

Zwischen Datenbank und Wörterbuch im Dictionnaire étymologique de l'ancien français *S. Dörr, M. Kiwitt*

Der Dictionnaire étymologique de l'ancien français (DEAF) ist ein Grundlagenwörterbuch des Altfranzösischen, das den Zeitraum von 842 bis in die Mitte des 14. Jahrhunderts umfasst. Als historisch wichtigste Vernakularsprache des europäischen Mittelalters ist das Altfranzösische in vielen Tausend Texten überliefert, die nicht nur aus Frankreich, sondern beispielsweise auch aus England, Italien, Zypern und dem Heiligen Land stammen. Seine Wirkungsmacht und sein Prestige lassen sich darüber hinaus auch am Einfluss der altfranzösischen Literatur auf alloglotte Literaturtraditionen - darunter etwa die mittelhochdeutsche Klassik - ablesen.

Der DEAF wurde in den sechziger Jahren des letzten Jahrhunderts von Georges Straka, Kurt Baldinger und Jean-Denis Gendron in Québec initiiert und erscheint seit Beginn der neunziger Jahre in einem Publikationsrhythmus von etwa einem Faszikel mit 192 Spalten pro Jahr. 25 Faszikel, in denen die Buchstaben G, H, I, J und K behandelt werden, sind bisher erschienen. Alle verfügbaren Materialien - Wörterbücher, Glossare, literarische und nichtliterarische Texte, Sekundärliteratur - werden im DEAF kritisch ausgewertet; derzeit umfasst der DEAF-Fichier etwa 1,5 Millionen Zettel, die auf ca. 12 Millionen Belege verweisen.

Seit 2008 hat die DEAF-Redaktion in Zusammenarbeit mit dem Karlsruher Institut für Programmstrukturen und Datenorganisation ein elektronisches Redaktions- und Publikationssystem (DEAFél) entwickelt, das am 2. Juli 2010 der Öffentlichkeit vorgestellt wurde. Im Hinblick auf die verbleibende Projektlaufzeit wurde hierbei ein zweistufiger Ansatz verfolgt: die Alphabetstrecke D-F aus wissenschaftshistorischen Gründen das "schwarze Loch" der galloromanischen Lexikographie - wird weitestgehend in gewohnter Form lexikographisch erschlossen; die Ergebnisse werden sowohl in gedruckter Form als auch im Internet publiziert (DEAFplus). Parallel dazu werden sämtliche Materialien zu den Buchstaben A-C und L-Z lemmatisiert und mit einer groben semantischen Struktur versehen, um auch die für diese Alphabetbereiche bereits erarbeiteten Daten langfristig zu konservieren und für eine weitergehende lexikographische Erschließung zugänglich zu machen (DEAFpré). Kernstück des elektronischen Redaktionssystems ist eine MySQL-Datenbank; weiterhin werden die mannigfaltigen in den Artikeln vorliegenden Informationen (etwa Belege, Datierungen, grammatikalische Kategorien, Definitionen, Sprach- und Dialektangaben) mit einem komplexen XML-Markup semantisch ausgezeichnet.

Neben Zeitersparnissen insbesondere bei der Klassifizierung und Strukturierung der Materialien und komfortablen Suchfunktionen, die weitaus umfassendere Abfragemöglichkeiten als ein herkömmlicher Index erlauben, bietet die computergestützte Redaktion und Publikation erstmals die Möglichkeit, die vom DEAF erarbeiteten wissenschaftlichen Ressourcen in einem für die langfristige Archivierung und Bereitstellung geeigneten digitalen Format zu speichern. Im Einzelnen erscheint für die folgenden Materialien eine digitale Archivierung wünschenswert:

- Primärdaten: Das Ausgangsmaterial für die Redaktion der Wörterbuchartikel liegt zunächst in handgeschriebenen - im Hinblick auf ihren Informationsgehalt und ihre Lesbarkeit ausgesprochen heterogenen - Zetteln vor, die von Hilfskräften mittels einer

Eingabemaske in einer XML-Struktur erfasst werden. Da aufgrund der hohen anfallenden Datenmengen (mehrere Tausend Zettel pro Monat) keine durchgängige manuelle Fehlerkorrektur möglich ist, wurde für zentrale Elemente - insbesondere für die Sigel in den Literaturangaben - ein erweiterbares Regelsystem geschaffen, das derzeit etwa 10.400 Ersetzungsregeln umfasst und die automatische Korrektur von 83 bis 97% der jeweils in einer Buchstabenstrecke auftretenden sigelspezifischen Eingabefehler erlaubt.

- Bibliographische Daten: Das Complément bibliographique des DEAF bietet eine umfassende kritische Bibliographie zum Altfranzösischen und wird ständig erweitert und aktualisiert. Derzeit beinhaltet es 6.410 Literaturverweise mit Angaben zu 953 mittelalterlichen Autoren, 2.663 Datierungen und 1.672 Lokalisierungen. Die XML-Struktur der Bibliographie erlaubt es, die elektronisch erfassten Zettel beim Import in die Datenbank mit aktuellen bibliographischen Angaben zu versehen und lässt sich darüber hinaus problemlos für die Online-Publikation in ein HTML-Format überführen. Ausgehend von den in XML vorliegenden bibliographischen Daten wurde das Complément bibliographique darüber hinaus zuletzt 2007 in gedruckter Form veröffentlicht.
- Wörterbuchartikel (DEAF und DEAFplus): Die mithilfe des elektronischen Redaktionssystem erarbeiteten Artikel zu den Buchstaben D, E und F (DEAFplus), die zunächst in einer MySQL-Datenbank gespeichert werden, können in verschiedene Formate - Latex für den Druck, HTML für die Online-Publikation, XML für die Archivierung - exportiert werden. Für die bereits publizierten Artikel der Alphabetstrecke G bis K, die lediglich als Image-Dateien vorliegen, ist eine Retrodigitalisierung wünschenswert.
- Strukturierte Materialien (DEAFpré): Die lemmatisierten und mit einer groben semantischen Struktur versehenen Materialien zu den Alphabetstrecken A-C und L-Z, die im Rahmen der verbleibenden Laufzeit nicht vollständig bearbeitet werden können, können analog zu den Artikeln des DEAFplus aus der Datenbank in HTML und in XML exportiert werden.

Mit der konsequenten Speicherung aller im Redaktionsprozess erarbeiteten Daten - vom Zettel über die Bibliographie bis zum fertigen Wörterbuchartikel - in XML erfüllt der DEAF die in den Empfehlungen der AG Elektronisches Publizieren umrissenen Anforderungen an die Standardisierung der Daten und ist für die Langzeitarchivierung seiner Materialien in einem digitalen Repository bestens gerüstet. Als Desiderat wäre in erster Linie die Bereitstellung eines zentralen Servers, auf dem die anfallenden Daten langfristig sicher abgelegt werden können, zu nennen.