Dictionnaire du Moyen Français (DMF), version 2012 (DMF 2012), ATILF – CNRS & Université de Lorraine, <<http://www.atilf.fr/dmf>>.

1 Einleitung

Die unter der Adresse <<http://www.atilf.fr/dmf/>> zur Verfügung stehende Online-Ressource *Dictionnaire du moyen français* (DMF) ist ein digitales Wörterbuch und zugleich Korpus, welches das Französische zwischen 1330 und 1500 dokumentiert. Bei seinem Erscheinen im Jahr 2003 schloss es eine nicht unerhebliche Lücke in der Lexikographie des Französischen. Inzwischen hat es sich als wertvolles und praktisches Werkzeug etabliert, das aus der Forschung zum Mittelfranzösischen kaum noch wegzudenken ist.

Wörterbücher wie der *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^{ème} au XV^{ème} siècle* (Gdf), das *Altfranzösische Wörterbuch* (TL) und das *Französische etymologische Wörterbuch* (FEW) bleiben selbstverständlich wichtige Quellen zur Erschließung des Mittelfranzösischen (auf die im DMF regelmäßig verwiesen wird), sie weisen jedoch bezüglich ihrer lexikographischen Verfahren, der zugrundeliegenden Materialien oder auch im Hinblick auf den berücksichtigten Zeitraum eine andere Konzeption als der DMF auf. Als Epochenwörterbuch liegt außerdem der *Dictionnaire étymologique de l'ancien français* (DEAF) in Teilbänden vor; seine Fertigstellung ist allerdings erst in einigen Jahren zu erwarten.

Das Projekt des DMF wurde in den 1980er Jahren von Paul Imbs und Robert Martin angeregt. Die grundsätzlichen Entscheidungen hinsichtlich Struktur, Orientierung und Arbeitsplan fielen ebenfalls in den 1980er Jahren. Ein wichtiger Teil der Arbeiten fand während der 1990er Jahre im an der Universität Nancy 2 verankerten *Institut national de la langue française* statt. Dieses wurde 2001 mit anderen Instituten zu einer Forschungseinrichtung mit dem Namen *Analyse et traitement informatique de la langue française* (ATILF) zusammengefasst. Der DMF stellt eines der Leitprojekte dieser seit 2013 von Éva Buchi geleiteten Einrichtung dar und soll die Untersuchung des Mittelfranzösischen durch ein Nachschlagewerk unterstützen, das den *Trésor de la langue française* (TLF) hinsichtlich des berücksichtigten Zeitraums ergänzt und ihm bezüglich Qualität und Vollständigkeit entspricht.

2 Konzeption

Die Beschränkung auf den Zeitraum 1330–1500 hatte teils praktische, teils sprachhistorische Gründe (cf. Martin 2012, 7). Da das 16. Jh. schon durch den *Dictionnaire de la langue française du seizième siècle* (DLFS) abgedeckt ist und dieses Jahrhundert zudem aufgrund der Erweiterung des Wortschatzes durch Entlehnungen aus dem Altgriechischen, Italienischen und auch Spanischen, das Einsetzen metasprachlicher Reflexion und, nicht zuletzt, die Verbreitung des Buchdrucks bedeutende Neuerungen mit sich brachte, bot sich 1500 als obere Grenze an. 1330 als unteres Limit ist selbstverständlich weniger offensichtlich. Der ausschlaggebende Grund war wohl, dass die dem TL zugrunde liegenden Texte kaum über dieses Datum hinausreichen. Auch der DEAF setzt als oberes Limit des Altfranzösischen die Mitte des 14. Jh. an. Anzumerken ist, dass schon ab der 2007 erschienenen Version das Korpus des DMF elf Texte enthält, die zum Teil einige Jahrzehnte nach 1500 entstanden sind (z. B. die *Histoire d'un voyage en terre de Brésil* von Jean de Lery aus dem Jahr 1578).

Anders als die oben erwähnten Wörterbücher war der DMF von Anfang an als digitale Ressource angelegt, obwohl in den Anfangsphasen des Projekts auch die parallele Herausgabe einer gedruckten Version ins Auge gefasst wurde (auf die man letztendlich verzichtete). Es handelt sich somit nicht um die (Retro-)Digitalisierung eines schon existierenden Werks in gedruckter Form, wie im Falle des DEAF, DLFS, Gdf oder TLF, die auch online abgefragt werden können, oder des TL, der 2002 vom Franz Steiner Verlag als DVD und CD-ROM publiziert wurde.

Der DMF erscheint, auch dank seiner digitalen Form, in regelmäßig überarbeiteten und erweiterten Versionen. Martin (2012, 3) spricht in diesem Zusammenhang von einer «lexicographie évolutive». Seit Juli 2012 ist eine neue Version verfügbar. Jede der vorhergehenden Ausgaben (DMF1 aus dem Jahr 2003, DMF2 aus dem Jahr 2007, DMF2009 und DMF2010) ist in sich abgeschlossen und kann bei Bedarf auf der Einstiegsseite des Wörterbuchprojektes aufgerufen werden. Ohne spezifische Vorauswahl wird der Benutzer direkt zur neuesten Ausgabe weitergeleitet. Die Abrufbarkeit aller vorhergehenden Versionen gründet auf dem dezidierten Wunsch der Herausgeber (cf. Martin 2012, 6), wie bei einem Buch bibliographische Verweise auf zeitlich klar definierte Ausgaben zu ermöglichen. Es handelt sich um eine beispielhafte Entscheidung, die bei vergleichbaren Wörterbuch- oder Korpusprojekten bisher kaum getroffen wurde. Die Versionen unterscheiden sich quantitativ und qualitativ. Für die meisten Benutzer ist sicher die Anzahl der Lemmata und der zur Verfügung stehenden Texte von Interesse. Die 32.779 Lemmata des DMF1 sind auf nunmehr 62.371 Lemmata im DMF2012 angewachsen. Die Anzahl der Texte hat sich allerdings ab dem DMF2 nicht mehr verändert und umfasst 242 Texte (davon elf aus der Zeit nach 1500).

In seiner Grundstruktur besteht der DMF aus drei miteinander verknüpften Datenbanken: das Wörterbuch, die spezifischen Wortschatzsammlungen und das Textkorpus. Die prinzipiellen Abfragemöglichkeiten orientieren sich an diesen Grundkomponenten. So kann man beispielsweise nach dem Verb *cuidar* im Wörterbuch, in einer der 21 Wortschatzsammlungen oder direkt in den Texten suchen. Die in den drei Komponenten enthaltenen Daten decken sich zum Großteil, aber nicht vollkommen. Vor allem umfasst das Wörterbuch zahlreiche Einträge, die weder in den Wortschatzsammlungen noch in den Texten zu finden sind, da sie aus zusätzlichen Quellen stammen. Auch die Wortschatzsammlungen enthalten viele Einträge, die nicht aus dem Textkorpus kommen.

Das Wörterbuch ist prinzipiell korpusbasiert, was bedingt, dass ausnahmslos alle Belege registriert werden, auch *hapax legomena* und Ähnliches. Keinen Eintrag haben hingegen Wörter aus anderen Sprachen (auch solche aus dem Okzitanischen, Frankoprovenzalischen und Anglonormannischen) gefunden. Lateinische Wörter sind nur dann registriert, wenn sie sich syntaktisch wie französische verhalten. Homonyme sind nach dem etymologischen Kriterium gruppiert, d.h.

zwei gleiche Formen mit unterschiedlicher Herkunft führen zu getrennten Einträgen.

Die 21 Wortschatzsammlungen, die übrigens den Ausgangspunkt und die ursprüngliche Grundlage des DMF darstellen (cf. Martin 2012, 6), berücksichtigen ganz unterschiedliche Gesichtspunkte. In einigen wird die Lexik eines Autors, etwa von Guillaume de Machaut oder Jean Froissart, erfasst. In anderen wird der Wortschatz eines einzelnen Werks, z.B. die *Cent nouvelles nouvelles*, analysiert. Im *Lexique de la langue scientifique* kann man Termini der Astrologie, Mathematik, Medizin und anderer Wissenschaften des 14. und 15. Jh. erforschen. Die Sammlung *Lexique de la littérature didactique* registriert den Wortschatz von Autoren wie Simon de Phares, Philippe de Mézières und Evrart de Conty.

Die 242 Texte des Korpus stammen mit wenigen Ausnahmen aus Druckausgaben, oft aus kritischen Editionen. Sie umfassen insgesamt 6.900.000 Wörter, was für ein Korpus mit historischen Dokumenten beträchtlich ist. Positiv fällt auf, dass hinsichtlich Genre und Texttypen eine beachtliche Streuung erreicht werden konnte. Natürlich nehmen die Werke bekannter Autoren wie Alain Chartier, Philippe de Commines, Guillaume de Machaut, Clément Marot, Nicole Oresme, Charles d'Orléans, Christine de Pizan und François Villon einen wichtigen Platz im Korpus ein. Viele dieser Autoren sind jedoch mit unterschiedlichen Genres vertreten. So liegen beispielsweise von Eustache Dechamps sowohl Werke in Versform (*Le miroir de mariage*) als auch in Prosa (*L'art de dictier*) vor. Jean Froissart ist nicht nur durch seine Chroniken sondern auch durch seine Lyrik (*La prison amoureuse*) vertreten. Zudem enthält das Korpus 95 anonyme Texte unterschiedlichster Natur, wie etwa die *Chartes et documents de l'abbaye de Saint-Magloire* oder den *Registre criminel du Châtelet de Paris*. Auch Reiseberichte, Memoiren, Predigten, Briefe und königliche Verordnungen haben Eingang in das Korpus gefunden. Leider sind bei einigen Texten (Werke von Guillaume de Machaut und Charles d'Orléans) die bibliografischen Angaben unvollständig.

3 Benutzeroberfläche

Die Benutzeroberfläche ist hinsichtlich Aufbau und Grafik zwar nicht auf dem aktuellen Stand des Webdesigns, aber doch einigermaßen ansprechend und funktionell. Der erste Kontakt mit dem Internetauftritt des Projekts desorientiert allerdings, denn es gibt irritierenderweise zwei Einstiegsseiten (*Accueil*), eine für den DMF und eine für die Ausgabe des DMF2012, die teilweise die gleichen Informationen enthalten; von ersterer führt kein direkter Link zu letzterer, von letzterer gibt es keinen Link zurück auf die des DMF. Man ist daher gezwungen, für diese Seiten gesonderte Lesezeichen anzulegen.

Für Recherchen startet man am besten sofort von der Einstiegsseite des DMF2012, da dort ein Auswahlmenü eine Übersicht über die verschiedenen Suchfunktionen bietet. Diese sind äußerst vielfältig. Die Abfragen im Wörterbuch können von unterschiedlichen Erstinformationen ausgehen: Lemma, Etymon, Wort aus einem Phraseologismus oder Eintrag in einem der zitierten Wörterbücher (Gdf, TL, FEW und weitere). Die Suchen werden durch praktische Assistenten unterstützt. Bei Eingabe der Form *cuit* in die Suchmaske wird man z. B. dank eines angeschlossenen Lemmatisierers auf drei Verben hingewiesen (*cuiden*, *cuire* und *quitter*) und zu den entsprechenden Wörterbucheinträgen geleitet. Die im DMF verwendete Bibliografie kann ebenfalls durchsucht werden.

Die Suchfunktionen in den Wortschatzsammlungen entsprechen im Prinzip den gerade erwähnten, nur dass zu Beginn einer Recherche einer oder mehrere Wortschätze auszuwählen sind, auf den oder die die Abfragen angewendet werden sollen. Bemerkenswert ist, dass man hier mit Hilfe von einzelnen Ausgangswörtern nicht nur nach Phraseologismen sondern auch nach 177 registrierten Sprichwörtern suchen kann.

Es ist anzunehmen, dass viele Benutzer den DMF weniger als Wörterbuch denn als Korpus verwenden. Als Beispiel sei hier Féron (2008) erwähnt. Auch hier wählt man zu Beginn der Recherche einen oder mehrere Texte als Suchbereich aus. Anschließend kann man die ausgewählten Texte nach Wortformen oder Lemmata durchsuchen. Auch komplexere Abfragen sind möglich: Man kann die Suche auf bestimmte Positionen im Satz (Anfang, Mitte, Ende) einschränken, Groß- und Kleinschreibung und Diakritika mitberücksichtigen sowie Abfragen mit regulären Ausdrücken (*regular expressions*) konstruieren. Bei der Suche nach Wortformen muss man selbstverständlich beachten, dass diese unterschiedlich verschriftet sein können (z. B. *je cuide*, *je cuyde*, *je quide* oder *je cuit*; *ce me samble*, *ce me semble* oder *se me semble*). Es besteht auch die Möglichkeit, nach Kookkurenzen von Wortformen oder Lemmata zu suchen. Die Ergebnisse werden wie üblich in der Form *keyword in context* (KWIC) dargestellt. Der Kontext kann nach oben und unten etwas erweitert werden. Wie bei vielen Korpusprojekten dieser Art ist es aufgrund von Urheberrechten jedoch nicht möglich, einen ganzen Text anzuzeigen und durchzulesen.

4 Fazit

Es handelt sich um eine ausgereifte Ressource, zu der es in lexikographischer und korpuslinguistischer Hinsicht nichts Gleichwertiges gibt und die noch dazu kostenfrei und ohne Registrierung zugänglich ist. Abgesehen von der Möglichkeit, in Zukunft den einen oder anderen Text hinzuzufügen, gibt es Spielraum

für einzelne Verbesserungen in den Abfragemöglichkeiten und in der Darstellung. Das Lemmatisierungsprogramm analysiert Wortformen und weist sie ihren jeweiligen Lemmata zu. So kann man beispielsweise nach der Form *cuit* und nach dem Lemma *cuidier* suchen. Eine Suche nach *cuit* als Form des Lemmas *cuidier* ist allerdings nicht möglich. Abgesehen davon ist das Textkorpus selbst (noch?) nicht lemmatisiert. Eine Suche nach den Formen des Lemmas *cuidier* liefert deshalb notgedrungen auch *cuit* von *cuire* und *quitter*.

5 Bibliografie

- DEAF = Baldinger, Kurt (ed.), *Dictionnaire étymologique de l'ancien français*, Tübingen/Berlin, Niemeyer/De Gruyter, 1974ss. (Bände G–K), <<http://www.deaf-page.de>>.
- DLFS = Huguet, Edmond, *Dictionnaire de la langue française du seizième siècle*, Paris, Champion/Didier, 1925–1967, <<http://www.classiques-garnier.com>>.
- Féron, Corinne, *Lexicographie du français médiéval et corpus informatisé. Le traitement lexicographique de verbes d'opinion et de connaissance à partir du corpus du DMF*, Corpus 7 (2008), 65–82, <<http://corpus.revues.org/index1606.html>>.
- FEW = Wartburg, Walther von, et al., *Französisches etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes*, 25 vol., Bonn et al., Klopp et al., 1922–2002.
- Gdf = Godefroy, Frédéric, *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IXème au XVème siècle*, 8 vol., Paris, Vieweg, 1880–1895, <<http://micmap.org/dicfro>>.
- Martin, Robert, *Présentation*, 2012, <<http://www.atilf.fr/dmf/PresentationDMF2012.pdf>>.
- TL = Tobler, Adolf/Lommatzsch, Erhard/Christmann, Hans Helmut, *Altfranzösisches Wörterbuch. Adolf Toblers nachgelassene Materialien. Bearb. u. hrsg. von Erhard Lommatzsch, weitergeführt von Hans Helmut Christmann*, Wiesbaden, Steiner, 1925–2008.
- TLF = Imbs, Paul (ed.), *Trésor de la langue française. Dictionnaire de la langue du XIXe et du XXe siècle (1789–1960)*, Paris, Éditions du Centre National de la Recherche Scientifique/Gallimard, 1971–1994, <<http://www.cnrtl.fr>>.

Prof. Dr. Stefan Schneider: Karl-Franzens-Universität Graz, Institut für Romanistik, Merangasse 70/3, A-8010 Graz, E-Mail: stefan.schneider@uni-graz.at