

I.2c Historische Sprachdatenbank ‚Simplex

Ein großes Problem bei dem Auffinden von Gesetzmäßigkeiten in historischen Sprachentwicklungen (und, damit verbunden, in Bezug auf die Organisation von Sprache im mentalen Lexikon des Menschen) ist das Verfolgen von (Laut-)entwicklungen über verschiedene Sprachstufen und Sprachen hinweg. Ein einfache Frage wie z.B. „Liste alle Wörter (zusammen mit ihren Entsprechungen im Deutschen und Niederländischen), die einen Kurzvokal im Altenglischen und Althochdeutschen hatten und einen Langvokal in Mittel- und Neuenglischen“ führt i.d.R. zu einer mehrwöchigen Sucharbeit, obwohl viele (auch etymologische) Lexika und philologische Werke inzwischen digitalisiert vorliegen (vgl. das Leidener *IEED-Projekt*). Hierzu sollen die Grundzüge einer Datenbankstruktur und eines Anfrageinterfaces entwickelt werden, das die vorhandenen Datenbestände zu durchsuchen gestattet und entsprechende Anfragen mit einer einfachen Eingabe ermöglicht, die nicht das Erlernen einer speziellen Anfragesprache erfordert. Ein wesentlicher Aspekt dieses Datenbanksystems ist es, dass keine relationale ‚Großdatenbank‘ aller vorhandenen Datenbestände geschaffen werden soll, sondern dass eine Art „Metacrawler“ die (vernetzten) Datenbestände durchsucht, dabei auch morphologische und semantische Bezüge auflösen kann (z.B. den Zusammenhang zwischen nd. ‚Frau‘ und ahd. ‚wib‘) und Bezüge über lautliche Ähnlichkeit (z.B. anord. ‚vifa‘) herstellen kann. Als Beispiel für die einfache Anfrageform kann das in Frankfurt entwickelte Interface zur Celex-Datenbank dienen, das auch zeigt, dass selbst die Suche nach Minimalpaaren in einem 500 000 Wörter umfassenden (nicht relationalen) Lexikon in wenigen Sekunden möglich ist. Die Entwicklung des Interfaces soll zugleich dazu dienen, die Zuverlässigkeit der vorhandenen Datenbestände zu überprüfen und den Weg für einen systematischen Ausbau der Datenbasis zu bahnen.