

Auswirkungen von Sprachkontaktphänomenen auf phylogenetische Algorithmen

Exposé zur Diplomarbeit

Mirko Hochmuth

17. November 2005

Betreuer: Prof. Ulf Leser
(HU Berlin, Inst. für Informatik)

Zeitraumen: 20. September 2005 - 20. März 2006

Motivation

Evolutionäre Zusammenhänge einer Menge von Taxa (Sprachen, Spezies, ...) werden zumeist durch *phylogenetische Bäume* repräsentiert und viele Verfahren wurden entwickelt, um baumartige Phylogenien zu (re-)konstruieren. Sie treffen jedoch alle einschränkende Annahmen, die die zugrundeliegenden Evolutionsprozesse zu stark vereinfachen. Insbesondere werden Effekte wie Sprachkontakt, Homoplasie, horizontaler Gentransfer oder Hybridisierung, die alle nicht in einer baumartigen Struktur dargestellt werden können, ausgeschlossen.

Diese Effekte erfordern eine Ausweitung des Baummodells auf *phylogenetische Netzwerke*. In letzter Zeit wurden bzw. werden vermehrt Verfahren zur Konstruktion solcher Netzwerke entwickelt (siehe z.Bsp. [1] & [2]).

Zielsetzung

Ziel dieser Diplomarbeit ist es, die Auswirkungen von Sprachwandelphänomenen, die in einem Baum nicht repräsentierbar sind, auf Algorithmen zur Erstellung von Sprachstammbäumen bzw. -netzwerken zu untersuchen.

Herangehensweise

Zunächst wird in Kooperation mit Linguisten ein Sprachwandelmodell aufgestellt und implementiert. Dieses soll sowohl baumkompatible Evolutionsprozesse wie systematischen Lautwandel als auch Sprachkontaktphänomene wie *Borrowing* berücksichtigen. Die Intensität jedes dieser Phänomene soll über einen Parameter regulierbar sein. Auf einer Art virtueller Landkarte werden ein oder mehrere Initialsprachen mit einem Vokabular von beispielsweise 200 Swadeshkonzepten platziert. Diese können sich dann auf neue Gebiete ausbreiten und dort unabhängig von der Muttersprache weiterentwickeln (und evtl. auch wieder austerben). Zwischen benachbarten Sprachen kann es zu Sprachkontakten kommen.

Das Modell muss anschliessend getestet und verfeinert werden, optimale Parameter müssen gefunden werden, die eine möglichst realitätsnahe Sprachevolution simulieren.

Mit diesem Modell sollen dann Simulationsphylogenien mit unterschiedlich starkem Grad an Netzwerkeffekten erzeugt werden. Diese müssen für die Weiterverarbeitung durch Phylogeniealgorithmen aufbereitet werden.

Um die Güte der verschiedenen Methoden beurteilen zu können müssen geeignete Abstandsmaße der rekonstruierten Phylogenien zur zugrundeliegenden, simulierten Phylogenie gefunden werden.

Literatur

- [1] Daniel H. Huson et al. Phylogenetic super-networks from partial trees. 2005.
- [2] Bernard M.E. Moret et al. Phylogenetic networks: Modeling, reconstructibility, and accuracy. *IEEE Transactions on computational Biology and Bioinformatics*, Vol. 1, No. 1, Jan-Mar, 2004.