

Ein umgekehrtes Lehnwörterbuch als Internetportal und elektronische Ressource: Lexikographische und technische Grundlagen

Peter Meyer, Stefan Engelberg

Institut für Deutsche Sprache

Mannheim

E-mail: meyer@ids-mannheim.de, engelberg@ids-mannheim.de

Abstract

Der vorliegende Beitrag stellt einen neuartigen Typ von mehrsprachiger elektronischer Ressource vor, bei dem verschiedene Lehnwörterbücher zu einem ‚umgekehrten Lehnwörterbuch‘ für eine bestimmte Gebersprache zusammengefasst werden. Ein solches Wörterbuch erlaubt es, die zu einem Etymon der Gebersprache gehörigen Lehnwörter in verschiedenen Nehmersprachen zu finden. Die Entwicklung einer solchen Webanwendung, insbesondere der zugrundeliegenden Datenbasis, ist mit zahlreichen konzeptionellen Problemen verbunden, die an der Schnittstelle zwischen lexikographischen und informatischen Themen liegen. Der Beitrag stellt diese Probleme vor dem Hintergrund wünschenswerter Funktionalitäten eines entsprechenden Internetportals dar und diskutiert einen möglichen Lösungsansatz: Die Artikel der Einzelwörterbücher werden als XML-Dokumente vorgehalten und dienen als Grundlage für die gewöhnliche Online-Ansicht dieser Wörterbücher; insbesondere für portalweite Abfragen werden aber grundlegende, standardisierte Informationen zu Lemmata und Etyma aller Portalwörterbücher samt deren Varianten und Wortbildungsprodukten (hier zusammenfassend als ‚Portalinstanzen‘ bezeichnet) sowie die verschiedenartigen Relationen zwischen diesen Portalinstanzen zusätzlich in relationalen Datenbanktabellen abgelegt, die performante und beliebig komplex strukturierte Suchabfragen gestatten.

Keywords: Lehnwörter, elektronische Lexikografie, mehrsprachige Ressource, Internetportal

1. Ein Lehnwörterbuchportal als ‚umgedrehtes Lehnwörterbuch‘

Ziel des vorgestellten Projekts ist ein Internet-Wörterbuchportal für Lehnwörterbücher, die Entlehnungen aus dem Deutschen dokumentieren. Dieses Portal ist dadurch gekennzeichnet, dass zum einen die eingestellten Wörterbücher als Einzelwerke veröffentlicht werden und zum anderen auf Portalebene komplexe Abfragen über sämtliche integrierte Wörterbücher hinweg formuliert werden können, zum Beispiel nach dem Weg einzelner deutscher Quellwörter über Mittlersprachen in die verschiedenen Zielsprachen, nach sämtlichen Lehnwörtern in bestimmten historischen Zeitspannen und geographischen Räumen, oder auch nach sämtlichen deutschen Lehnwörtern, die bestimmte Charakteristika aufweisen (z. B. Wortart, semantische Klasse). Das Portal realisiert damit – nicht in den Einzelwörterbüchern, aber in seiner Gesamtheit – als umgekehrtes Lehnwörterbuch das Konzept eines neuen

Wörterbuchtyps.¹ Während es in der Sprachkontaktlexikographie – etwa in Fremdwörterbüchern – üblich ist, Entlehnungsprozesse aus der Perspektive der Zielsprache zu beschreiben, erfasst das geplante Portal aus der Perspektive der Quellsprache die Wege, die deutscher Wortschatz in andere Sprachen genommen hat (Engelberg, 2010). Gegenwärtig wird am Institut für Deutsche Sprache (Mannheim) im Rahmen eines über 18 Monate laufenden und vom Beauftragten der Bundesregierung für Kultur und Medien geförderten Pilotprojektes die grundsätzliche Softwarearchitektur des Portals entwickelt und implementiert sowie die Integration dreier Lehnwörterbücher in das Portal vorgenommen, und zwar zu deutschen Entlehnungen im Polnischen (Vincenz & Hentschel, 2010), zu deutschen Entlehnungen im Teschener Dialekt des Polnischen (Menzel & Hentschel, 2005) und zu deutschen

¹Wiegand (2001) spricht in diesem Zusammenhang von aktiven bilateralen Sprachkontaktwörterbüchern. Wörterbücher dieses Typs sind extrem selten, vgl. auch (Engelberg, 2010). (Görlach, 2001) ist das einzige nennenswerte Beispiel.

Entlehnungen im Slovenischen (Striedter-Temps, 1963). Da das Portal auf Offenheit bezüglich der Integration weiterer Ressourcen konzipiert ist, können jederzeit weitere Lehnwörterbücher integriert werden. Entsprechende Wörterbücher zu Entlehnungen aus dem Deutschen existieren zu relativ vielen Sprachen (Englisch, Japanisch, Portugiesisch, Schwedisch, Serbokroatisch, Tok Pisin, Ukrainisch, Usbekisch, ...). Hier wären entsprechende Kooperationen anzustreben und Rechtsfragen zu klären.²

2. Nutzen eines Lehnwörterbuchportals

Das Portal soll sowohl für Laien wie für Wissenschaftler nutzbar sein. Die Laiennutzung kann über einfache Suchanfragen erfolgen, die wissenschaftliche Nutzung orientiert sich an der Möglichkeit komplexer Suchanfragen und an direkten Schnittstellen (Webservices). Dabei wird sowohl die sprachwissenschaftliche Sprachkontaktforschung wie auch die historisch, soziologisch oder anthropologisch ausgerichtete Kulturkontaktforschung Nutzen aus dem Portal ziehen.

Im Rahmen der wissenschaftlichen Nutzung soll das Portal nicht nur philologisch motivierte, interpretative Einzelstudien unterstützen, sondern durch die in ihm realisierte Kumulation von Daten auch spezifische neuartige, zum Teil quantitativ orientierte Forschungsfragen ermöglichen. Dazu gehören Untersuchungen

- zum Zusammenhang zwischen bestimmten Typen von soziokulturellen Entwicklungen (Herrschaftswechsel, Migration, Technologieschub) und Zeitverlaufstypen der Entlehnungsfrequenzen von Lexemen (wie etwa eine plötzliche oder eine eher graduelle quantitative Zunahme von Entlehnungen),
- zu Faktoren und Prozessen der Etablierung von Lehnwörtern,³
- dazu, ob verschiedene Typen des Sprachkontakts typische quantitative und zeitliche Verteilungsmuster von Lehnwörtern hervorbringen,⁴

- zur Lebensdauer von Lehnwörtern (insoweit die integrierten Wörterbücher auch das Verschwinden oder die Obsoleszenz von Entlehnungen verzeichnen), abhängig von onomasiologischen, grammatischen und anderen Faktoren, vgl. etwa (Schenke, 2009; Hentschel, 2009),
- zu Lehnwortketten (z. B. Deutsch > Polnisch > Weißrussisch > Russisch > Usbekisch) im Zusammenhang mit onomasiologischen und quantitativen Faktoren,
- zu „Germanoversalien“, d. h. etwa dazu, ob bestimmte phonologische, morphologische oder semantische Eigenschaften deutscher Lexeme besonders entlehnungsfördernd sind.

3. Grundsätzliche Überlegungen zur lexikographischen Datenstruktur des Portals

Hinsichtlich der Datenorganisation des Lehnwörterbuchportals lassen sich auf einer konzeptionellen Ebene grob drei Bereiche unterscheiden: (1) Lexikographische Grundlage des Portals sind einzelne Lehnwörterbücher traditionellen Zuschnitts, die nach den fremdsprachigen Lehnwörtern einer bestimmten Nehmersprache lemmatisiert sind. (2) Um sprach- und wörterbuchübergreifende Suchen im Portal zu ermöglichen, muss über diese Datengrundlage eine möglichst dünne Zugriffsstruktur gelegt werden, die von den Idiosynkrasien der Einzelwörterbücher abstrahiert. (3) Für die Etyma der Gebersprache muss eine ‚Metalemmaliste‘ erstellt werden, deren Einträge jeweils über die unter Punkt 2 genannte Abstraktionsschicht untereinander und mit zugehörigen Artikeln der Einzelwörterbücher vernetzt sind.

Die folgenden Unterabschnitte stellen die in den drei genannten Bereichen auftretenden lexikographischen und technischen Anforderungen und Probleme ausführlicher dar, bevor im letzten Abschnitt die technische Umsetzung ihres Zusammenspiels erörtert wird.

²Zum Teil ist die Beschreibungssprache in diesen Wörtern die Quellsprache (z. B. Usbekisch, Portugiesisch), so dass im Falle entsprechende Übersetzungen erforderlich wären.

³Solche Studien können auf lexikographischer und sprachübergreifender Basis. Ergebnisse aus korpusbasierten Arbeiten zum lexikalischen Entrenchment von Entlehnungen komplementieren, vgl. (Chesley & Baayen, 2010).

⁴Sprachkontakttypen wären etwa (i) langandauernder Kontakt

an Bevölkerungsgrenzen (Deutsch – Slowenisch, Deutsch – Polnisch), (ii) Kontakt durch Emigration mit Sprachinselbildung (Deutsch – Rumänisch, Deutsch – Russisch, Deutsch – Amerikanisches Englisch) und Kontakt durch Elitenaustausch (Deutsch – Japanisch, Deutsch – Russisch, Deutsch – Britisches Englisch, Deutsch – Tok Pisin).

3.1. Die Ebene der Einzelwörterbücher

Die zugrundeliegenden Lehnwörterbücher werden im Regelfall bereits existierende Werke sein, die nicht von vornherein für ein Lehnwörterbuchportal des hier diskutierten Typs entwickelt worden sind. Technische Minimalanforderung für die Verwendung im Portal ist, dass die Wörterbücher in geeigneter Form digitalisiert bzw. retrodigitalisiert als XML-Dokumente vorliegen.⁵ Auch eine Bilddigitalisierung ist denkbar, sofern zu jedem Artikel zusätzlich ein XML-Dokument mit den portalrelevanten lexikographischen Daten (und gegebenenfalls Verweisen auf Bildkoordinaten im Digitalisat) vorliegt. Angesichts der enormen Vielfalt möglicher Makro- und Mikrostrukturen in Wörterbüchern ist es nicht praktikabel, für das Portal ein festes XML-Schema vorzugeben, in das sich die XML-Repräsentationen aller Wörterbücher überführen lassen müssen. Es wird jedoch, um weitgehend automatisierte Verarbeitung zu ermöglichen, vom XML-Schema für die Einzelartikel eines jeden Wörterbuchs jeweils verlangt, dass es möglichst weitgehend von Layout- und Präsentationsaspekten abstrahiert, etwa im Sinne der TEI.dictionaries-Richtlinien; vgl. (Burnard & Bauman, 2010). Es gibt gute Gründe, die XML-Digitalisate der Ausgangswörterbücher selber nicht mit portalrelevanten Informationen anzureichern. Abgesehen von urheberrechtlichen Erwägungen und dem angestrebten Erhalt der Wörterbücher als digitalen Einzelpublikationen ist es so möglich, dass an den Einzelwörterbüchern völlig unabhängig von ihrer Nutzung im Lehnwörterbuchportal weiterhin Veränderungen und Erweiterungen von den Autoren des betreffenden Werks vorgenommen werden.

Ähnlich wie bei anderen Portalen können ganze Wörterbuchartikel oder Teile davon (XML-Dokumente bzw. XML-Fragmente) beispielsweise durch

⁵ Aus expositorischen Gründen wird hier auf der Ebene der Einzelwörterbücher durchgehend von einer XML-basierten Datenhaltung ausgegangen, so wie sie im Projekt selber tatsächlich verwendet wird. Technisch lassen sich die Mikrostrukturen von Wörterbüchern natürlich auch in relationalen Datenbankschemata abbilden, was aus Performanzgründen ratsam sein kann. Andererseits können einige moderne Datenbankmanagementsysteme (z. B. Oracle) XML-Daten mit fester Struktur intern ohnehin relational repräsentieren. Vgl. z. B. (Müller-Spitzer & Schneider, 2009) für das OWID-Portal als ein konkretes Beispiel zur texttechnologischen Umsetzung von XML-Verarbeitung in einem Wörterbuchportal.

XSL-Transformationen in eine geeignete HTML-Präsentation überführt werden. Dies ist die Grundlage für eine wörterbuchspezifische Online-Ansicht der Einzelwörterbuchartikel, vgl. (Engelberg & Müller-Spitzer, 2011) für eine ausführlichere Darstellung.⁶ Die XML-Repräsentation ermöglicht außerdem im Prinzip beliebig komplexe Suchvorgänge auf den Einzelwörterbüchern, da konkrete Informationen über Abfragesprachen wie XPath und XQuery aus den Artikeln ausgelesen werden können. Allerdings sind solche XML-basierten Abfragen häufig datenbankseitig mit hohen Verarbeitungskosten versehen und daher für performante Webanwendungen kaum praktikabel. Dies ist ein wesentlicher Grund, die für wörterbuchspezifische sowie portalweite (wörterbuchübergreifende) Suchen relevanten Informationen zusätzlich in separaten relationalen Datenbanktabellen vorzuhalten. Diese zusätzlichen Tabellen ermöglichen nicht nur ungleich performantere Datenbankabfragen, sie dienen auch, wie im folgenden ausgeführt wird, dazu, von den Spezifika der Einzelwörterbücher zu abstrahieren.

3.2. Wörterbuchübergreifende Abstraktionsschicht

Im Normalfall werden die einzelnen Lehnwörterbücher hinsichtlich ihrer Artikel- und Lemmatisierungsstruktur sowie der für Periodisierung und Lokalisierung der Entlehnung verwendeten Begriffe und Angabeformate nicht vollständig kompatibel sein. Auch hinsichtlich der zugrunde gelegten grammatischen Beschreibungssprache kann es Differenzen geben. Der hier vorzustellende Ansatz zur Lösung dieser Probleme stellt insbesondere für wörterbuchübergreifende Suchen eine eigene, relational aufbereitete Datenschicht bereit, die für das Portal relevante Informationen zu allen vorliegenden lexikalischen Einheiten aus den verschiedenen Wörterbüchern in portaleinheitlicher Weise erfasst. In einer wörterbuchübergreifenden Datenbanktabelle werden daher alle Lemmata, alle in den betreffenden Artikeln genannten (diasystematischen, ggf. auch orthographischen) Ausdrucksvariantender Lemmata sowie sämtliche in Einzelartikeln aufgeführten Derivate

⁶ In der skizzierten Weise wird auch bei dem am Institut für deutsche Sprache entwickelten OWID-Wörterbuchportal verfahren (<http://www.owid.de/index.html>).

und Komposita der Lemmata als je eigene Entitäten – im Folgenden als ‚Portalinstanzen‘ bezeichnet – behandelt, also in jeweils einer separaten Tabellenzeile aufgeführt. Eine Tabellenzeile spezifiziert außer dem Wörterbuch, aus dem die Instanz (also das gegebene Lemma bzw. die gegebene Ausdrucksvariante, das Derivat oder Kompositum) stammt, u.a. folgende weiteren Informationen (Attribute), sofern das Wörterbuch diese zur Verfügung stellt: (a) eine räumliche, zeitliche und diasystematische Einordnung des Entlehnungsvorganges; (b) grammatische Informationen, insbesondere Wortart; (c) ggf. eine semantische/onomasiologische Kategorisierung. Außerdem muss jeweils angegeben werden, ob es sich bei der betreffenden Instanz um die Lemmavariante des zugehörigen Wörterbuchartikels handelt, so dass sich aus der Tabelle der Instanzen die Lemmalisten der Einzelwörterbücher ableiten lassen. Falls ein verwendetes Lehnwörterbuch innerhalb eines Artikels z. B. Lesarten unterscheidet, für die unterschiedliche Etymologien diskutiert werden, sind diese in je separaten Portalinstanzen zu kodieren, da von makrostrukturellen Eigenheiten der Einzelwörterbücher abstrahiert werden muss.

Bei hinreichend komplexer und rigider XML-Kodierung eines Lehnwörterbuchs können die Portalinstanzen weitestgehend automatisiert aus den Originalartikeln extrahiert werden. Die Portalinstanzen sollten keine Informationen aus den Lehnwörterbüchern duplizieren; daher enthalten sie außerdem Verweise auf den zugehörigen Artikel und gegebenenfalls auf das dem relevanten Artikelausschnitt entsprechende XML-Element, so dass sämtliche weiteren für die Instanz relevanten Informationen mechanisch aus dem Ursprungsartikel gewonnen und z.B. für eine HTML-basierte Darstellung aufbereitet werden können. Damit portalweite, wörterbuchübergreifende Suchvorgänge möglich sind, müssen zur Erstellung der Portalinstanzen die Angaben der Ausgangswörterbücher zur zeitlichen und räumlichen Einordnung des Entlehnungsvorganges sowie grammatische Informationen in ein einheitliches konzeptuelles Schema überführt werden. Neben komplexen Technologien wie Raum- und Zeitontologien stehen für das Pilotprojekt einfachere Lösungen wie die wörterbuchspezifisch definierte Abbildung von Sprachstufenangaben auf standardisierte Jahresintervalle zur Verfügung. Auch der

Einsatz von Georeferenzierungsverfahren ist in einer späteren Ausbaustufe des Projektes denkbar, um kartographische Visualisierungen zu ermöglichen. Wichtig ist, dass Portalinstanzen mit Informationen angereichert werden können, die keinerlei Entsprechung im zugrundeliegenden Lehnwörterbuch haben. So kann jede Instanz einem Synset einer WordNet-artigen Ressource zugeordnet oder anderweitig semantisch klassifiziert werden, um Abfragen mit onomasiologischer Komponente zu ermöglichen. Schwierig ist dies sicherlich besonders in Wortschatzbereichen, aus denen intensiv und bis hin in fachsprachliche Details entlehnt wurde (z. B. Bergbau, Chemie, Religion).

Auch die Einführung von zusätzlichen Portalinstanzen kann sinnvoll sein; ist etwa ein deutsches Wort über das Polnische in das Russische gelangt, kann der womöglich im polnischen Lehnwörterbuch des Portals gar nicht verzeichnete polnische ‚Zwischenschritt‘ als eigene Portalinstanz hinzugefügt werden.

3.3. Metalemmaliste und etymologische Information

Die lexikographisch und linguistisch anspruchsvollste und zum Großteil manuell zu erstellende Datenschicht ist die Erarbeitung einer Metalemmaliste der Etyma der Gebersprache. Da Lehnwörterbücher häufig mehrere diasystematische bzw. Wortbildungsvarianten der Etyma angeben (darunter auch bloß rekonstruierte Formen) und verschiedene mögliche Etymologisierungen diskutieren, muss – auch angesichts der Probleme mit verschiedenen Transkriptionen – ein möglichst allgemeiner Ansatz gewählt werden. In der von uns gewählten Lösung werden für die in den Einzelwörterbüchern genannten Etymonformen – als *tertia comparationis* des umgekehrten Lehnwörterbuchs – jeweils ebenfalls Portalinstanzen angelegt, die in der Datenbanktabelle mit einem speziellen Attribut als (deutsche) Etymonformen gekennzeichnet werden. Im folgenden bezeichnen wir solche Portalinstanzen kurz als ‚Etymoninstanzen‘. Taucht ein deutsches Lexem in mehreren Wörterbüchern als Herkunftswort auf, wird für jedes Wörterbuch eine eigene Etymoninstanz angelegt, da die Identifikation dieser Instanzen ja erst in einem nachgelagerten lexikographischen Arbeitsschritt auf der Portalebene geschieht. Entscheidend ist daher die Identifizierung von

Gruppen „zusammengehöriger“ Etymoninstanzen. In der von uns vorgeschlagenen Datenmodellierung wird für jede solche Gruppe eine wörterbuchunabhängige Etymon-Instanz erstellt, die unter verschiedenen lexikographischen Gesichtspunkten ein besonders geeigneter Kandidat für ein Metalemma ist, also prototypischerweise ein heute noch gebräuchliches, standardsprachliches deutsches Simplex. Dieses ‚Meta-Etymon‘ kann sinnvoll insbesondere in einer Stichwortliste aller deutschen Etyma des Portals verwendet werden. Alle synchronen oder diachronen Varianten, Wortbildungsprodukte/-bestandteile usw. eines Etymons werden dann auf die im folgenden Abschnitt geschilderte Weise mit ihren zugehörigen Meta-Etyma vernetzt. Es kann wünschenswert sein, zusätzliche Meta-Etyma aufzunehmen, etwa, damit der Benutzer zu einem deutschen Simplex auch dann Entlehnungen von daraus gebildeten Komposita findet, wenn dieses Simplex selber in keinem Wörterbuch als Herkunftswort geführt wird.

4. Zur Architektur der Webanwendung

Die Einführung einer Tabelle von Portalinstanzen ermöglicht die saubere Entkopplung der Portalerstellung von der Ebene der Einzelwörterbücher. Typische portalbezogene Suchvorgänge operieren i.a. ausschließlich auf dieser Abstraktionsschicht.

4.1. Kodierung und Verwaltung der Vernetzungen zwischen Portalinstanzen

Portalinstanzen müssen untereinander vernetzt werden, etwa zur Modellierung von Wortbildungsrelationen. Eine besondere Rolle spielen etymologische Angaben, die als Vernetzungen von Portalinstanzen auf Etymoninstanzen kodiert werden können. Der häufigste Fall ist die Vernetzung von Portalinstanzen, die demselben Quellwörterbuch zugeordnet sind. Um Verkettungen von Entlehnungsvorgängen zu modellieren oder ‚Identitätsbeziehungen‘ zwischen Etymoninstanzen sowie zwischen Lemmata in sehr eng verwandten Sprachformen zu formulieren, müssen aber auch Vernetzungen zwischen aus verschiedenen Quellen stammenden Portalinstanzen angesetzt werden.

Zur Modellierung der Vernetzungen zwischen Artikeln und Instanzen könnten im Prinzip standardisierte Repräsentationsformate wie RDF und die dafür

entwickelten Speicher- und Zugriffstechnologien verwendet werden, vgl. (Hitzler, Krötzsch & Rudolph, 2009). Da aber die Vernetzungsstruktur des Portals sehr regelmäßig ist, ziehen wir eine einfachere Lösung vor, die alle Vernetzungen in einer separaten relationalen Datenbanktabelle als geordnete Paare aus einer Quell- und einer Zielinstanz repräsentiert. Jede Vernetzung von Portalinstanz P auf Portalinstanz Q wird per Attribut einem bestimmten Typ zugeordnet; unter anderem sind folgende Typen vorgesehen: (i) P ist Variante von Q (dabei können Varianten verschiedenen Typs unterschieden werden, z.B. orthographisch/synchron/diachron); (ii) P ist Derivat von Q; (iii) P ist Kompositum zu Q; (iv) P hat Q als Etymon; (v) P ist dasselbe Lexem / dieselbe Lexemvariante wie Q (wenn in einer Entlehnungskette das Lehnwort P selber wieder als Grundlage eines Entlehnungsprozesses gedient hat, wird für dieses Lehnwort eine zweite Portalinstanz Q angesetzt, die das Wort in seiner Rolle als Ausgangswort für die weitere Entlehnung repräsentiert); (vi) P gehört im jeweiligen Einzelwörterbuch zum Lemma bzw. Meta-Etymon Q.

Weitere Attribute von Vernetzungen sind die Quelle der Vernetzungsinformation sowie eine einfache, ordinalskalierte Kategorisierung der in der Quelle selber angegebenen Verlässlichkeit dieser Information.

Die Vernetzungen bilden einen gerichteten azyklischen Graphen (DAG). Bei typischen Suchvorgängen müssen im DAG Pfade von ggf. vorab nicht bekannter Länge ermittelt werden – etwa, um Entlehnungsketten zu finden oder ausgehend von einem Meta-Etymon E nach Derivaten/Varianten/... von Entlehnungen beliebiger Derivaten/Varianten/... von E zu suchen. Um performante SQL-Abfragen auf den Tabellen durchführen zu können, wird in der Vernetzungstabelle der transitive Abschluss der Vernetzungsrelationen oder eine geeignete Teilmenge davon abgebildet, d.h. es werden – zumindest auf der Ebene der Meta-Etyma und Einzelwörterbuch-Lemmata – auch ‚indirekte‘ Vernetzungen gespeichert und als solche etikettiert. Die Verwaltung der Verweisstrukturen zwischen den Datenschichten muss softwaregestützt erfolgen.⁷

⁷ Änderungen an den Einzelwörterbüchern ziehen entsprechende Änderungen in den relationalen Instanzen- und Vernetzungstabellen nach sich, die in den meisten Fällen

4.2. Präsentation

Der Benutzer kann die Einzelwörterbücher mit jeweils eigener (neben der Suchformular-/Artikelansicht ausschnittsweise angezeigten) Lemmaliste und Suchfunktionalität nutzen. Die Etymoninstanzen bilden die Grundlage für ein separates umgekehrtes Lehnwörterbuch, also das Portalwörterbuch der deutschen Herkunftswörter, dessen Lemmaliste aus den Meta-Etyma erstellt wird. Suchvorgänge in diesem Portalwörterbuch erzeugen eine Liste von Verweisen auf passende Artikel in den Einzelwörterbüchern.

5. Literatur

- Burnard, L., Bauman, S. (2010): TEI P5: Guidelines for Electronic Text Encoding and Interchange. Text Encoding Initiative. Online: <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>.
- Chesley, P., Baayen, R.H. (2010): Predicting new words from newer words: Lexical borrowings in French. *Linguistics* 48 (4), pp. 1343-1374.
- Engelberg, S. (2010): An inverted loanword dictionary of German loanwords in the languages of the South Pacific. In A. Dykstra & T. Schoonheim (Eds.), *Proceedings of the XIV EURALEX International Congress* (Leeuwarden, 6-10 July 2010). Ljouwert (Leeuwarden): Fryske Akademy, pp. 639-647.
- Engelberg, S., Müller-Spitzer, S. (erscheint 2011): Dictionary portals. In R. Gouws, U. Heid, W. Schweickard, & H.E. Wiegand (Eds.), *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie. Bd. 4*. Berlin, New York: de Gruyter.
- Görlach, M. (Ed.) (2001): *A Dictionary of European Anglicisms: a Usage Dictionary of Anglicisms in Sixteen European Languages*. Oxford etc.: Oxford University Press.
- Hentschel, G. (2009): Intensität und Extensität deutsch-polnischer Sprachkontakte von den mittelalterlichen Anfängen bis ins 20. Jahrhundert am Beispiel deutscher Lehnwörter im Polnischen. In Stolz, Ch. (Ed.): *Unsere sprachlichen Nachbarn in Europa. Die Kontaktbeziehungen zwischen Deutsch und seinen Grenznachbarn*. Bochum: Brockmeyer, pp. 155-171.
- Hitzler, P., Krötzsch, M., Rudolph, S. (2009): *Foundations of Semantic Web Technologies*. Boca Raton, FL etc.: Chapman & Hall/CRC Textbooks in Computing.
- Menzel, T., Hentschel, G., unter Mitarbeit von P. Jančák und J. Balhar (2005): *Wörterbuch der deutschen Lehnwörter im Teschener Dialekt des Polnischen. Studia slavica Oldenburgensia, Band 10* (2003). Oldenburg: BIS-Verlag. 2., ergänzte und korrigierte elektronische Ausgabe. Online: <http://www.bkge.de/14451.html>.
- Müller-Spitzer, C., Schneider, R. (2009): Ein XML-basiertes Datenbanksystem für digitale Wörterbücher. Ein Werkstattbericht aus dem Institut für Deutsche Sprache. *it - Information Technology* 4/2009, pp. 197-206.
- Schenke, M. (2009): Sprachliche Innovation – lokale Ursachen und globale Wirkungen. Das ‚Dynamische Sprachnetz‘. Saarbrücken: Südwestdeutscher Verlag für Hochschulschriften.
- Striedter-Temps, H. (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.
- Vincenz, A. de, Hentschel, G. (2010): *Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache. Von den Anfängen des polnischen Schrifttums bis in die Mitte des 20. Jahrhunderts. Studia slavica Oldenburgensia, Band 20*. Oldenburg: BIS-Verlag. Online: <http://www.bis.uni-oldenburg.de/bis-verlag/wdlp>.
- Wiegand, H.E. (2001): Sprachkontaktwörterbücher: Typen, Funktionen, Strukturen. In: B. Iglá, P. Petkov & H.E. Wiegand (Eds.). *Theoretische und praktische Probleme der Lexikographie. 1. Internationales Kolloquium zur Wörterbuchforschung am Institut Germanicum der St. Kliment-Ohridski-Universität Sofia, 7. bis 8. Juli 2000 (= Germanistische Linguistik, 161-162)*. Hildesheim, Zürich, New York: Georg Olms Verlag, pp. 115-224.

automatisch durch Datenbanktrigger durchgeführt werden können. Durch solche Trigger können auch Konsistenzprüfungen durchgeführt werden, die manuellen Anpassungsbedarf feststellen und melden.