

## 49. Jahrestagung des Instituts für Deutsche Sprache

12. bis 14. März 2013

### Lexikostatistik 2.0

Gerhard Jäger

In der Mitte des 20. Jahrhunderts gab es diverse Versuche, die Klassifikation von Sprachen mit Hilfe von Wortlisten, die dem Grundvokabular der betreffenden Sprachen entnommen sind, zu automatisieren. Eine besonders ambitionierte Variante dieses lexikostatistischen Programms war die Glottochronologie, die mit numerischen Methoden die Zeittiefe der Divergenz verwandter Sprachen zu bestimmen suchte. Diese Methoden wurden und werden in der historischen Sprachwissenschaft gemeinhin kritisch diskutiert, da sich die erzielten Ergebnisse häufig als fehlerhaft erwiesen. In den letzten Jahren erleben wir einen neuen Aufschwung lexikostatistischer und glottochronologischer Ansätze. Die Erfolgsaussichten dafür sind heute aus drei Gründen wesentlich besser als vor einem halben Jahrhundert:

- Uns stehen große Mengen an sprachvergleichenden Daten in elektronischer Form zur Verfügung,
- es gibt einfachen Zugang zu schneller Computer-Hardware, und
- die Bioinformatik hat für strukturell analoge Probleme ein mächtiges Instrumentarium an statistischen und algorithmischen Werkzeugen entwickelt.

In dem Vortrag werden die Grundannahmen der aktuellen computati-  
onellen historischen Linguistik und der computationellen Dialektometrie dis-  
kutiert und einige neue Resultate dargestellt. Daran schließen sich Über-  
legungen über das Potential statistischer und computationeller Methoden  
für die Sprachgeschichtsforschung an.