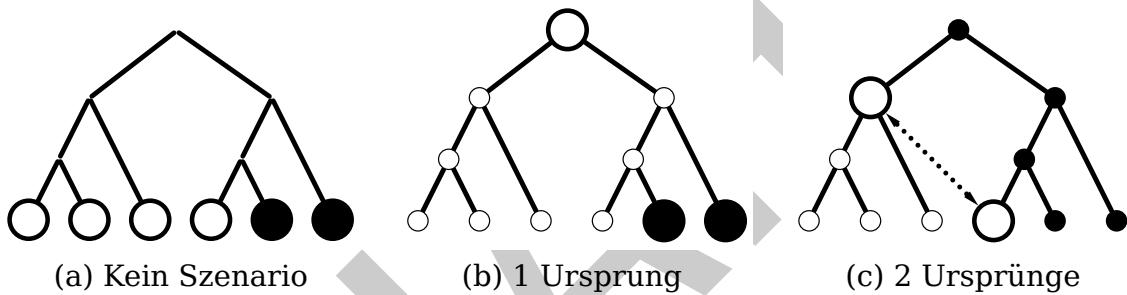
3.1 Gain-Loss Mapping

Die Gain-Loss-Mapping-Methode stammt aus der Biologie, sie kann jedoch auch mehr oder weniger direkt für Sprachen angewandt werden. Die Grundidee der Methode ist es, phyletische Muster (*phylogenetic patterns*) genauer zu untersuchen. Ein phyletisches Muster ist eine Matrix-Repräsentation von Kognatensets in einer bestimmten Menge von Sprachen. Dabei gibt die Matrix wieder, ob ein Kognatenset einen Reflex in der jeweiligen Sprache aufweist. Für eine bestimmte Sprache

werden also zwei verschiedene Zustände hinsichtlich eines Kognatensets unterschieden: *presence* (1) oder *absence* (0). Je nachdem, welche Kognatensets nun analysiert werden, kann man verschiedene Muster vergleichen, wie die folgende Darstellung zweier phylogenetischer Muster zeigt.

Language	Spanish	Portuguese	French	English	German	Swedish
“mountain”	montaña	montanha	montagne	mountain	Berg	berg
Pattern mountain	1	1	1	1	0	0
Pattern berg	0	0	0	0	1	1

Die Gain-Loss-Mapping-Methode versucht nun, mit Hilfe eines Referenzbaumes, zu erklären, wie bestimmte phyletische Muster entstanden sind. Dabei wird von einem sehr einfachen Prozess ausgegangen, in dessen Verlauf Sprachen Reflexe eines bestimmten Kognatensets entweder verlieren, oder erwerben können. Ein Gain-Loss-Szenario stellt dabei eine spezifische Hypothese bezüglich der Entwicklung eines phyletischen Musters auf. Dabei sind für verschiedene phyletische Muster durchaus verschiedene Gain-Loss-Szenarien denkbar, wie man an den folgenden Abbildungen (b) und (c) sehen kann, die beide das Muster für *mountain* „erklären“.

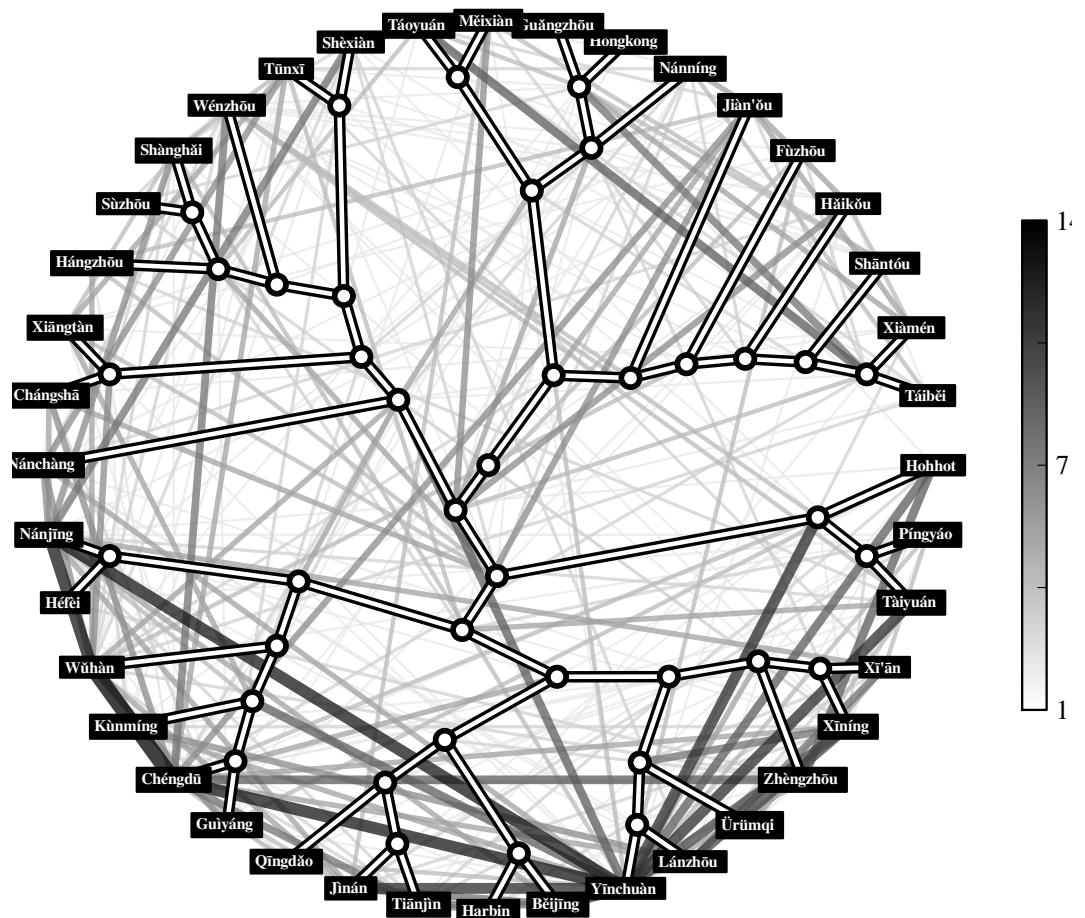


Ein Gain-Loss-Szenario unterscheidet nicht nur die beiden Zustände von Sprachen hinsichtlich eines Kognatensets, es zeigt auch an, wo sich Zustände im Verlauf der Entwicklung (repräsentiert durch den Referenzbaum) *verändern*, also *Übergänge* stattfinden. Dabei gibt es zwei mögliche Übergänge: Ursprung (*origin, gain*) und Schwund (*loss*). Wenn nun für ein Gain-Loss-Szenario zwei unabhängige Ursprünge angesetzt werden, dann ist es plausibel anzunehmen, dass die Ursprünge nicht unabhängig voneinander stattgefunden haben, sondern mitunter durch eine Entlehnung entstanden sind.

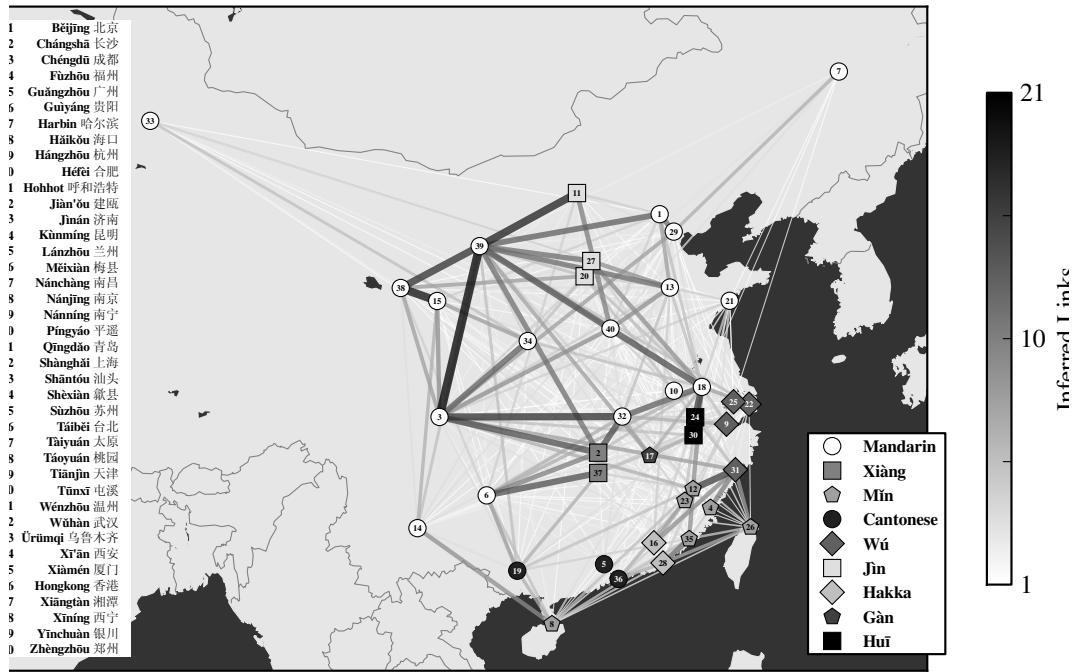
War das soweit verständlich? Wenn ja, ist die Annahme plausibel, und trifft sie auf das Beispilmuster *mountain* auch zu?

3.2 Beispiel

Wendet man diese Methoden auf große Datensätze an, kann man zeigen, welche Knoten in den Referenzbäumen am häufigsten unabhängig voneinander Ursprungsereignisse aufweisen. Das Ergebnis ist ein Netzwerk, in dem die vertikale Entwicklung durch den Referenzbaum dargestellt wird, und die horizontale durch „Interferenzkanten“. Das folgende Beispiel beruht auf lexikalischen Daten für 40 verschiedene chinesische Dialekte (Hou 2004).



In ähnlicher Weise können diese Ergebnisse dann auch verwendet werden, um mögliche areale Beziehungen zwischen Sprachen zu ermitteln, wie der geographische Plot der 40 chinesischen Dialekte zeigt.



Worin mögen die Schwächen in der Gain-Loss-Mapping-Methode liegen?

Literatur

- Haspelmath, M. and U. Tadmor, eds. (2009). *World Loanword Database*. URL: <http://wold.livingsources.org>.
- Hock, H. H. and B. D. Joseph (1995 [2009]). *Language history, language change and language relationship. An introduction to historical and comparative linguistics*. 2nd ed. Berlin and New York: Mouton de Gruyter.
- Hóu, J., ed. (2004). *Xiàndài Hànyǔ fāngyán yīnkù* [Phonological database of Chinese dialects] 现代汉语方言音库 [Phonological database of Chinese dialects]. Shanghai: Shànghǎi Jiàoyü.
- McMahon, A., P. Heggarty, R. McMahon, and N. Slaska (2005). "Swadesh sublists and the benefits of borrowing: An Andean case study". In: *Transactions of the Philological Society* 103, 147–170.
- Nelson-Sathi, S., J.-M. List, H. Geisler, H. Fangerau, R. D. Gray, W. Martin, and T. Dagan (2011). "Networks uncover hidden lexical borrowing in Indo-European language evolution". In: *Proc B* 278.1713, 1794–1803.
- Trask, R. L., comp. (2000). *The dictionary of historical and comparative linguistics*. Edinburgh: Edinburgh University Press.
- Weinreich, U. (1953 [1974]). *Languages in contact. With a preface by André Martinet*. 8th ed. The Hague and Paris: Mouton.

Quantitative Aspekte von Bedeutungswandel

1 Bedeutungsprobleme

1.1 Was ist Bedeutung?

Bedeutung für Sprachforscher

Die Aufgabe, die Bedeutung eines Wortes festzustellen, mag für den Logiker besagen: seinen begrifflichen Inhalt möglichst genau erfassen und abgrenzen, eine Definition liefern, die nach altbewährtem Rezept die übergeordnete Art und die unterscheidenden Merkmale angibt. Für den Sprachforscher bedeutet sie etwas anderes und in der Regel weit Schwierigeres, denn er ist schon hier, wie in allen weiteren Stadien seiner Arbeit, genötigt, neben den klar erfaßbaren logischen auch psychologische Faktoren von oft schwer greif- und wägbarer Natur in Betracht zu ziehen. (Sperber 1923: 1)

Ungeachtet der Tatsache, ob Sperber Recht hat, oder nicht. Wenn man an einen dieser unheimlich berühmten Sätze von Saussure denkt, sollte es eigentlich leicht sein, zu sagen, worin der Unterschied zwischen der Bedeutung eines Wortes in der Logik und der Bedeutung eines Wortes in der Linguistik besteht. Welcher Satz ist gemeint?

Wenn Bedeutung allein keinen Sinn mehr macht

Bedeutung ist nicht alles...

„Worte sind Zeichen für Begriffe.“ Dieser Satz bedarf – wie schon dargelegt – in mehr als einer Hinsicht der Einschränkung. Aber auch wenn man den Sinn des Wortes „Begriff“ so weit faßt wie nur irgend möglich, giebt der Satz keinesfalls eine erschöpfende Definition. Worte sind noch anderes und mehr als Zeichen für Begriffe. Sie enthalten Werthe, die nichts mit dem zu schaffen haben, was wir bisher an den Wortbedeutungen beachtet und untersucht haben; Werthe, auf denen gerade die feinsten Wirkungen der Sprache beruhen. Es empfieilt sich, sie von dem gewöhnlichen, dem begrifflichen Wortsinne abzusondern und ihm gegenüber zu stellen. (Erdmann 1900: 78)

Sinn macht Bedeutung

Es liegt nun nahe, mit einem Zeichen (Namen, Wortverbindung, Schriftzeichen) außer dem Bezeichneten, was die Bedeutung des Zeichens heißen möge, noch das verbunden zu denken, was ich den Sinn des Zeichens nennen möchte, worin die Art des Gegebenseins enthalten ist. Es würde danach in unserm Beispiele zwar die Bedeutung der Ausdrücke „der Schnittpunkt von a und b“ und „der Schnittpunkt von b und c“ dieselbe sein, aber nicht ihr Sinn. Es würde die Bedeutung von „Abendstern“ und „Morgenstern“ dieselbe sein, aber nicht der Sinn. (Frege 1892: 26f)

Bedeutung hat Hintersinn

Um eine bequeme Verständigung zu erzielen, wird es sich empfehlen, wie ich schon andeutete, scharf zwischen dem begrifflichen Inhalt und der Gesamtbedeutung des Wortes zu unterscheiden; zwischen dem begrifflichen Inhalt, der alle objectiven Merkmale einschließt, und der allgemeinen Wortbedeutung, die außer dem Begriff noch alle anderen Werthe enthält, die das Wort zum Ausdruck bringt. Diese Werthe sondere ich also von der Wortbedeutung ab, stelle sie dem Begriff gegenüber und fixiere sie sprachlich als „Nebensinn“ und „Ge-fühlswerth“ (Stimmungsgehalt). Nach dieser Auffassungs- und Ausdrucksweise

ist es dann eindeutig zu sagen, dass Leu und Löwe einerseits und Hose und Beinkleid andererseits Worte von verschiedener Bedeutung aber gleichem begrifflichen Inhalt seien. (Erdmann 1900: 80)

Die oben wiedergegebenen Zitate stehen repräsentativ für typische Debatten, die in der Semantik geführt wurden und immer noch geführt werden. Welche Debatten verbergen sich hinter den letzten beiden Zitaten und mit welchen Termini würde man das, was dort gesagt wird, heute bezeichnen?

1.2 Mehrdeutigkeit

Zunächst gehören hierher alle die Fälle, in denen die lautliche Übereinstimmung bei Verschiedenheit der Bedeutung nur auf Zufall beruht, wie bei nhd. *acht* = *diligentia* [Aufmerksamkeit] - *proscriptio* [Bann]- *octo* [acht]. [...] lautlich besteht [...] Identität, und derjenige, welcher einen solchen Lautcomplex ausser Zusammenhang aussprechen hört, hat kein Mittel zu erkennen, welche von den verschiedenen damit verknüpften Bedeutungen der Sprechende im Sinne hat. [...] Wirkliche Mehrheit von Bedeutungen muss man aber auch in sehr vielen Fällen anerkennen, wo nicht bloss lautliche, sondern auch etymologische Identität besteht. Man vergleiche z.B. nhd. *fuchs* *vulpes* - Pferd von fuchsig Farbe - rothaariger Mensch - schlauer Mensch - Goldstück - Student im ersten Semester, *boc* *hircus* - Bock der Kutsche - Fehler, *futter* *pabulum* - Überzug oder Unterzug, [...]. In den meisten der angeführten Fälle ist es ohne geschichtliche Studien überhaupt nicht möglich den ursprünglichen Zusammenhang zwischen den einzelnen Bedeutungen zu erkennen, und dieselben verhalten sich dann gar nicht anders zu einander, als wenn die lautliche Identität nur zufällig wäre. (Paul 1880[1886]: 68)

Die folgenden Beispielsätze sind mehrdeutig, wobei die Mehrdeutigkeit jeweils auf die Mehrdeutigkeit eines Lexems zurückzuführen ist. Identifiziere das mehrdeutige Wort in den Sätzen und charakterisiere es mit Hilfe der Termini "Polysemie", "Homonymie", "Homophonie", "Homographie" und "Vagheit".

1. ER SCHEINT DAS HINDERNIS UMFahren ZU WOLLEN.
2. ER IST AN DER RUHR ERKRANKT.
3. HAST DU DIE ARME GESEHEN?
4. HAST DU DIE FLIEGEN GESEHEN?
5. ICH HAB EIN NEUES SCHLOSS.
6. ER HATTE LEIDER KEINE STIMME MEHR.
7. MEIN OPA WAR SEHR STRENG.
8. DAS WIRD ZU LANG.
9. WAS FÜR EIN DUMMER AUGUST.
10. e:ə hat aŋst di: fələ tsu: veːli:ən

2 Bedeutungswandel

2.1 Bedeutungswandel und Lautwandel

Während der lautwandel durch eine widerholte unterschiebung von etwas unmerklich verschiedenem zu stande kommt, wobei also das alte untergeht, zugleich mit der entstehung des neuen, ist beim bedeutungswandel die erhaltung des alten durch die entstehung des neuen nicht ausgeschlossen. Er besteht immer in einer erweiterung oder einer verengung des umfangs der bedeutung, denen eine verarmung oder bereicherung des inhalts entspricht. Erst durch die aufeinanderfolge von erweiterung und verengung kann eine von der ursprünglichen völlig verschiedene bedeutung sich bilden. (Paul 1880[1886]: 66)

Wir haben bereits in einer früheren Sitzung von den grundlegenden Unterschieden zwischen Bedeutungs- und Lautwandel gesprochen. Worin bestehen diese? Welches waren die Schlagworte, mit denen die unterschiedlichen Prozesse beschrieben wurden?

2.2 Bedeutungswandel und lexikalischer Wandel

Wenn es keinen semantischen Wandel gäbe, würden sich die Lexika der Sprachen der Welt nicht ändern und in allen Zeiten gleich bleiben. Wörter würden zwar ihre Form ändern, es würde aber immer eine ungebrochene Tradition identischer Denotation geben. Da dies nicht so ist, wandeln sich die Lexika unserer Sprachen ständig. Wörter gehen verloren, wenn die Sprecher sie nicht mehr verwenden, neue Wörter treten ins Lexikon ein, wenn neue Konzepte geschaffen werden, sei es durch Entlehnung oder durch Neuschöpfung. Diese Prozesse des *Worterwerbs* (*word gain*) und des *Wortverlustes* (*word loss*) sind sehr häufig und können zuweilen sogar direkt durch die Sprecher einer Sprache bemerkt werden.

Der wichtigste Prozess des lexikalischen Wandels ist der Prozess der *lexikalischen Ersetzung* (*lexical replacement*). Lexikalische Ersetzung meint den Prozess, durch den ein Wort *A*, das normalerweise verwendet wird, um eine Bedeutung *x* auszudrücken, diese Bedeutung nicht weiter ausdrückt, da zur selben Zeit ein Wort *B*, das zuvor die Bedeutung *y* ausdrückte, fortan verwendet wird, um die Bedeutung *x* auszudrücken.

Nenne ein Beispiel für den Prozess der lexikalischen Ersetzung. Worin besteht der Unterschied dieses Prozesses zum semantischen Wandel?

2.3 Das ewig Abstrakte zieht uns hinab

Sehr häufig entdeckt man, daß einem Wort bloß deshalb ein bedeutungswandel zugeschrieben wird, weil seine *tatsächliche* Bedeutung sich nicht oder nicht völlig mit derjenigen deckt, die der etymologische Wortsinn anzudeuten scheint. Z. B. liegt in dem Wort *Schneider* kein Hinweis darauf, daß das, was der so bezeichnete Handwerker schneidet, gerade zu Kleidern bestimmte Stoffe sein müssen. Es sieht vielmehr ganz so aus, als habe hier ein Wort, das ursprünglich ganz allgemein „Mensch, welcher (etwas beliebiges) schneidet“ bedeutete, sekundär eine Einschränkung seines Begriffsumfanges erlitten, die zur Entwicklung seines heutigen engbegrenzten Sinnes führte. Oder: *Gewehr* bedeutet etymologisch nicht „Handfeuerwaffe von einem bestimmten Typus“ sondern „irgendetwas, womit man sich wehrt“ usw. ins Unendliche. Schon Paul hat mit Recht darauf hingewiesen, daß in solchen Fällen immer erst untersucht werden muss, ob die durch die Etymologie angedeutete allgemeinere Bedeutung jemals bestanden

hat, oder ob nicht das betreffende Wort vom ersten Augenblick an nur die engere Bedeutung besessen hat, eine Einschränkung seines Begriffsumfanges also gar nicht erfolgt ist. [...] Die erste Bedingung für die Ansetzung eines Bedeutungsübergangs ist doch offenbar, daß mindestens zwei Bedeutungen eine ursprüngliche und eine abgeleitete, im Sprachbewußtsein wirklich vorhanden sind oder waren. Bevor man von einem Bedeutungswandel spricht, muß man sich daher immer zuerst überzeugen, daß nicht die eine dieser Bedeutungen auf einem unbefugten Rückschluß aus dem buchstäblichen Sinn des Wortes beruht. (Sperber 1923: 11f)

Wow, wieder ein langes Zitat. Scheint aber wichtig zu sein, wenn der Dozent das in voller Länge abdruckt. Er will wahrscheinlich auf die Indizien hinaus, die erlauben, Bedeutungswandel zu rekonstruieren, also ...

3 Netzwerke aus sprachübergreifenden Polysemien

Wir könnten uns fragen, warum wir uns überhaupt mit Bedeutungswandel beschäftigen. Könnte man diese ganze Thematik nicht einfach ignorieren? Nun, von einem allgemeinen Standpunkt her kann sie natürlich ignoriert werden, und zwar dann, wenn wir die historische Linguistik generell als ein unsinniges Betreiben ablehnen. Wenn wir uns diese Ansicht jedoch nicht zu eigen machen, ist eine genauere Erforschung des semantischen Wandels von großer Bedeutung für die historische Linguistik: Sie hilft uns, wenn wir sie sinnvoll betreiben, im Rahmen der *semantischen Rekonstruktion* die ursprünglichen Bedeutungen von Wörtern zu rekonstruieren. Sie hilft uns, allgemeine Tendenzen semantischen Wandels zu ermitteln, falls es diese gibt. Und sie kann uns helfen, Kognaten zwischen genetisch verwandten Sprachen auch dort zu identifizieren, wo sich die Bedeutung von Formen gewandelt hat.

Moment mal, die Lexikostatistik beruht doch auf dieser Annahme, dass es Basisvokabular gibt, also eine Liste von stabilen Konzepten. In dem Zusammenhang wäre es doch auch wichtig, wenn wir mehr über Bedeutungswandel wüssten, denn dann könnten wir ja ...

3.1 Milch und Euter

Was haben „Milch“ und „Euter“ miteinander gemein? Konzeptuell stehen sie in enger Beziehung, da „Milch“ das ist, was aus dem „Euter“ rauskommt. Auch sprachlich können beide Wörter ähnlich sein, denn in viele Sprachen werden sie durch ein und dieselbe Wortform ausgedrückt. Historisch kann man sich vorstellen, dass die Beziehung zwischen „Milch“ und „Euter“ als semantischer Wandel auftritt. Im Standardchinesischen ist das Wort [niou³⁵nai²¹⁴] „Milch“ beispielsweise ein Kompositum aus [niou³⁵] „Kuh“ und [nai²¹⁴] „Milch“, das ursprünglich selbst „Euter“ und „Brust“ bedeutete.

Wie nennt man das Phänomen in der Linguistik, wenn konzeptuell verwandte Bedeutungen in einer Sprache durch dasselbe Wort ausgedrückt werden? Welches Problem ist mit diesem Terminus verbunden?

3.2 Polysemie als Indiz für semantischen Wandel

Der Terminus *Polysemie* wurde zuerst von Bréal (1897: 154) verwendet, der dabei explizit betonte, dass Polysemie eine Folge semantischen Wandels sei. Semantischer Wandel wird heute meist als ein Prozess angesehen, der eine polyseme Phase einschließt (Traugott and Dasher 2002). Dies heißt wiederum, das Fälle von Polysemie umgekehrt als Indiz für semantische Wandelprozesse genommen werden können. Natürlich kann man die Wandelprozesse nicht direkt aus Polysemien ablesen, man kann jedoch durch die Suche nach Polysemien diejenigen Konzepte identifizieren, die in konzeptueller Nähe zueinander stehen, weil sie häufig entweder den Ausgangs- oder den Endpunkt von Wandelprozessen darstellen. Die Idee, Polysemien explizit zu verwenden, um einen quantitativen Zugang zum semantischen Wandel zu bekommen, wurde in einer Reihe von Arbeiten von verschiedenen Autoren verfolgt (Croft et al. 2009, Cysouw 2010, François 2008, Perrin 2010, Steiner et al. 2011).

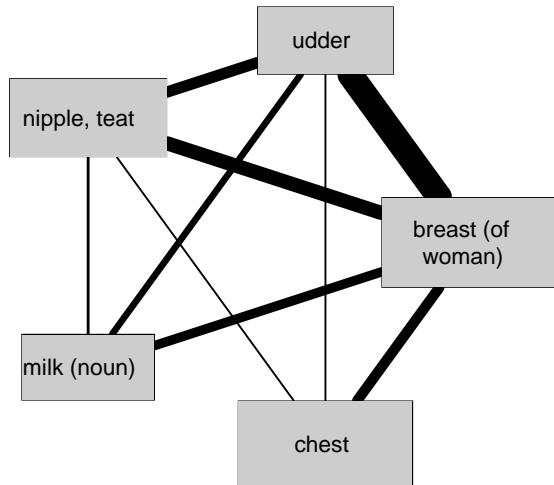
Diese Darstellung war jetzt ein bisschen unscharf. Was genau verbirgt sich hinter den Ansätzen, die Polysemie als Indiz für semantischen Wandel nehmen? Wie muss man sich das vorstellen?

3.3 Netzwerke

Wenn man genug Daten hat, ist es recht einfach, Konzeptnetzwerke aus sprachübergreifenden Polysemien zu rekonstruieren. Alles, was man braucht, sind semantisch alinierte Wortlisten, die in möglichst viele verschiedene Sprachen übersetzt wurden. Wenn wir uns beispielsweise eine Wortliste wie die folgende anschauen, dann fällt auf, dass in dieser Liste zwei „Polysemien“ auftauchen, nämlich Russisch *derevo* „tree + wood“ und Deutsch *Erde* „earth, land + ground, soil“. Wenn wir uns nun ein Netzwerk vorstellen, in dem alle Konzepte in unserer Wortliste Knoten sind, dann können wir Kanten zwischen allen Knoten ziehen, die in einer der Sprachen durch die gleiche Wortform ausgedrückt werden. Das Gewicht dieser Kanten kann ferner bestimmt werden, indem man zählt, wie oft ein bestimmtes Konzeptpaar durch dieselbe Wortform in den Sprachen auf der Wortliste ausgedrückt wird.

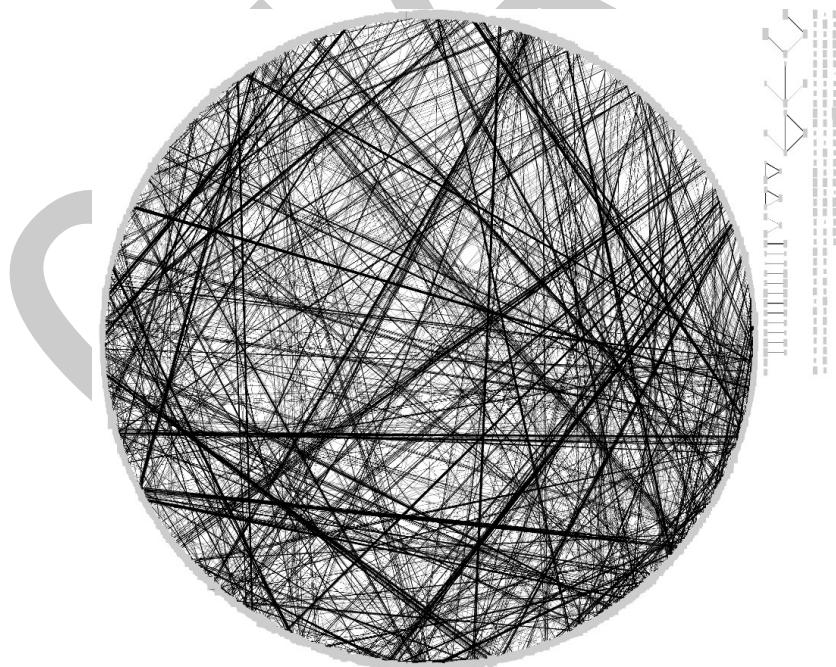
Key	Concept	Russian	German	...
1.1	world	mir, svet	Welt	...
1.21	earth, land	zemlja	Erde , Land	...
1.212	ground, soil	počva	Erde , Boden	...
1.420	tree	derevo	Baum	...
1.430	wood	derevo	Wald	...
...

Gewichte von Kanten werden in Netzwerken durch die Dicke der Kanten visualisiert. Somit ergibt sich für Konzepte rund um „Euter“ und „Milch“ beispielsweise unten abgebildete Netzwerk. Worin besteht das prinzipielle Problem des Wortlisten-Ansatzes, wenn man sich die traditionelle Definition von "Polysemie" ins Gedächtnis ruft?



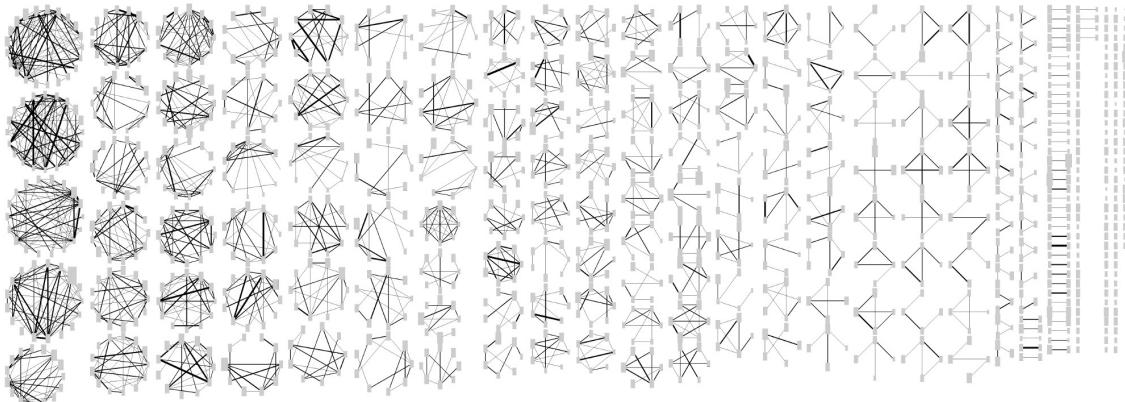
3.4 Netzwerkanalyse

Polysemienetzwerke allein bringen wenig, wenn man keine Möglichkeit hat, sie richtig zu analysieren. Denn Polysemienetzwerke können mitunter sehr komplex sein, so dass durch das bloße Betrachten dieser Netzwerke keine neuen Erkenntnisse gewonnen werden können. Die folgende Grafik zeigt beispielsweise ein Netzwerk, das aus der Analyse von 1289 Konzepten, übersetzt in 195 Sprachen aus 44 Sprachfamilien erstellt wurde **List2013**. Sehen kann man in einem solchen Netzwerk nicht viel.



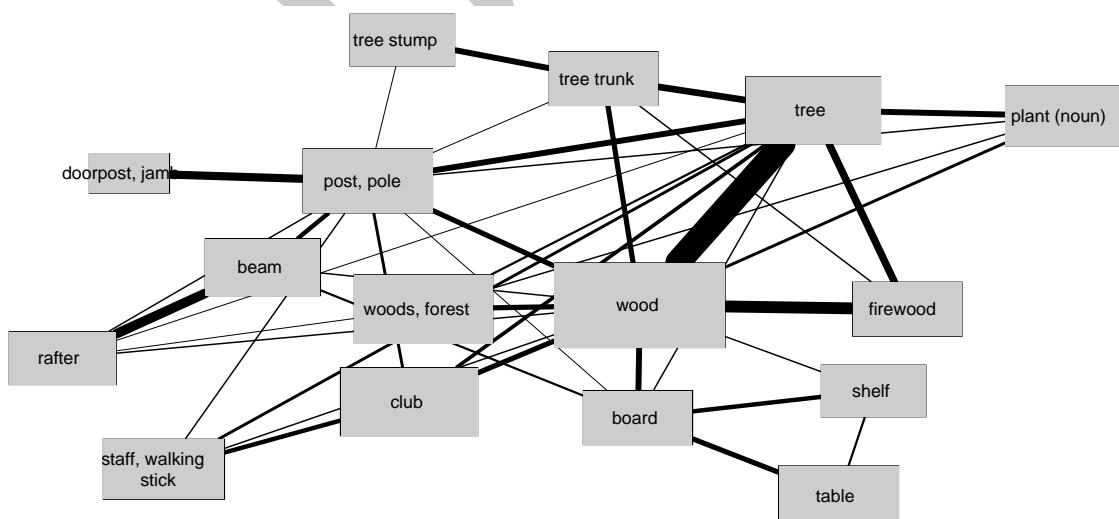
Netzwerkanalysen mit Hilfe von speziellen Algorithmen helfen im Falle von Polysemiedaten jedoch, die Struktur von Netzwerken genauer zu analysieren und spezifische Gruppen (sogenannte *Communities*) zu isolieren, die für die Struktur des Netzwerks bedeutsam sind. Communities sind definiert als „groups of vertices [Knoten] within which the connections are dense but between which they are sparser“ (Newman 2004: 4). Eine Analyse des Netzwerks aus **List2013** mit Hilfe des Algorithmus zur Community-Entdeckung von Girvan and Newman (2002), teilt

dieses beispielsweise in 337 Communities auf, wobei 104 relativ groß sind (5 und mehr Knoten), und einen Großteil der Konzepte abdecken (879 von 1289, 68%).



Das Schöne an dieser Analyse ist, dass die meisten dieser großen Communities Konzepte auf intuitiv einleuchtende, sinnvolle Art gruppieren. Eine Community, bspw. deckt Bedeutungen ab, die mit der Bedeckung von Körpern zu tun haben („Feder“, „Haar“, „Rinde“, etc.), eine weiter Community deckt Bedeutungen ab, die mit „Lernen“ zu tun haben („studieren“, „zählen“, „versuchen“), und eine Community deckt gruppiert Fahrzeuge („Kanu“, „Boot“, „Wagen“).

Unten ist eine Community aus dem Netzwerk abgebildet, in der Bedeutungen gruppiert werden, die den Konzepten "Baum" und "Holz" nahe stehen. Was lässt sich durch die Betrachtung des Netzwerkes erfahren? Welche Schlüsse kann man aus dem Netzwerk ziehen, welche nicht? Wie kann man sich die konkrete Verwendung von dieser Art von Netzwerken im Dienste von Algorithmen zur Kognatenerkennung vorstellen?



Literatur

Bréal, M (1897). *Essai de sémantique. Science des significations*. Paris: Hachette.
 Internet Archive: [essaidesmantiq00bruoft](http://www.archive.org/details/essaidesmantiq00bruoft); English translation: Bréal, M (1900).
Semantics. Studies in the science of meaning. Trans. from the French by H Cust.

- With an intro. by JP Postgate. New York: Henry Holt & Company. Internet Archive: semanticssstudie02postgoog.
- Croft, W, C Beckner, L Sutton, T Wilkins J.and Bhattacharya, and D Hruschka (2009). *Quantifying semantic shift for reconstructing language families*. Paper, presented at the conference "83rd Annual Meeting of the Linguistic Society of America" (San Francisco).
- Cysouw, M (2010). "Semantic maps as metrics on meaning". *Linguistic Discovery* 8.1, 70–95.
- Erdmann, KO (1900). *Die Bedeutung des Wortes*. Leipzig: Eduard Avenarius.
- François, A (2008). "Semantic maps and the typology of colexification: intertwining polysemous networks across languages". In: *From polysemy to semantic change*. Ed. by M Vanhove. Amsterdam: Benjamins, 163–215.
- Frege, G (1892). "Über Sinn und Bedeutung". *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50.
- Girvan, M and ME Newman (2002). "Community structure in social and biological networks". *Proceedings of the National Academy of Sciences of the United States of America* 99.12, 7821–7826.
- Newman, MEJ (2004). "Analysis of weighted networks". *Physical Review E* 70.5, 056131.
- Paul, H (1880 [1886]). *Principien der Sprachgeschichte*. 2nd ed. Halle: Max Niemeyer. prinzipiendersp01paulgoog: ia.
- Perrin, LM (2010). "Polysemous qualities and universal networks, invariance and diversity". *Linguistic Discovery* 8.1, 259–280.
- Sperber, H (1923). *Einführung in die Bedeutungslehre*. Bonn and Leipzig: Kurt Schroeder.
- Steiner, L, PF Stadler, and M Cysouw (2011). "A pipeline for computational historical linguistics". *Language Dynamics and Change* 1.1, 89–127.
- Traugott, EC and RB Dasher (2002). *Regularity in semantic change*. Cambridge: Cambridge University Press.