

VANYA DIMITROVA/HEIKE RENNER-WESTERMANN

VERNETZUNG DES LINGUISTIK-PORTALS MIT LINGUISTIC LINKED OPEN DATA: DIE ROLLE DES BLL-THESAURUS

Abstract: Das Linguistik-Portal ist ein internetbasiertes Rechercheinstrument für die allgemeine und vergleichende Sprachwissenschaft, das einen einheitlichen Zugang zu fachspezifischen Informationen bietet. Eine innovative Komponente des Portals stellt die Vernetzung mit Linguistic Linked Open Data (LLOD) dar: Dadurch wird eine Brücke zwischen linguistischen Repositorien innerhalb der LLOD-Cloud und bibliografischen Daten geschlagen und eine Verbindung zu weiteren Sprachressourcen ermöglicht.

Die Anbindung an die Cloud erfolgt, indem der Schlagwort-Thesaurus der Bibliography of Linguistic Literature (BLL) für LOD aufbereitet und mit relevanten Ontologien verlinkt wird. Der BLL-Thesaurus wird seit über 40 Jahren als integraler Bestandteil der gleichnamigen Bibliografie entwickelt. Darüber hinaus liefert er die Grundlage für die thematische Klassifikation und die normierten Schlagwörter des Linguistik-Portals. Das neue Anwendungsszenario stellt jedoch spezifische technische und konzeptuelle Anforderungen, die nur durch eine formale und inhaltliche Aufarbeitung erfüllt werden können.

Keywords: Linguistik-Portal, Bibliography of Linguistic Literature (BLL), Thesaurus, Ontologie, Linked Open Data, Terminologie-Modellierung

1. Linguistik-Portal und LLOD: Einführung

Das Linguistik-Portal (www.linguistik.de) ist ein Fachportal für die allgemeine und vergleichende Sprachwissenschaft sowie die Linguistiken der europäischen und außereuropäischen Einzelphilologien. Es entstand zwischen 2012 und 2017 an der Universitätsbibliothek Frankfurt in Kooperation mit dem Institut für Deutsche Sprache (IDS) Mannheim, dem Linguistik-Server Essen (LinseLinks) und der AG Angewandte Computerlinguistik des Instituts für Informatik an der Goethe-Universität Frankfurt. Aktuelle Entwicklung und Ausbau finden im Rahmen des DFG-Förderprogramms „Fachinformationsdienste für die Wissenschaft“ statt.

Das Portal bietet Wissenschaftlern, Studierenden und fachlich Interessierten einen breit angelegten, einheitlichen Zugang zu fachspezifischen, wissenschaftlichen Ressourcen: konventionelle, gedruckte oder elektronische Sekundärliteratur, digitale Informationsressourcen im Internet und Forschungsprimärdaten. Zu den Kernmodulen des Linguistik-Portals zählen ein virtueller Katalog für die gleichzeitige Recherche in mehreren Fachkatalogen, Bibliografien, Open-Access-Repositorien und den portaleigenen Modulen, ein Fachinformationsführer zu Internetquellen (Link-Verzeichnis), ein

Verzeichnis von Forschungsprojekten und ein Verzeichnis von Online-Wörterbüchern sowie ein möglichst umfassender Zugang zu linguistisch relevanten Online-Zeitschriften und -Datenbanken.

In der linguistischen Forschung der letzten Jahre ist eine deutliche Tendenz zu beobachten: Forschende rekurren zunehmend auf große textbasierte Datenmengen, die Relevanz von Forschungsdaten und vor allem von Sprachkorpora steigt kontinuierlich. Gleichzeitig wächst disziplinenübergreifend die Akzeptanz für Ansätze des Semantic Web,¹ z. B. die Anwendung von Repräsentationsformalismen, die die Interoperabilität möglichst vieler Ressourcen gewährleisten und durch maschinelle Greifbarkeit eine automatisierte Informationsgewinnung ermöglichen. So steigt auch die Anzahl der linguistisch relevanten Ressourcen, die nach Linked-Data-Prinzipien² aufbereitet sind und unter einer offenen Lizenz zur Verfügung stehen: Die Linguistic Linked Open Data Cloud (LLOD-Cloud),³ die auf eine Initiative der Open Linguistics Working Group der Open Knowledge Foundation zurückgeht, vereint zahlreiche Korpora, Wörterbücher, Wissensdatenbanken, Thesauri, Ontologien sowie Metadatensammlungen und Bibliografien.

Vor diesem Hintergrund starteten 2016 die Arbeiten an der Vernetzung des Linguistik-Portals mit der LLOD-Cloud. Mittlerweile konnte im Portal eine Suchfunktionalität für die Recherche nach fachrelevanten LLOD-Ressourcen implementiert werden. Durch die Vernetzung wurde es umgekehrt auch möglich, ausgehend von Terminologien in der LLOD-Cloud zielgerichtet auf thematisch klassifizierte Publikationen im Portal zu verlinken.

Bei der Anbindung des Linguistik-Portals an die LLOD-Cloud spielt der Schlagwort-Thesaurus der Bibliography of Linguistic Literature⁴ (BLL-Thesaurus) eine wesentliche Rolle. Dafür gibt es mehrere Gründe: Als integraler

¹ Das Semantic Web ist eine Initiative von Sir Tim Berners-Lee, die seit dem Ende der 1990er Jahre vom World Wide Web Consortium (W3C) vorangetrieben wird. Das Semantic Web ist als Erweiterung des vorhandenen World Wide Web zu verstehen. Der Grundgedanke ist, die bereitgestellten Informationen nicht nur in einer für den Menschen lesbaren, sondern zusätzlich auch in einer für Maschinen verarbeitbaren Form zu repräsentieren (vgl. Hitzler et al. 2008).

² „Linked Data baut also auf den bestehenden Web-Standards Uniform Resource Identifiers (URI) und Hypertext Transfer Protocol (HTTP) auf und ergänzt diese durch das Datenmodell RDF [Resource Description Framework] zur Repräsentation von Information, SPARQL [SPARQL Protocol and RDF Query Language] zur Abfrage von RDF-Daten und RDFS sowie OWL [Web Ontology Language] zur Schaffung von Vokabularen/Ontologien.[...] Linked Open Data sind offene Daten (Open Data), die gemäß Linked-Data-Prinzipien (Linked Data) unter Nutzung der entsprechenden offenen W3C-Standards bereitgestellt werden.“ (Online: <http://dini-ag-kim.github.io/glossar/glossar.html>, Stand: Dezember 2018).

³ <http://linguistic-lod.org/lod-cloud> (Stand: Dezember 2018).

⁴ www.blldb-online.de (Stand: Dezember 2018).

Teil der Bibliografie ist der Thesaurus in großem Umfang mit Titeldaten verknüpft; gleichzeitig liefert er auch das Terminologie-Inventar für die Klassifikation und die Erschließung in den portaleigenen Modulen. Für die Anbindung an die Cloud musste der BLL-Thesaurus für Linked Open Data (LOD) aufbereitet und mit entsprechenden Terminologie-Repositoryen verknüpft werden. Als Einstiegspunkt diente das Referenzmodell der Ontologies of Linguistic Annotations⁵ (OLiA) (siehe Chiarcos/Sukhareva 2015), ein zentraler Verbindungsknoten für verschiedene Klassifikationssysteme in der LLOD-Cloud.

Im Folgenden werden die interne Struktur und die historische Entwicklung des BLL-Thesaurus kurz vorgestellt. Die LOD-Aufbereitung wird im Detail beschrieben. Dabei werden die allgemeine Vorgehensweise sowie die methodologischen Herausforderungen thematisiert. Designstrategien und praktische Lösungen werden anhand konkreter Beispiele illustriert.

2. **BLL-Thesaurus: Eigenschaften und Hintergrund**

Die BLL ist eine international rezipierte Fachbibliografie und Datenbank, die an der UB Frankfurt seit 1971 kompiliert und veröffentlicht wird. Mit über 480.000 verzeichneten Publikationen ist sie eine der wichtigsten bibliografischen Informationsquellen für die allgemeine Linguistik mit ihren verschiedenen Teilbereichen und Nachbardisziplinen sowie die anglistische, germanistische und romanistische Sprachwissenschaft.

Der BLL-Thesaurus entstand als fester Bestandteil der Bibliografie und wurde als spezialisiertes Indexierungs- und Recherche-Werkzeug konzipiert. Er ist hierarchisch aufgebaut und bilingual in Deutsch und Englisch angelegt. Die Strukturierung der Information folgt einem begriffsorientierten Ansatz: Jeder Datensatz des Thesaurus soll einem einzelnen Konzept⁶ entsprechen, dem jeweils mehrere Benennungen zugeordnet werden können. In der Regel stehen die deutsche und die englische Schlagwort-Bezeichnung gleichberechtigt nebeneinander. Eine Ausnahme bilden die Fälle, in denen es keine oder keine gebräuchliche deutsche Benennung gibt oder der fremdsprachliche Terminus

⁵ <http://acoli.cs.uni-frankfurt.de/resources/olia> (Stand: Dezember 2018).

⁶ „Konzept“ wird an dieser Stelle und im weiteren Verlauf des Textes synonym zu „Begriff“ verwendet. Die aus der allgemeinen Terminologielehre bekannte Unterscheidung zwischen der abstrakten Einheit (Begriff oder Konzept) und ihrem sprachlichen Ausdruck (Benennung oder Terminus) gilt auch im Rahmen des Thesaurus. „Begriff“ steht dabei für die „Denkeinheit, die aus einer Menge von Gegenständen unter Ermittlung der diesen Gegenständen gemeinsamen Eigenschaften mittels Abstraktion gebildet wird“ (DIN 2342:2011-08, S. 5). Unter „Benennung“ versteht man hingegen die „sprachliche Bezeichnung eines Allgemeinbegriffs aus einem Fachgebiet“ (ebd., S. 11).

sich auch im deutschsprachigen Raum etabliert hat (z. B. *Small clause*,⁷ *Split question*, *Movement (synt.)*).

Der Thesaurus kodiert Vorzugsbezeichnungen, spezifische Verweisbeziehungen wie Synonyme/Quasi-Synonyme (Siehe-Verweise) oder Verwandtschaftsbeziehungen (Siehe-auch-Verweise), Ober- und Unterbegriffe. Dabei sind jedoch keine polyhierarchischen Strukturen erlaubt. So ist z. B. *Psycholinguistik* dem Klassifikationsbegriff *Disziplinen* untergeordnet und dem Schlagwort *Neuro-linguistik* nebengeordnet. *Psycholinguistik* ist der Oberbegriff von *Sprachbewusstheit*, was die deutsche Vorzugsbezeichnung für *Language awareness* ist. Als Verweise werden *Sprachbewusstsein* und *Metalinguistic awareness* angegeben.

Gleichlautende Bezeichnungen werden mithilfe von Homonymzusätzen disambiguiert wie z. B. *Elsässisch* in der Bedeutung „im Elsass verbreitete alemannische Dialekte“ versus *Elsässisch (französischer Dialekt)* in der Bedeutung „im Elsass gesprochene Varietät des Französischen“. Homonymzusätze dienen des Weiteren dazu, die spezifische Perspektive der Betrachtung (z. B. *Symbol (graphem.)*, *Symbol (semiot.)* und *Symbol (psycholing.)*) oder den theoretischen Rahmen (z. B. *Analogie (Metaphertheorie)*) zu kodieren. In der Regel stehen die Vorzugsbenennungen im Nominativ Singular (z. B. *Verb*, *Objekt*). Vereinzelt kommen allerdings auch Pluralformen vor, und zwar bei Sammelbezeichnungen, die sich als Konvention etabliert haben (z. B. *Wortarten*, *Satzteile*), oder wenn Sprachfamilien bzw. Sprachgruppen benannt werden (z. B. *Afro-Asiatische Sprachen*, *Kwa-Sprachen*).

Jedes Schlagwort verfügt über eine eindeutige Kennzeichnung in Form einer Identifikationsnummer (Pica Produktionsnummer, PPN), die das einzige unveränderliche Element eines Datensatzes darstellt. Die dazugehörige Notation (z. B. 04.08.08.014 für *Objekt*) gibt die Position in der hierarchischen Struktur wieder. Die Tiefe der Unterordnungshierarchie variiert von Zweig zu Zweig und erreicht maximal sieben Stufen.

Zum Zeitpunkt der Drucklegung umfasst der BLL-Thesaurus über 8.000 Schlagwörter, die in fünf Hauptzweige aufgeteilt sind: *Disziplinen*, *Ebenen der Sprachbeschreibung*, *Indoeuropäische Sprachen*, *Nicht-indoeuropäische Sprachen* und *Allgemeines* (siehe Abbildung 1).

Der Zweig *Disziplinen* deckt die Subdisziplinen⁸ der Linguistik ab und besteht aus 3.417 Schlagwörtern. Für die Erschließung von Sprachvarietäten (Sprachfamilien, Sprachen, Dialekte etc.) stehen 2.305 BLL-Sprachbezeichner zur Ver-

⁷ Thesaurus-Schlagwörter werden im Folgenden *kursiv* gesetzt.

⁸ Der Zweig *Disziplinen* umfasst 35 Teilzweige wie *Psycholinguistik* (176 Schlagwörter), *Neuro-linguistik* (42 Schlagwörter), *Soziolinguistik* (185 Schlagwörter), *Sprachtypologie* (26 Schlagwörter) etc.

fügung: 839 im Zweig *Indoeuropäische Sprachen* und 1.466 im Zweig *Nicht-indoeuropäische Sprachen*. Der Zweig *Ebenen der Sprachbeschreibung* (2.067 Datensätze) widmet sich der grammatischen Terminologie. Er besteht aus zehn Teilzweigen: *Phonetik* (143), *Phonologie* (221), *Graphemik* (156), *Morphologie* (198), *Lexikologie* (353), *Lexikografie* (67), *Onomastik* (136), *Syntax* (281), *Textlinguistik* (378) und *Semantik* (133). Weitere 313 Schlagwörter, die sonstige Themen wie *Forschungsprojekte*, *Institute/Organisationen*, *Ausbildung*, *Länder/Regionen* etc. abdecken, sind im Zweig *Allgemeines* zu finden.

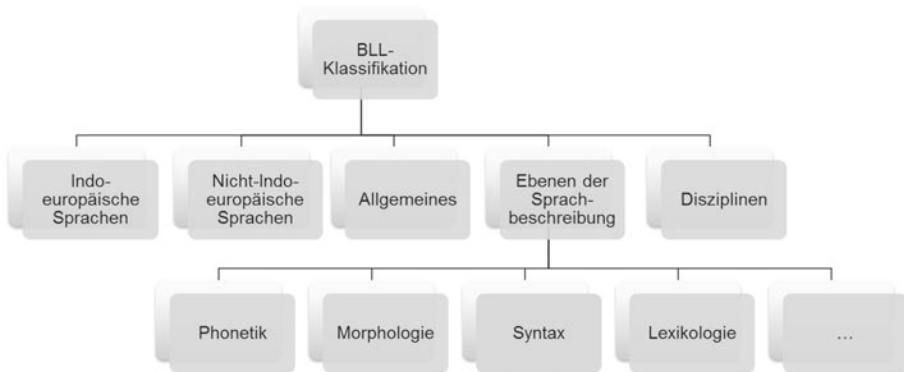


Abb. 1: Die oberste Ebene der Thesaurus-Hierarchie⁹

Der BLL-Thesaurus wird kontinuierlich an die laufende Entwicklung des Fachs angepasst. Dies geschieht hauptsächlich durch das Ergänzen von Schlagwörtern, Siehe-Verweisen oder Siehe-auch-Verweisen. So wurden im Jahr 2014 über 230 neue Schlagwörter wie *Argument sharing*¹⁰ oder *Parasitäres Partizip*¹¹ hinzugefügt. Das Hinzufügen von neuen Schlagwörtern ist dabei nicht das Ergebnis theoretischer Überlegungen, sondern „empirisch“ motiviert. Es basiert auf mindestens einer Publikation, in der das jeweilige Phänomen thematisiert wird. Die BLL erfasst auch Themenbereiche, die noch nicht intensiv erforscht sind; deswegen finden sich im Thesaurus auch Begriffe, die noch nicht zum etablierten „Standardwissen“ gehören. Das Schlagwort *Inflektiv* bezeichnet bspw. besondere Verbformen wie „seufz“ oder „grübel“, wie sie vor allem

⁹ Alle verwendeten Abbildungen stammen aus eigener Darstellung.

¹⁰ „Argument sharing“ ist ein Phänomen, das sich auf die Argumentstruktur von Serialverbkonstruktionen bezieht (vgl. Veenstra 1993).

¹¹ Wood (2013) verwendet den Begriff „parasitic participle“ für „the occurrence of two perfect participles with only one perfect auxiliary (have)“ (ebd., S. 60). Das syntaktische Phänomen gilt als weit verbreitet in den germanischen Sprachen.

in Comics zu finden sind.¹² Löschungen kommen selten vor und werden typischerweise dann vorgenommen, wenn zwei verwandte Schlagwörter zu einer neuen Kategorie zusammengelegt werden. 2014 flossen z. B. die Schlagwörter *Geografie (Fachsprache)* und *Geodäsie (Fachsprache)* in das neue Schlagwort *Geowissenschaften (Fachsprache)* ein. Da die eindeutige Identifikationsnummer (PPN) im Falle einer Löschung blockiert wird und nicht wiederverwendet werden darf, entstehen dadurch keine formalen Inkonsistenzen.

Die Entwicklung des BLL-Thesaurus betrifft nicht nur den Umfang und die Granularität des Vokabulars, sondern auch seine Anwendungsfälle. Die primäre Aufgabe des Thesaurus ist seit über 40 Jahren die Dokumentation von linguistisch relevanten Publikationen im Kontext der Fachbibliografie. Seit 2012 wird der BLL-Thesaurus auch im Rahmen des Linguistik-Portals verwendet: Er liefert die Grundlage für die thematische Klassifikation sowie die normierten Schlagwörter, die der intellektuellen Erschließung in den portaleigenen Modulen dienen. Seit 2015 fungiert der BLL-Thesaurus zudem als Anknüpfungspunkt bei der Vernetzung des Linguistik-Portals mit der LLOD-Cloud.

Das jüngste Anwendungsszenario stellte spezifische technische und konzeptuelle Anforderungen. Um sie zu erfüllen, musste der BLL-Thesaurus formal und inhaltlich aufgearbeitet werden.

3. BLL-Thesaurus als LOD

Die Vorbereitung der LOD-Edition begann mit einer stichprobenartigen Analyse der Schlagwörter sowie der im Thesaurus kodierten Beziehungen. Die anvisierte Anwendung machte auch einen Vergleich der Thesaurus-Struktur mit dem Datenmodell der bereits erwähnten Ontologie OLiA erforderlich.

Als technisches Rahmenkonzept verwendet OLiA die Web Ontology Language (OWL), die auf der Prädikatenlogik erster Stufe basiert.¹³ Für die Herstellung der Interoperabilität zwischen dem BLL-Thesaurus und OLiA ist daher

¹² Forster/Borgwaldt/Neef (2012) beschreiben „inflective“ als „a non-conventional inflection form mostly found in comics or sms/chat communication. In German and English, the inflective corresponds to the bare verb stem, e. g., *seufz*, *gähm*, derived from the verbs *seufzen* and *gähnen*, or the respective English items *sigh*, *cough* derived from *to sigh* and *to cough*.“ (ebd., S. 124).

¹³ RDF und RDF(S) verfügen nur über beschränkte Ausdrucksmittel, die für die Darstellung komplexer Zusammenhänge nicht ausreichen. OWL ist hingegen eine ausdrucksstarke Repräsentationssprache, die die formale Überprüfung der Gültigkeit von Aussagen oder den Zugriff auf implizites Wissen auf der Basis logischer Schlussfolgerungen ermöglicht. Für eine Einführung in OWL vgl. Hitzler et al. (2008).

ein OWL-Modell des Thesaurus erforderlich. Nur auf dieser Grundlage können valide Links zwischen den Ressourcen erstellt werden.

Die Überprüfung der Thesaurus-Struktur zeigte jedoch, dass eine automatisierte formale Überführung in OWL (d. h. die Darstellung der Thesaurus-Begriffe als OWL-Klassen und der Unter-Oberbegriff-Relationen mittels der `SUBCLASSOF`-Property)¹⁴ Inkonsistenzen hervorrufen kann. Denn die hierarchischen Beziehungen zwischen den Schlagwörtern erfüllen nur bedingt die formalen Kriterien einer Ontologie: Sie sind weniger axiomatisch, basieren häufig auf thematischen Assoziationen und nicht auf konzeptueller Subsumtion. So ist z. B. das Schlagwort *Nominalsyntax* dem Thesaurus-Begriff *Satzteile* untergeordnet, was mit der Klassenstruktur von OWL und der Semantik der `SUBCLASSOF`-Property nicht vereinbar ist.

Darüber hinaus wurde festgestellt, dass in vielen Fällen Begriffsdefinitionen fehlen. Dies ist gerade in den Geisteswissenschaften problematisch, da sich zuweilen hinter einer Bezeichnung Konzepte aus unterschiedlichen Denkmodellen verbergen. Zudem wurden mehrdeutige Vorzugsbenennungen nicht immer disambiguiert. In welchem Sinne das Schlagwort *Komitativ*¹⁵ verwendet werden soll, ist bspw. nur anhand seiner Position in der Thesaurus-Hierarchie erkennbar – im vorliegenden Fall als Unterbegriff von *Aktionsart* (*morph.*). Ob das Schlagwort in den zurückliegenden Jahren der Bibliografie-Erstellung tatsächlich nur mit dieser Bedeutung verwendet wurde, kann einzig eine Durchsicht der erfassten Publikationen zeigen.

Die Notwendigkeit, das natürlich gewachsene Thesaurus-Design zu prüfen, war unbestreitbar.

Bei der Größe der BLL-Datenbank, d. h. der Anzahl der mit den Schlagwörtern verknüpften Titel und der Anzahl der Verknüpfungen innerhalb der Schlagwörter, war jedoch eine komplette Überarbeitung des aktiv eingesetzten Thesaurus mit den für das Projekt vorgesehenen Ressourcen nicht realisierbar. Die Lösung bestand darin, für die LOD-Aufbereitung des BLL-Thesaurus ein mehrstufiges Konzept zugrunde zu legen und den Thesaurus physisch bzw. datentechnisch von seinem ontologischen Modell zu trennen. Dies sorgte für größere Flexibilität und erlaubte bei der Modellierung eine

¹⁴ Im Folgenden werden die ontologischen Klassen und Properties durch KAPITÄLCHEN ausgezeichnet.

¹⁵ In Bußmann (Hg.) (2002) werden drei Bedeutungen von Komitativ aufgelistet:

(1) Aktionsart des Verbs, durch die eine Handlung bezeichnet wird, die als „begleitende“ Handlung zu einer anderen vollzogen wird [...] (2) Tiefenkasus für den Begleiter eines anderen Mitspielers [...] (3) Kasus in den Finno-Ugrischen Sprachen für die unter (2) genannte semantische Funktion. (ebd., S. 354)

selektive Bearbeitung der jeweils benötigten Thesaurus-Bereiche, für die in der LLOD-Cloud bereits passende Verknüpfungsmöglichkeiten vorlagen. Für diese Domänen konnten anschließend LOD-Funktionalitäten implementiert werden. Das mittelfristige Ziel ist, nach und nach weitere Thesaurus-Zweige in die BLL-Ontologie zu integrieren.

Für die konkrete Umsetzung wurden drei Modelle erstellt, die hierarchisch aufeinander aufbauen:

- eine SKOS-Version des BLL-Thesaurus,¹⁶
- ein OWL-Modell,
- ein Linking-Modell, das die Verbindung zum OLiA-Referenzmodell herstellt.

Die SKOS-Version repräsentiert LOD-konform die Originalstruktur des BLL-Thesaurus, wird automatisiert erstellt und beinhaltet alle BLL-Schlagwörter. Die jeweiligen Identifikatoren (Uniform Resource Identifiers, URIs) werden auf Basis der internen PPN generiert. Im Falle einer Löschung behalten alle Aussagen bezüglich des gelöschten Schlagwortes ihre Gültigkeit, es wird lediglich als veraltet (OWL:DEPRECATED mit dem Wert „true“) markiert. Das SKOS-Modell stellt die Vorzugsbezeichnungen (mittels `skos:prefLabel`), die hierarchischen Relationen (mittels `skos:broader`),¹⁷ die BLL-interne numerische Notation (mittels `rdf:value`) sowie die Siehe-auch-Verweise (mittels `rdfs:seeAlso`) dar.

Das OWL-Modell des Thesaurus, das BLL-Ontologie genannt wurde, wird intellektuell erstellt und enthält nur ausgewählte Thesaurus-Domänen. Die BLL-Ontologie referenziert dabei auf die SKOS-Version und ergänzt diese um eine annotierte ontologische Struktur, die den OWL-Standards gerecht wird und sich auch für Reasoning-Verfahren eignet. Dadurch wird Kompatibilität mit OLiA und anderen Ontologien, die dasselbe Framework verwenden, erreicht. Dies ermöglicht aber auch weitere Anwendungsfälle wie z. B. eine Ontologie-basierte Suche.

Durch die Anwendung der Prinzipien der Ontologie-Bildung und die damit einhergehende, gründliche intellektuelle Überprüfung wird es leichter, Inkonsistenzen in der Terminologie und der Struktur des Thesaurus aufzude-

¹⁶ SKOS steht dabei für „Simple Knowledge Organisation System“, eine formale Sprache, die im Kontext des Semantic Web standardmäßig genutzt wird, um Thesauri und andere Dokumentationssprachen zu veröffentlichen und zu kombinieren (siehe Miles/Bechhofer 2009).

¹⁷ Die `skos:broader`-Relation ist nicht so rigide definiert wie die `subclassOf`-Relation und wird für die Modellierung von Thesauri empfohlen (siehe Pastor-Sanchez/Martinez/Rodriguez-Muñoz 2009).

cken und – mithilfe der Ausdrucksstärke von OWL – klarer nachzuzeichnen. Die gewählte modulare Architektur und die klare Trennung zwischen den Modellen (SKOS-, OWL- und Linking-Modell) erlaubt eine zusätzliche (ontologische) Formalisierung, ohne den aktiv eingesetzten Thesaurus direkt umstrukturieren zu müssen. Da die Thesaurus-Hierarchie eine substantielle Auswirkung auf das BLL-Datenbank-Retrieval haben kann, ist bei umfangreicheren Umstrukturierungen große Vorsicht geboten.¹⁸

Im Vorfeld der OWL-Modellierung stellte sich noch die Frage nach der Notwendigkeit, eine selbstständige neue Ontologie zu erstellen: Wäre es nicht möglich, den Thesaurus als Submodul einer bereits existierenden linguistischen Ontologie darzustellen, z. B. als eine Community of Practice Extension (COPE, siehe Farrar/Lewis 2007) der General Ontology of Linguistic Description (GOLD, siehe Farrar/Langendoen 2003)?

Die Entscheidung, den BLL-Thesaurus als eigenständige Ressource zu modellieren, beruhte auf der Tatsache, dass auch in der LLOD-Community unterschiedliche theoretische Ansätze und Anwendungsbereiche zu abweichenden Konzeptualisierungen und inkompatiblen Modellen der linguistischen Domäne führen. Eine einzige und einheitliche Linguistik-Ontologie scheint vorläufig unerreichbar zu sein:

An ontology for all of linguistics is, at this point, unachievable and would require deep consensus as to how language is conceptualized. Still, we will undoubtedly come across descriptions that are incompatible with one another due to different theoretical assumptions, in other words, disparate conceptualizations. (Farrar/Langendoen 2010, S. 47)

GOLD ist bspw. eine Ontologie der deskriptiven Linguistik, die für die Bedürfnisse der Sprachdokumentation entworfen wurde mit dem Ziel, Interoperabilität zwischen Sprachdaten zu sichern. Der BLL-Thesaurus wird hingegen für die Indexierung und Dokumentation von linguistisch relevanten Publikationen verwendet und strebt nach der Abdeckung des gesamten Fachgebiets, inklusive Sprachen, Disziplinen, Subdisziplinen sowie der unterschiedlichsten theoretischen Ansätze. Der Umfang des BLL-Thesaurus und die Anzahl der darin enthaltenen Schlagwörter (ca. 8.000) sind in ihrer Größenordnung nicht vergleichbar mit allgemeinen terminologischen Repositori-

¹⁸ Selbstverständlich fließen die aus der Ontologie-Arbeit gewonnenen Erkenntnisse – soweit wie möglich – in eine Umarbeitung des BLL-Thesaurus ein. Bei den auf diese Weise veranlasseten Änderungen handelt es sich vor allem um Aktualisierungen der Vorzugsbenennungen und Ergänzungen fehlender Definitionen, aber auch um Änderungen der hierarchischen Struktur. Für die Bibliografie-Erstellung sind allerdings in vielen Bereichen „flachere“ Hierarchien von Vorteil.

en oder Ontologien, die in der Regel auf wenige gut dokumentierte Bereiche fokussieren. So weist GOLD ca. 500 Klassen und 76 Properties nach, wobei hier der Schwerpunkt auf Morphosyntax liegt.¹⁹

3.1 Modellierung in OWL: BLL-Ontologie

In der BLL-Ontologie werden die Thesaurus-Schlagwörter als Klassen (OWL:CLASS) modelliert. Zur Organisation der Begriffe werden die Relationen RDFS:SUBCLASSOF und OWL:EQUIVALENTOF verwendet. Im Unterschied zum Thesaurus sind hier polyhierarchische Strukturen grundsätzlich erlaubt. Von einer ontologischen Repräsentation der im BLL-Thesaurus vorhandenen Verweisbeziehungen wird vorerst abgesehen.²⁰

Die Arbeit an der BLL-Ontologie fing mit dem Thesaurus-Zweig *Ebenen der Sprachbeschreibung* an. Dieser Zweig deckt den Bereich der linguistischen Metasprache ab, der auch im Fokus des OLiA-Referenzmodells steht.

1.165 BLL-Konzepte flossen in die ontologische Modellierung ein. Die Teilzweige *Syntax*, *Morphologie*, *Lexikologie* und *Phonologie* wurden komplett bearbeitet, die Schlagwörter aus dem *Graphemik*-Zweig zu 75 % und die aus dem *Semantik*-Zweig zu 25 %. Dabei wurde jedes Schlagwort in seiner Beschaffenheit und strukturellen Einordnung inhaltlich geprüft. Das Ergebnis dieser Überprüfung bestimmte seine formale Darstellung.

Die Verifizierung der Begriffsdefinitionen erfolgte Thesaurus-intern oder durch Vergleich mit weiteren terminologischen Repositorien, Nachschlagewerken sowie Online-Quellen. Falls die Definitionen widersprüchlich oder ambig waren, wurden zusätzlich die mit dem jeweiligen BLL-Schlagwort indextierten bibliografischen Einträge überprüft.

Die Überprüfung der Hierarchien erfolgte durch die folgenden einfachen Testverfahren: Substitutionstest (S ist eine Subklasse von O, wenn in jedem beliebigen Satz S durch O ersetzt werden kann) und/oder Paraphrasierungstest (S ist eine Subklasse von O, wenn „S ist ein O“ gültig ist).

¹⁹ <http://linguistics-ontology.org> (Stand: Dezember 2018).

²⁰ Die Gründe dafür sind praktischer Natur: Für die anvisierte Funktionalität hat die Modellierung der Verweise keine Priorität, und die Überprüfung des genauen Status einschließlich Unterscheidung zwischen Synonymen, Quasi-Synonymen und echten Äquivalenzen gestaltet sich oft schwierig. Eine zukünftige Erweiterung der Ontologie ist allerdings nicht ausgeschlossen.

Bei der Erstellung der Klassenhierarchie wurden die BLL-Konzepte sowohl thematisch (Syntax, Morphologie, Morphosyntax etc.) als auch ihrer Natur entsprechend (Kategorie, Prozess, Merkmal oder Relation) gruppiert. Diese Reorganisation wurde durch die Erstellung von neuen, in der Originalstruktur des Thesaurus nicht vorhandenen Klassen unterstützt. Klassen wie MORPHOLOGISCHERPROZESS,²¹ MORPHOSYNTAKTISCHESMERKMAL oder SYNTAKTISCHEKATEGORIE dienen somit als strukturelle Anker.

Die konkrete Umsetzung der taxonomischen Reorganisation wird im Folgenden anhand eines Beispiels illustriert.

Im BLL-Thesaurus hat das Schlagwort *Nebensatz* 16 unmittelbare Unterbegriffe: *dass-Satz*, *Relativsatz*, *Konditionalsatz*, *Temporalsatz*, *Kausalsatz*, *Finalsatz*, *Konzessivsatz*, *Komparativsatz*, *Komplementsatz*, *Adversativsatz*, *Embedding*, *Konsekutivsatz*, *Lokalsatz*, *Modalsatz*, *Subjektsatz* sowie *Sonstige Nebensätze*. Die Überprüfung der Begriffe und der Relationen zeigte, dass die Bedingungen für ontologische Konsistenz in 14 Fällen erfüllt sind. Zehn Konzepte wurden allerdings nicht als unmittelbare Subklassen von *Nebensatz* modelliert. Da sie gemeinsame Eigenschaften teilen – sie bezeichnen Nebensätze, die im Hauptsatz die syntaktische Funktion eines Adverbials erfüllen –, wurden *Konditionalsatz*, *Temporalsatz*, *Kausalsatz*, *Finalsatz*, *Konzessivsatz*, *Komparativsatz*, *Adversativsatz*, *Konsekutivsatz*, *Lokalsatz* und *Modalsatz* als Subklassen der neu geschaffenen Klasse *ADVERBIALERNEBENSATZ* definiert.

Die Beziehung zwischen *Embedding* und *Nebensatz* wurde hingegen als problematisch betrachtet (Paraphrasierungstest: „*Ein Embedding ist ein Nebensatz“). Deswegen bekam das Konzept eine neue hierarchische Einordnung und wurde als Subklasse von *SYNTAKTISCHERPROZESS* definiert (Paraphrasierungstest: „Ein Embedding ist ein syntaktischer Prozess“).

Für *Sonstige Nebensätze* fehlte eine eindeutige Definition. Da das Schlagwort im Thesaurus weiterhin Verwendung findet, wurde es in die Ontologie aufgenommen, aber als Unterklasse von *NICHTKLASSIFIZIERTESLINGUISTISCHESKONZEPT* definiert. *NICHTKLASSIFIZIERTESLINGUISTISCHESKONZEPT* ist eine Hilfskonstruktion, die dazu dient, BLL-Schlagwörter aufzunehmen, die zwar überprüft wurden, für die aber im aktuellen Entwicklungsstand der Ontologie keine taxonomische Einordnung gefunden werden konnte.

Das Ergebnis dieser und weiterer Arbeitsschritte ist der Abbildung 2 zu entnehmen. Während im Thesaurus *Nebensatz* dem Schlagwort *Satzarten* unmittelbar untergeordnet ist, wurde das Konzept in der Ontologie als Subklasse

²¹ Für die neu erstellten ontologischen Klassen gilt: Falls ein Label aus mehreren Wörtern besteht, werden diese ohne Leerzeichen zusammengeschrieben.

VON FINITERTEILSATZ bzw. TEILSATZ modelliert. Dies entspricht der traditionellen Grammatik, in der Satzart (auch Satztyp) sich auf die Unterscheidung der Formtypen des Satzmodus bezieht (Bußmann (Hg.) 2002, S. 582).

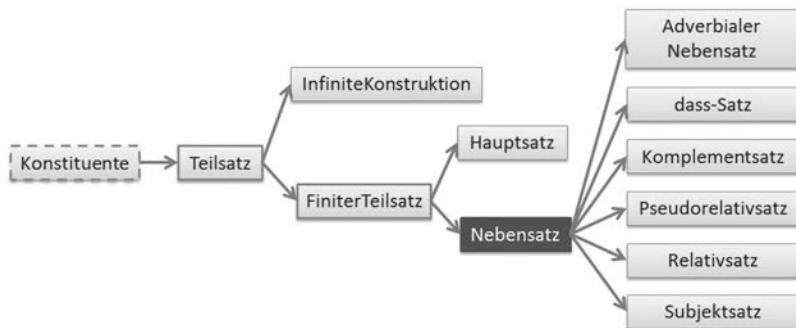


Abb. 2: Die Position der Klasse NEBENSATZ in der BLL-Ontologie

Bei der Überprüfung der Thesaurus-Hierarchie wurde festgestellt, dass in vielen Fällen ein Umbenennen der Begriffe ausreicht, um die Bedingungen für ontologische Konsistenz zu erfüllen. Formal erfolgte eine Umbenennung, indem das ursprüngliche Thesaurus-Konzept mittels `OWL:EQUIVALENTCLASS` durch eine neu definierte Klasse ersetzt und anschließend in der Ontologie mittels `OWL:DEPRECATED TRUE` als veraltet deklariert wurde. So wurde der Thesaurus-Begriff *Syntax* im OWL-Modell durch die Klasse `SYNTAKTISCHESPHÄNOMEN` repräsentiert, der Begriff *Satzteile* durch `KONSTITUENTE` und der Begriff *Wortarten* durch `MORPHOSYNTAKTISCHEKATEGORIE`.

Die Durchsicht der Schlagwörter zeigte, dass einige intrinsisch mehrdeutig sind oder im Laufe der Zeit ambig verwendet wurden. Im Thesaurus wird bspw. das Schlagwort *Instrumental* sowohl für den Kasus (den Instrumental) als auch für die semantische Rolle (das Instrument) verwendet. Das Schlagwort *Komposition* deckt sowohl den morphologischen Prozess ab als auch das Ergebnis dieses Prozesses, das Kompositum.

Komplexeres Wissen lässt sich in OWL durch logische Konstruktoren, die auf Klassen angewendet werden, ausdrücken. Zur Verfügung stehen Konjunktion, Disjunktion und Negation, d. h. logisches *und*, *oder* und *nicht*, realisiert durch die OWL-Sprachelemente `OWL:INTERSECTIONOF`, `OWL:UNIONOF` und `OWL:COMPLEMENTOF`. Mit ihrer Hilfe lassen sich atomare Klassen²² zu komplexen Klassen verbinden (Hitzler et al. 2008, S. 135).

²² „Atomar“ werden die Klassen genannt, die durch einen Klassenbezeichner beschrieben werden und nicht weiter in ihre Bestandteile zerlegbar sind.

Die Methode, die für die Modellierung von *Instrumental* und anderen ambigen Thesaurus-Schlagwörtern erarbeitet wurde, greift auf das OWL-Instrumentarium zurück. *Instrumental* wurde zunächst als Subklasse von MEHRDEUTIGDEFINIERTESKONZEPT eingeordnet. Zusätzlich wurden zwei neue Klassen definiert: INSTRUMENT, eine Subklasse von SEMANTISCHEROLLE sowie INSTRUMENTALKASUS, eine Subklasse von KASUS. Anschließend wurde INSTRUMENTAL mit einer neu erstellten komplexen Klasse gleichgestellt und zwar mit der Vereinigung (mittels owl:UNIONOF) von INSTRUMENT und INSTRUMENTALKASUS.

Die Zulassung polyhierarchischer Beziehungen ermöglicht auch eine präzisere Modellierung von Überlappungen. Grammatische Phänomene mit Ausprägungen auf verschiedenen Beschreibungsebenen können daher als Subklassen von Schnittmengen (owl:INTERSECTIONOF) dargestellt werden. So wurde z. B. *Kasus (morph.)* der Konjunktion von KASUS und MORPHOLOGISCHESPHÄNOMEN untergeordnet, während *Kasus (synt.)* als Subklasse der Schnittmenge von KASUS und SYNTAKTISCHESPHÄNOMEN definiert wurde.

Nachdem die Mehrheit der Thesaurus-Begriffe aus dem Zweig *Ebenen der Sprachbeschreibung* bearbeitet worden war, konnte mit der Verknüpfung von BLL-Ontologie und OLiA-Referenzmodell begonnen werden. Zum Zeitpunkt der Erstveröffentlichung²³ bestand die BLL-Ontologie aus 1.419 Klassen, für die insgesamt 2.511 SUBCLASSOF-Beziehungen sowie 133 EQUIVALENTCLASS-Axiome definiert waren.

3.2 Verknüpfung der BLL-Ontologie mit OLiA: BLL-OLiA-Linking

OLiA wurde mit dem Ziel entwickelt, die Interoperabilität möglichst vieler linguistischer Daten und Metadaten zu sichern. Das frei verfügbare Repository ist modular aufgebaut (siehe Abbildung 3). Zu seinen Hauptkomponenten zählen ein zentrales Referenzmodell (OLiA Reference Model), zahlreiche Annotationsmodelle (OLiA Annotation Models) sowie die Verknüpfungen zwischen den Annotationsmodellen und dem Referenzmodell (OLiA Linking Models) (Chiarcos/Sukhareva 2015).

Zusätzlich ist das OLiA-Referenzmodell mit mehreren externen Klassifikationssystemen (External Reference Models wie TDS²⁴ oder GOLD) verlinkt. Somit dient das Referenzmodell als Mediator zwischen verschiedenen Annotations schemata und weiteren linguistischen Ontologien.

²³ Die BLL-Ontologie wurde als Teil der LOD-Edition des BLL-Thesaurus am 22.12.2016 unter <http://data.linguistik.de/bll> veröffentlicht.

²⁴ TDS steht hier für die Ontologie, die für die Datenintegration im Rahmen des Projekts „Typological Database System“ entwickelt wurde (Online: <http://language.link.let.uu.nl/tds/index.html>, Stand: Dezember 2018).

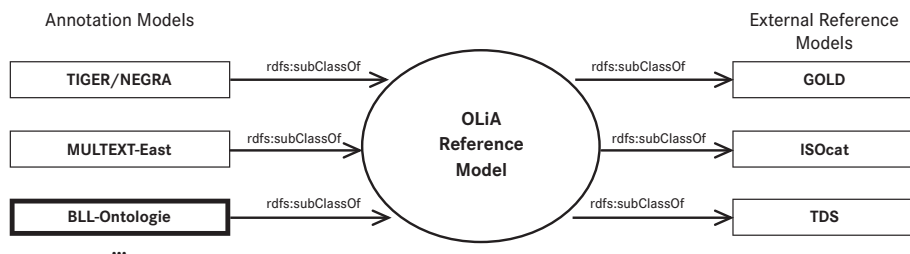


Abb. 3: Modulare Architektur von OLiA

Das OLiA-Referenzmodell definiert Begriffe linguistischer Metasprache und liefert eine Taxonomie von Kategorien aus den Bereichen Morphologie, Syntax und Morphosyntax. Sein Kernvokabular besteht aus 881 Klassen und 51 Properties. Die Annotationsmodelle sind ressourcenspezifische Submodule und stellen in der Regel formalisierte Versionen der integrierten Tag-Sets dar. Die Linking-Modelle hingegen enthalten nur die Verknüpfungen zwischen den Annotationsmodellen und dem Referenzmodell.

Analog zu den Annotationsmodellen erfolgte die Integration der BLL-Ontologie in die modulare Architektur von OLiA mittels eines Linking-Modells (BLL-OLiA-Linking). Dabei handelt es sich im Grunde um eine etwas abgewandelte Form des Mapping,²⁵ eines klassischen Instruments der Zusammenführung von verschiedenen kontrollierten Vokabularen.

Für die Begriffe aus dem Referenzmodell wurde intellektuell in der BLL-Ontologie nach dem entsprechenden oder naheliegenden Eintrag gesucht. Um sicherzustellen, dass sich hinter identischen oder ähnlichen Bezeichnungen auch die gleichen Konzepte verbergen, wurden die Definition und die hierarchische Anordnung der einzelnen OLiA-Begriffe geprüft und mit denen der jeweiligen BLL-Klasse verglichen.

Bei der Erstellung des BLL-OLiA-Linking wurde – wie bei allen anderen OLiA-Submodulen – OWL als formales Framework verwendet. Die Links zwischen den Ontologien wurden mittels `SUBCLASSOF`-Relationen realisiert, d. h. die zu verlinkende BLL-Klasse wurde der entsprechenden OLiA-Klasse untergeordnet. Dabei wurden sowohl Links zu atomaren als auch zu komplexen OLiA-Klassen hergestellt.

²⁵ „Bei einem Mapping bleiben die beteiligten Begriffssysteme in ihrer Originalform bestehen. Es werden lediglich Querverbindungen, sogenannte Mappingrelationen, zwischen ihren Konzepten abgebildet und es wird von einem Vokabular auf das andere zugegriffen.“ (Dunkel 2017, S. 254 f.).

Die Verknüpfungen zwischen der BLL-Klasse APPELLATIVUM und dem OLiA-Konzept COMMONNOUN sowie zwischen der BLL-Klasse ORDINALADJEKTIV und der OLiA-Klasse RELATIONALADJECTIVE stellen Links zwischen atomaren Klassen dar.

Aufgrund unterschiedlicher Konzeptualisierung und Granularität fanden manche BLL-Klassen jedoch kein eindeutiges Pendant im OLiA-Referenzmodell, sondern gleich mehrere Entsprechungen. Basierend auf der Schlagwort-Definition im BLL-Thesaurus ist z. B. die BLL-Klasse WURZEL (MORPH.) so konzipiert, dass sie sowohl die Wortwurzel (die nicht weiter zerlegbare Grundform eines Wortes) als auch den Wortstamm (den Bestandteil eines Wortes, der als Ausgangsbasis für die Flexion und Derivation dient) abdeckt. Im OLiA-Referenzmodell sind hingegen beide Kategorien repräsentiert und zwar durch die Klassen ROOT und STEM. Entsprechend wurde die Verknüpfung hier realisiert, indem die BLL-Klasse einer komplexen OLiA-Klasse, die die Disjunktion von ROOT und STEM darstellt, untergeordnet wurde.

Mehrere BLL-Begriffe ohne Entsprechung zu einer atomaren OLiA-Klasse konnten in anderen Fällen mit einer komplexen OLiA-Klasse verknüpft werden, die die Schnittmenge zweier OLiA-Konzepte mittels OWL:INTERSECTIONOF ausdrückt. So wurde bspw. die BLL-Klasse ADVERBIALES ADJEKTIV mit der Überlappung der OLiA-Kategorien ADJECTIVE und ADVERBIAL verlinkt.

Das BLL-OLiA-Linking enthält insgesamt 262 Verknüpfungen zwischen BLL-Klassen und OLiA-Begriffen.

4. Fazit und Ausblick

Mit der Erstellung des BLL-OLiA-Linking ist eine Grundlage für die Realisierung einer LOD-basierten Suchfunktion im Linguistik-Portal entstanden. Die LLOD-Cloud wird zunächst nach Datensätzen durchsucht, die eine Verknüpfung mit dem OLiA-Referenzmodell aufweisen. Die Metadaten dieser Ressourcen werden unter Verwendung des BLL-OLiA-Linking mit BLL-Schlagwörtern angereichert und in das Linguistik-Portal importiert.

Die LOD-basierte Suchfunktionalität ermöglicht einen zentralen Zugang zu den verteilten linguistischen Ressourcen in der Cloud. Durch die nahtlose Integration der Ergebnisse in den bestehenden Suchindex des Linguistik-Portals wird zudem ein niederschwelliger Zugriff auf LLOD-Ressourcen ermöglicht: Portalnutzer formulieren ihre Suchanfragen auf gewohnte Weise in der ihnen geläufigen Terminologie; die „Übersetzung“ der Suchbegriffe in das andere Vokabular findet im Hintergrund über die Verknüpfung der BLL-Schlagwörter zu OLiA statt.

Die LOD-Edition des Thesaurus umfasst neben der SKOS-Version und der BLL-Ontologie auch den BLL-Index, eine Datei, die die Zuordnung der frei verfügbaren BLL-Titeldaten²⁶ zu den BLL-Schlagwörtern beinhaltet. Die Daten werden unter <http://data.linguistik.de/bll> maschinenlesbar oder als Download zur Verfügung gestellt.²⁷ Die Veröffentlichung des BLL-Index ermöglicht dabei ein zweites Anwendungsszenario: ausgehend von linguistischen Ressourcen in der Cloud auf thematisch klassifizierte Publikationen im Linguistik-Portal zu verlinken.

Die Dateien der LOD-Edition können je nach Bedarf für weitere Anwendungen eingesetzt werden. So können die SKOS-Version oder die BLL-Ontologie bspw. als Ausgangspunkt für Mappings zu weiteren Vokabularien dienen. Gleichzeitig wird die Vernetzung des Portals mit LOD im laufenden Projekt weiter ausgebaut, indem die BLL-Sprachbezeichner formalisiert und mit passenden Repositorien verlinkt werden. Basierend auf dem Verfahren, das für die Bearbeitung der Thesaurus-Begriffe aus dem Bereich linguistische Metasprache entwickelt wurde, werden nun die Schlagwörter aus den Thesaurus-Zweigen *Indoeuropäische Sprachen* und *Nicht-Indoeuropäische Sprachen* in die BLL-Ontologie integriert und mit Lexvo (siehe de Melo 2015) und Glottolog (siehe Nordhoff/Hammarström 2011), zwei frei verfügbaren Metadaten-Repositorien in der LLOD-Cloud, verlinkt.

Die Vernetzung des Linguistik-Portals mit der LLOD-Cloud und insbesondere die Aufbereitung des BLL-Thesaurus für LOD liefern ein Beispiel dafür, wie die praktische Implementierung zentraler Ideen des Semantic Web funktionieren kann. Durch die Entwicklung des Semantic Web wurden verbale Dokumentationssprachen wiederentdeckt: Idee und Umsetzung des Konzepts „Linked Data“ ermöglichen, das volle Potenzial kontrollierter Vokabulare auszunutzen (Gastmayer/Wannags/Neubert 2016, S. 218).

Eine formal und strukturell konsistente Modellierung der linguistischen Terminologie ist allerdings mit spezifischen Herausforderungen verbunden. Schwierigkeiten bereitet vor allem die Polysemie linguistischer Termini. Zudem verändern sich im Rahmen des fachlichen Diskurses oft die Inhalte und die Bezeichnungen der Begriffe. Eine umfassende Repräsentation des Fachgebiets ist daher nur durch kontinuierliche Pflege, Erweiterung und Anpassung des Vokabulars möglich.

²⁶ Alle bibliografischen Einträge, die vor 2008 erschienen sind (ca. 360.000 Titeldaten), sind frei verfügbar.

²⁷ Das BLL-OLiA-Linking wird getrennt gehostet und ist unter www.purl.org/olia/bll-link.rdf verfügbar.

Literatur

- Bußmann, Hadumod (Hg.) (2002): Lexikon der Sprachwissenschaft. 3., aktual. u. erw. Aufl. Stuttgart: Kröner.
- Chiarcos, Christian/Sukhareva, Maria (2015): OLiA – Ontologies of Linguistic Annotation. In: Semantic Web Journal 6, 4, S. 379-386.
- de Melo, Gerard (2015): Lexvo.org: Language-related information for the linguistic linked data cloud. In: Semantic Web Journal 6, 4, S. 393-400.
- DIN 2342:2011-08 = Deutsches Institut für Normung (DIN) (Hg.) (2011): Begriffe der Terminologielehre. Berlin: Beuth.
- Dunckel, Pea (2017): Zusammenführung mehrerer Thesauri zu einem gemeinsamen Begriffssystem. Probleme und Lösungsansätze. In: Information. Wissenschaft & Praxis 68, 4, S. 253-262.
- Farrar, Scott/Langendoen, Terry (2003): A linguistic ontology for the semantic web. In: GLOT International 7, 3, S. 97-100.
- Farrar, Scott/Langendoen, Terry (2010): An OWL-DL implementation of GOLD. An ontology for the semantic web. In: Witt, Andreas/Metzing, Dieter (Hg.): Linguistic modeling of information and markup languages. Contributions to language technology. (= Text, Speech and Language Technology Series 40). Dordrecht: Springer, S. 45-66.
- Farrar, Scott/Lewis, William D. (2007): The GOLD community of practice: An infrastructure for linguistic data on the web. In: Language Resources and Evaluation 41, 1, S. 45-60.
- Forster, Iris/Borgwaldt, Susanne/Neef, Martin (2012): Form follows function: Interjections and onomatopoeica in comics. In: Writing System Research 4, 2, S. 122-139.
- Gastmeyer, Manuela/Wannags, Max-Michael/Neubert, Joachim (2016): Relaunch des Standard-Thesaurus Wirtschaft – Dynamik in der Wissensrepräsentation. In: Information. Wissenschaft & Praxis 67, 4, S. 217-240.
- Hitzler, Pascal/Krötzsch, Markus/Rudolph, Sebastian/Sure, York (2008): Semantic web. Grundlagen. Berlin/Heidelberg: Springer.
- Miles, Alistair/Bechhofer, Sean (2009): SKOS Simple Knowledge Organization System. Reference. W3C Recommendation 18 August 2009. (Online: www.w3.org/TR/2009/REC-skos-reference-20090818, Stand: Dezember 2018).
- Nordhoff, Sebastian/Hammarström, Harald (2011): Glottolog/Langdoc: Defining dialects, languages, and language families as collections of resources. In: Proceedings of the First International Workshop on Linked Science 23.-24.10.2011. Bonn, S. 53-58.
- Pastor-Sanchez, Juan-Antonio/Martinez, Francisco Javier/Rodriguez-Muñoz, José Vicente (2009): Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. In: Information Research 14, 4.
- Veenstra, Tonjes (1993): Serial verb constructions, parameter settings and thematic restrictions on argument sharing. In: Linguistics in the Netherlands 10, 1, S. 153-164.
- Wood, Jim (2013): Parasitic participles in the syntax of verbal *rather*. In: Lingua 137, S. 59-87.