Interoperability of Language-related Information: Mapping the BLL Thesaurus to Lexvo and Glottolog

V. Dimitrova, C. Fäth, C. Chiarcos, H. Renner-Westermann, F. Abromeit
Goethe University Frankfurt, Germany
{v.dimitrova@ub, faeth@em, chiarcos@em, h.renner-westermann@ub, abromeit@em}.uni-frankfurt.de

Abstract

Since 2013, the thesaurus of the Bibliography of Linguistic Literature (BLL Thesaurus) has been applied in the context of the *Lin|gu|is|tik* portal, a hub for linguistically relevant information. Several consecutive projects focus on the modeling of the BLL Thesaurus as ontology and its linking to terminological repositories in the Linguistic Linked Open Data (LLOD) cloud. Those mappings facilitate the connection between the *Lin|gu|is|tik* portal and the cloud. In the paper, we describe the current efforts to establish interoperability between the language-related index terms and repositories providing language identifiers for the web of Linked Data.

After an introduction of Lexvo and Glottolog, we outline the scope, the structure, and the peculiarities of the BLL Thesaurus. We discuss the challenges for the design of scientifically plausible language classification and the linking between divergent classifications. We describe the prototype of the linking model and propose pragmatic solutions for structural or conceptual conflicts. Additionally, we depict the benefits from the envisaged interoperability - for the *Lin|gu|is|tik* portal, and the Linked Open Data Community in general.

Keywords: *Lin|gu|is|tik* portal, Bibliography of Linguistic Literature (BLL), thesaurus, language identifiers, Linguistic Linked Open Data (LLOD), Glottolog, Lexvo

1. Introduction

The *Lin|gu|is|tik* portal (`www.linguistik.de`) is a hub for linguistically relevant information developed by the University Library Frankfurt and the Applied Computational Linguistics lab at the Goethe University Frankfurt. As a main service it provides a research tool that comprises selected online sources, databases, open access documents, bibliographies and catalogs for linguistic literature. Recently, the portal has been connected with the Linguistic Linked Open Data (LLOD) cloud[1]. A novel, LOD-based search function has been developed, and numerous language resources have been integrated.

The thesaurus of the Bibliography of Linguistic Literature[2] (BLL Thesaurus) serves as a connecting point. It has been modeled according to LOD principles and linked to a linguistic ontology covering the domains of morphology, syntax and morphosyntax.

In this paper, we describe current efforts to enhance the functionality by establishing interoperability between the BLL Thesaurus and two LLOD terminological repositories that provide language identifiers: Lexvo[3] and Glottolog[4] .

2. Motivation and previous work

The LLOD cloud comprises lexical-conceptual resources (dictionaries, knowledge bases), corpora, terminological repositories (thesauri, ontologies), and metadata collections. Published under an open license, these resources can be of great benefit to the users of the *Lin|gu|is|tik* portal. Thus, we decided to establish a connection between both platforms and make as much language resources as possible visible and searchable via the *Lin|gu|is|tik* portal.

The concept we developed builds on the interoperability of linguistic terminology and the interconnected nature of the resources in the cloud (Chiarcos et al. (2016)). Since the BLL Thesaurus provides the majority of the subject headings used for indexation within the *Lin|gu|is|tik* portal, it plays a key role for the implementation.

The BLL Thesaurus is a hierarchically categorized bilingual thesaurus of domain-specific index terms in German and English. Since 1971, the thesaurus has been used in the context of the Bibliography of Linguistic Literature. The subject terms are, therefore, interlinked with a significant amount of bibliographical references.

As of February 2018, the BLL Thesaurus comprises 7,965 subject terms organized in five top-level branches. The main branch *Domains*[5] covers the subdisciplines of linguistics (e.g., *Psycholinguistics*, *Sociolinguistics*) and lists 3,350 subject terms. The branch *Levels* includes the levels of language description (e.g., *Syntax*, *Phonology*) and consists of 2,037 subject terms. 312 subject terms are subsumed under the branch *General topics*. Additionally, the BLL Thesaurus provides 2,242 subject terms for the encoding of language-related information.

In a previous project (finalized in December 2016), interoperability between the BLL Thesaurus and the Ontologies of Linguistic Annotations (OLiA) (Chiarcos and Sukhareva (2015)) was established. The implementation involved the conversion of the BLL Thesaurus in Simple Knowledge Organization System (SKOS)[6] format as well as the modeling of the subject terms as ontological classes using the Web Ontology Language (OWL)[7]. The building of the ontological model (i.e., the BLL Ontology) focused initially on the thesaurus branch *Levels* of language description (Dim-

[1] `http://linguistic-lod.org/llod-cloud`
[2] `http://www.blldb-online.de`
[3] `http://www.lexvo.org/`
[4] `http://glottolog.org/`

[5] Thesaurus subject terms are represented in *italics*, and ontological classes or properties in `typewriter font`.
[6] `https://www.w3.org/2004/02/skos/`
[7] `https://www.w3.org/TR/owl-ref/`

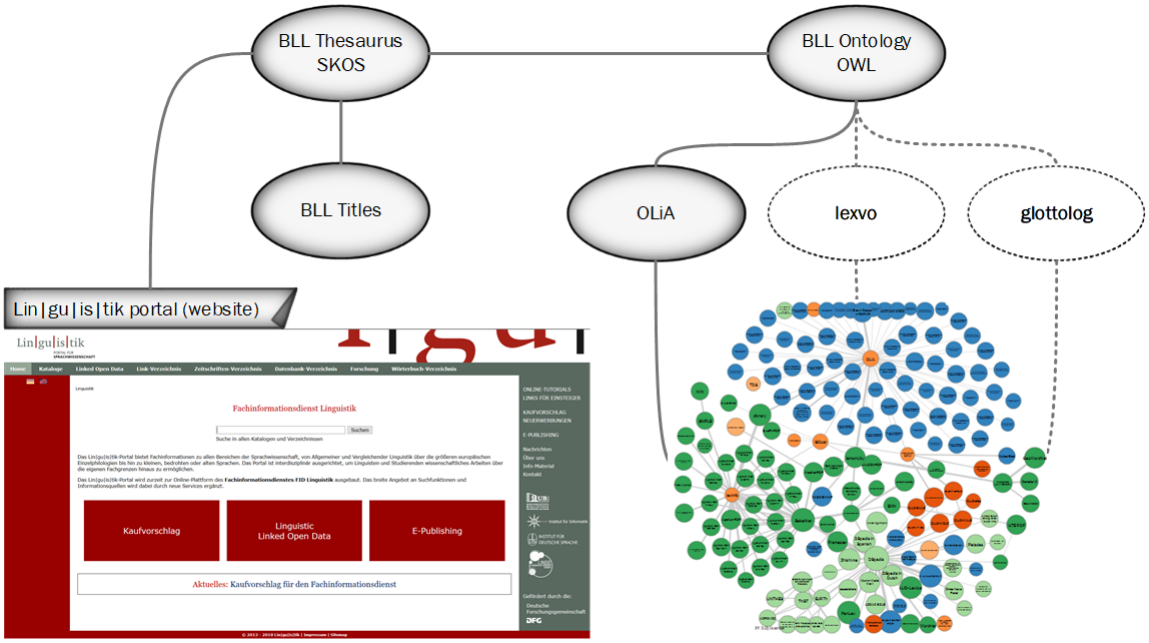

Figure 1: Project architecture

itrova et al. (2016)). By the end of the project, circa 1,100 subject terms covering mainly the domains of syntax and morphology were integrated in the BLL Ontology[8]. Subsequently, the BLL Ontology was linked to the OLiA Reference Model, and on this basis a search algorithm was developed.

Figure 1 features a schematic representation of the connection between the *Lin|gu|is|tik* portal and the LLOD cloud together with the existing (OLiA) and the prospective (Lexvo, Glottolog) links between the platforms. Via the links between the BLL Ontology and OLiA, resources using annotation models interoperable with OLiA have already been integrated into the *Lin|gu|is|tik* portal. Currently, we focus on the branches of the BLL Thesaurus containing language-related information and are working on the integration of the relevant subject terms into the BLL Ontology and their mapping to the LLOD vocabularies Lexvo and Glottolog. Thus, we will achieve broader coverage, enable finer-grained queries, and include more LLOD resources into the portal.

3. Language Identifiers for the Semantic Web

The problems of defining a language make an exact enumeration of the world's languages extremely difficult. Depending on the data source and the classification criteria, the number of the world's languages varies significantly.

The 20th edition of Ethnologue[9], for example, lists 7,099 living languages grouped in 145 language families. As of January 2018, Glottolog defines 7,389 spoken L1 languages classified into 241 families and 188 isolates. The international standard ISO 639-3[10] provides currently 7,858 three-letter codes denoting individual languages or macrolanguages.

3.1. Lexvo

Lexvo is an online service that publishes lexical and language information in both human-readable and machine-readable form. It provides language identification as well as language descriptions. Since 2008, Lexvo has defined URIs of the form `http://www.lexvo.org/id/iso639-3/eng` for the languages covered by the ISO 639-3 standard[11] (de Melo, 2015). The ISO 639-3 standard does not provide codes for dialectal or other substandard forms: each identifier is supposed to denote all spoken or written varieties of the respective language. For language families and language collections, Lexvo defines URIs based on ISO 639-5. The codes provided by this standard are, however, also not exhaustive[12].

For each language, Lexvo delivers extensive descriptions (e.g., multilingual language names, scripts, geographic regions) that are expressed using properties and classes from the Lexvo Ontology[13]. Additionally, the Lexvo Ontology provides properties for notions of identity and near-identity aimed at mitigating with long-standing problems in the Linked Data world (de Melo, 2013). Lexvo not only defines global IDs (URIs) for language-related objects, but

[8]The SKOS version of the BLL Thesaurus and the BLL Ontology can be retrieved from `http://data.linguistik.de/`

[9]`https://www.ethnologue.com/`

[10]`http://www-01.sil.org/iso639-3/`

[11]Unfortunately, the provided identifiers are not completely up-to-date. Since the last update took place in January 2014, changes to the ISO 639-3 codes that took place afterwards are not included in the database.

[12]As of August 2008, ISO 639-5 defined 114 collective codes, covering thus just a portion of the established language families and subfamilies.

[13]`http://lexvo.org/ontology`

also ensures that these identifiers are dereferenceable and highly interconnected as well as externally linked to a variety of resources on the Web, e.g., DBpedia, YAGO (de Melo, 2015).

3.2. Glottolog

The ISO 639-3 standard offers practical solutions for many language identification needs. For the research of worldwide linguistic diversity, however, it is not granular enough. The goal of Glottolog is to collect and formalize information about languages and language resources (Nordhoff and Hammarström, 2011). It provides an exhaustive bibliographical coverage of the world's lesser-known languages that serves as an empirical ground for extensional definitions of languages and language classification.

Applying a novel approach, Glottolog defines a language by the set of documents which describe it. Relations between languages can thus be modeled in a set-theoretic manner. Glottolog introduces the term *languoid* as a cover term for dialect, language and language family. Every languoid is seen as a set that has its own URI and is annotated for ancestors, siblings, children, names, codes, geographic location and references. Subset and superset relations represent genealogical relationships.

Currently, Glottolog lists 320,559 references and defines 23,495 languoids. Languoids are modeled using SKOS and RDFS, and linked to ontologies like GOLD, Lexvo and geo. Furthermore, links to MultiTree, LL-Map, Ethnologue, ODIN, WALS, OLAC, and Wikipedia are provided (Nordhoff, 2012).

3.3. BLL Thesaurus

The language-related information in the BLL Thesaurus is organized in two top-level branches: *Indo-European languages* (825 items) and *Non-Indo-European languages* (1,417 items).

The structure and the granularity of the thesaurus are an outcome of the bibliography's specialization. English, German, and Romance linguistics belong to the focal areas of the BLL. The broad bibliographical coverage of these areas and the required detailed indexing explain the disproportional numbers: *German*, for example, has 115 subterms on five different hierarchical levels, while the subbranch *Afro-Asiatic languages* encompasses only 86 subterms[14].

As Table 1 shows, the subject terms representing language-related information are heterogeneous in nature. Generally, they can be grouped in three main types: individual language/variety, collection, and descriptive type. The first type refers to subject terms denoting a single language variety[15]. The language may be living, ancient, reconstructed or extinct as well as artificial or a sign language. A subject term may also denote a historical stage of a language, or refer to a code-switching phenomenon or register.

[14]In this family, Ethnologue lists 379 living languages. According to Glottolog, the family consists of 374 languages, not including the Omotic subfamily.

[15]Within the BLL Thesaurus, there is no consistent formal representation of the status of a language variety, i.e., whether it is considered a language or a dialect.

The second type denotes collections such as language families, subfamilies, dialectal groups or groupings by region. The remaining subject terms are of a mixed, descriptive type: they not only identify the language, but also describe the context of use (e.g., spatial or temporal aspects).

Type	**Description**	**BLL language identifier**
individual language / variety	living	Romanian
	ancient	Gothic
	extinct	Dalmatian
	historical	Old High German
	constructed	Klingon
	dialectal	Pantiscu
	code-switching	Trasianka
	register	Tok Master
	sign language	New Zealand Sign Language
	linguistic reconstruction of a common ancestor	Proto-Slavic
collection	language family	Afro-Asiatic languages
	language subfamily	Celtic languages
	geographical designation	Caucasian languages
	dialectal group	Scanian dialects
descriptive	individual language in a spatial context	German in Romania
	individual language in a temporal context	15th-18th century Italian

Table 1: Languages in the thesaurus

Compared to the other two models, the BLL Thesaurus seems to have more in common with Glottolog than with Lexvo. While Lexvo supplies a list of global IDs based on the codes defined by the ISO 639 registration authorities, the thesaurus subject terms are defined by an editorial team and encoded in a hierarchically structured way. The method applied for the addition of new BLL language identifiers resembles to some extent the resource-based approach of Glottolog. As a general rule, a new subject term can be included in the thesaurus only if the phenomenon or language in question has already been encountered in scientific publications indexed in the bibliography.

While the taxonomies within the relevant thesaurus branches follow mainly bibliographical principles and are only loosely based on the genetic relatedness between the language varieties, the Glottolog family trees are defined solely on genealogical principles.

The three repositories differ not only in structure, but also in scope and granularity, and none of them covers the other completely. Initial sampling showed that some BLL index terms might well find equivalents in Lexvo, but not in Glottolog (e.g., *Norn*, *Vandalic*). And the other way around, for some BLL language identifiers we can only find matches in Glottolog, but not in Lexvo (e.g., *Elfdalian*, *Hottentot Pidgin Dutch*). In the coverage of dialects, language families and subfamilies, Lexvo and Glottolog differ fundamentally. Determined to make as many BLL subject terms as possible interoperable, we conceptualized a mixed linking model between BLL, Lexvo and Glottolog.

4. Linking the BLL Ontology to Lexvo and Glottolog

When working on the interoperability of the subject terms from the thesaurus branch *Levels*, we applied the methodology introduced by Chiarcos et al. (2016) and briefly outlined in Section 2. This method involved two main steps: the remodeling of the thesaurus subject terms as ontological classes and their linking to the corresponding OLiA classes.
We decided to represent the subject terms as OWL classes for several reasons. First of all, applying OWL constraints facilitates the development of a consistent representation of the domain terminology and helps to uncover problematical modeling. OWL provides description logical operators to represent and to (partially) resolve conceptual overload and ambiguity as observed in the BLL Thesaurus[16]. Furthermore, the establishment of valid links to terminological repositories that adopt OWL as their primary formal framework requires an OWL modeling. And moreover, a fully-fledged ontology is suitable for reasoning and can be used to develop an ontology-based search function.
Since the OLiA Reference Model applied similar modeling principles, the linking between the BLL Ontology and OLiA was implemented by assigning BLL ontological concepts corresponding OLiA superconcepts by means of `rdfs:subClassOf` properties.
The BLL bibliographical entries were modeled in the BLL Ontology as instances of OWL classes (representing the corresponding BLL subject terms). With the thesaurus subject terms being empirically grounded in the bibliography, they could be interpreted – on an abstract conceptual level - as collections of references to linguistically relevant publications. The subject term *Auxiliary verb*, for example, could be seen as an abstraction of all the bibliographical references that concern this morphosyntactical category.
As we started focusing on the language-related information within the BLL Thesaurus, however, we had to reconsider some of the previous design decisions.

4.1. Instances

Terminological repositories that provide language identifiers often model those as instances of ontological classes. Lexvo, for example, uses the class `lvont:language` as defined in the Lexvo ontology, and Glottolog applies `dcterms:LinguisticSystem`, `gold:Language`, `gold:LanguageSubfamily`, `gold:LanguageFamily` and `gold:Dialect`.
We, by contrast, model the BLL language identifiers as ontological classes (e.g., a class `German` with subclasses `HighGerman` and `LowGerman`) following the previously described methodology. Thus, we provide a consistent ontological representation for all subject terms, avoid splitting the thesaurus into heterogeneous fractions, and are able to use the same standard reasoning principles across all branches.
However, the establishment of links between BLL classes and Lexvo or Glottolog instances may lead to formal inconsistencies. Therefore, we made several adjustments to our model.
First of all, we assume a new additional layer of individuals as specific realizations of each concept (in the SKOS version of the thesaurus) or class (in the BLL Ontology).
Within the framework of OLiA, a similar approach is employed: in the annotation models, there is a layer of instances referencing the actual tags of the annotated tag set[17]:

```
:VAINF a :AuxiliaryInfinitive;
  olia_sys:hasTag "VAINF"^^xsd:string .
```

Applying this approach to BLL not only improves its compatibility with OLiA on a conceptual level, but also facilitates a formally consistent representation of the links to Lexvo and Glottolog.
Because of these changes, we had to reconsider the modeling of the bibliographical entries. They are no longer represented as instances of the BLL classes, corresponding to the subject terms used for indexation. Instead, they are now modeled as individuals of the newly created class `bll:Title`. The relationship to the respective subject terms is expressed on the instance level by means of the `foaf:topic` property.

4.2. Class hierarchy

In the BLL Ontology, the classes are represented in a hierarchically structured way. When modeling the BLL language identifiers we face challenges specific to the domain of language classification.
The classification of the world's languages is a notoriously controversial field where political and social aspects often play a more important role than scientific facts. That is why, within the field of linguistics, the question whether a specific variety must be considered a language or a dialect is no longer of primary importance. With regard to the relationships between the languages, many linguists consider the family tree model the only approach of scientific relevance. However, due to lack of sufficient data, it is hardly applicable to all known human languages.
When describing or classifying languages, different traditions or naming conventions can pose further difficulties. The language presently known as Occitan, for example, has been described throughout the centuries as Provençal, Langue d'oc, Limousin or Southern Gallo-Romance (Blanchet and Schiffman, 2004).
The ontological modeling of the BLL language identifiers will take the expectations of the linguistic community into account, but it will not be based exclusively on genealogical principles. We aim at a classification that reflects established conventions and traditions, and, simultaneously, complies with the class structure underlying OWL. Furthermore, the nature and specificity of the BLL Thesaurus should be preserved where appropriate.
The BLL Ontology has been designed in a way that allows enhancement and seamless integration of additional thesaurus branches. The inclusion of the branches

[16]The subject term *Accusative*, for example, captures not only the case and its morphological marking, but also different syntactic aspects of the phenomenon.

[17]The example is taken from the annotation model of the STTS-Tagset, retrieved from `http://purl.org/olia/stts.owl`

Indo-European languages and *Non-Indo-European languages* is facilitated by the definition of a new top-level class `LanguageRelatedTerm` that serves as a structural anchor. The second hierarchical level comprises classes representing well established language families (e.g., `Indo-European languages`) as well as classes designating groupings based on typological criteria (e.g., `CreoleOrPidgin`), geographic location (e.g., `Australian languages`), or modality (e.g., `SignLanguage`).

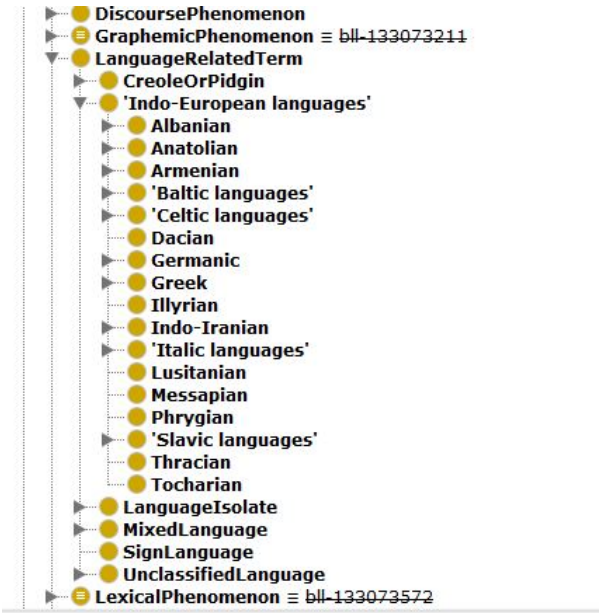

Figure 2: Ontological modeling of the thesaurus branch *Indo-European languages*: hierarchical structure.

Figure 2 illustrates the position `Indo-European languages` takes in the class hierarchy, its superclass and subclasses. Numerous structural reorganizations have to be undertaken in order to create a model that adheres to scientific insights and well established conventions. *Romance*, for example, represents a subconcept of *Indo-European languages* in the BLL Thesaurus. In the BLL Ontology, however, it is modeled as a subclass of `Italic languages` depicting the generally accepted classification for this language group.

Where appropriate, further elaboration of the hierarchical structure takes place. For example, in order to reflect the conventional subgrouping of the Slavic languages in East, South and West Slavic, we defined an intermediate level between the node `Slavic languages` and the individual representatives of the group (e.g., `Russian`, `Bulgarian`, and `Slovak`), and subdivided the languages accordingly.

The BLL Ontology adopts some of the general principles applied in the thesaurus. For example, there is no explicit differentiation between languages, dialects or historical varieties. Historical forms are usually coded as subterms of the respective language and normally "occupy" the same hierarchical level as the dialects of the given language.

4.3. Properties (for interlinking)

When establishing links between BLL concepts and language identifiers provided by Lexvo and Glottolog, we have to determine the nature of the relationship first and then find a fitting relational property for its formal representation. According to preliminary analyses, the entities in the different repositories demonstrate not only genuine identity and near-identity, but also more specific forms of similarity. In order to avoid constraint violations, we apply the property `owl:sameAs` only if the strict form of identity is guaranteed. Although the Lexvo Ontology already provides properties for notions of near-identity, those properties do not seem to suffice our modeling purposes. The simple hierarchy of `lvont:somewhatSameAs` and `lvont:nearlySameAs` is insufficiently distinguishable regarding the properties' strength and use cases. Hence we decided to extend the Lexvo Ontology with a mereologically defined set of properties.

Closely following the W3C best practices[18], we propose a more specific hierarchy for the relations between the language identifiers (Figure 3) which we define as subproperties of the Lexvo property `lvont:somewhatSameAs`. The `bll:overlaps` property is the most general of these and only states that there is at least a subset of entities (e.g., dialectal varieties, individual languages) which is contained in the definitions of both terms interlinked by it. It can therefore be asserted as a symmetric property. The `bll:hasPart` property and its inverse `bll:partOf` describe a transitive, full containment relation where one term contains all elements of the other term but not vice-versa .

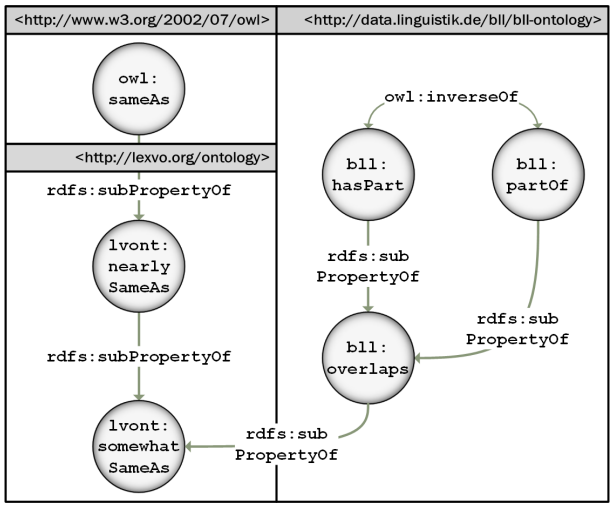

Figure 3: Properties for interlinking language terms

Here are some examples for the prototypical use of the newly defined properties.

The BLL concept Luwian, for instance, is a "container" subject term: it refers to the two varieties of Luwian known after the scripts in which they are written, namely

[18] `https://www.w3.org/2001/sw/BestPractices/OEP/SimplePartWhole/simple-part-whole-relations-v1.5.html`

Cuneiform Luwian and Hieroglyphic Luwian. Since ISO 639-3 defines distinct codes for both varieties, the following links between BLL and Lexvo are proposed[19]:

bll:Luwian `bll:hasPart` **lexvo:Cuneiform Luwian**
bll:Luwian `bll:hasPart` **lexvo:Hieroglyphic Luwian**

The property `bll:overlaps` will be applied mainly when linking BLL concepts with Glottolog language families or subfamilies. For example, both BLL and Glottolog define Anatolian as a subconcept of Indo-European. The BLL concept Anatolian, however, has five subconcepts (Hittie, Luwian, Lycian, Lydian, and Palaic), while the corresponding Glottolog subfamily Anatolian consists of ten individual languages (including the aforementioned five). Despite identical labels and hierarchical position, `owl:sameAs` is not an option here, since both concepts are not isomorphic on a graph-theoretic level. Therefore, the link will be established by means of `bll:overlaps`. But what if a mereological statement is not possible? The BLL language identifier Italo-Albanian, for example, is conceptually related to the Glottolog's languoid Arbëreshë Albanian. As divergent labels often indicate difference in the semantics, both terms cannot be equated. In similar cases Nordhoff and Hammarström (2011) use `skos:closeMatch`. Instead, we employ `lvont:nearlySameAs` as recommended by the Lexvo Ontology.

At the current stage, it is not possible to predict how often the mereological approach will be actually applicable. Taking into account the complexity of the field, using highly underspecified properties such as `lvont:somewhatSameAs` would be, of course, easier to implement. We prefer, however, more specific properties that semantically enrich the data and allow for more elaborate applications. After the linking to Lexvo and Glottolog has been completed, we will evaluate the newly defined properties with regard to frequency of usage and confidence of the relations.

5. Summary and Outlook

In this paper, we described the current efforts to expand the LLOD interface of the *Lin|gu|is|tik* portal, and the required extension of the existing architecture comprising the BLL Thesaurus and the BLL Ontology. In addition to the already implemented ontological representation of the branch *Levels* of language description and its linking to OLiA, we established a framework for the inclusion of the thesaurus branches containing language-related information. By modeling the BLL language identifiers as a hierarchy of ontological classes, we maintain a high level of consistency within the BLL Ontology. By defining individuals as specific realizations of the underlying concepts, we further improve the interoperability. Now, the LOD representation of the BLL can easily be connected to other terminological repositories or ontologies - on the instance level as well as on the class level.

Furthermore, we define a set of additional properties for interlinking language identifiers. We use the gradations of `owl:sameAs` specified in the Lexvo Ontology as a basis, extend them with mereologically defined subproperties and thus enable flexible and precise linking.

The implementation of the OWL modeling of the BLL language identifiers and their linking to Lexvo and Glottolog is work in progress[20]. As of February 2018, circa 85% (700 items) of the subject terms from the thesaurus branch *Indo-European languages* could be hierarchically organized and integrated in the BLL Ontology. For the necessary restructuring, 52 additional classes with no equivalents in the BLL Thesaurus were defined. Approximately 60% of the concepts subsumed under the ontological class `Indo-European languages` could be linked to at least one LLOD repository.

The targeted interoperability between the language-related information in the BLL Thesaurus and the LLOD repositories Lexvo and Glottolog will result in mutual benefits for both the *Lin|gu|is|tik* portal and the LLOD cloud.

Through the established links, the LOD-based search in the *Lin|gu|is|tik* portal will be enhanced: LOD resources that use Lexvo or Glottolog identifiers will become visible and searchable via the portal. Furthermore, the LOD-based search functionality will allow a fine-grained selection from the level of language families down to dialects using either the original hierarchy of the BLL Thesaurus or the manually annotated BLL Ontology. Additionally, the users of the *Lin|gu|is|tik* portal will profit from the integration of the bibliographical data listed by Glottolog.

Building on the LOD principles, we will gain access to information that can facilitate further functions. Glottolog, for instance, provides information about the geographical distribution of languages, and the spatial data can be used as basis for the development of a geographical search. Also, encyclopedic information about language varieties can be integrated through the links between Lexvo and DBpedia.

As of the LLOD community, it will benefit from the inclusion of a significant source of bibliographic material: the BLL lists currently more than 460,000 entries, and the records published before 2001 (circa 250,000 titles) are freely available as RDF. This can be very useful for a platform like Glottolog that employs a resource-based definition of language.

The interoperability of language-related information can also facilitate new applications. Presently, we are working on the implementation of an extended extraction algorithm for LLOD entries which automatically indexes existing resources with information about the languages they cover and the annotation models they use.

Acknowledgments

We would like to thank three anonymous reviewers for insightful feedback and helpful comments.

The research described in this paper is supported by the German Research Foundation (DFG) in the context of the

[19] For better readability in this quasi-Turtle example some local names have been replaced by their corresponding labels.

[20] The updated BLL Ontology and the mapping will be published under `http://data.linguistik.de/` by the end of the project.

project Specialised Information Service Linguistics (*Fachinformationsdienst Linguistik*, funding period 2017-2019).

6. References

Blanchet, P. and Schiffman, H. (2004). Revisiting the sociolinguistics of Occitan: a presentation. *International Journal of the Sociology of Language*, 169:3–24.

Chiarcos, C. and Sukhareva, M. (2015). OLiA - Ontologies of Linguistic Annotation. *Semantic Web Journal*, 6:379–386.

Chiarcos, C., Fäth, C., Renner-Westermann, H., Abromeit, F., and Dimitrova, V. (2016). Lin|gu|is|tik: Building the Linguist's Pathway to Bibliographies, Libraries, Language Resources and Linked Open Data. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 4463–4471, Portorož, Slovenia, May. European Language Resources Association (ELRA).

de Melo, G. (2013). Not Quite the Same: Identity Constraints for the Web of Linked Data. In Subbarao Kambhampati (Conference Chair), et al., editors, *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, pages 1092–1098, Bellevue, Washington, USA, July. Association for the Advancement of Artificial Intelligence.

de Melo, G. (2015). Lexvo.org: Language-Related Information for the Linguistic Linked Data Cloud. *Semantic Web Journal*, 6:393–400.

Dimitrova, V., Fäth, C., Chiarcos, C., Renner-Westermann, H., and Abromeit, F. (2016). Building an Ontological Model of the BLL Thesaurus: First Steps Towards and Interface with the LLOD Cloud. In John P. McCrae, et al., editors, *Proceedings of the 5th Workshop on Linked Data in Linguistics: Managing, Building and Using Linked Language Resources (LDL-2016)*, pages 50–58, Portorož, Slovenia, May. European Language Resources Association (ELRA).

Nordhoff, S. and Hammarström, H. (2011). Glottolog/Langdoc: Defining dialects, languages, and language families as collections of resources. In Tomi Kauppinen, et al., editors, *Proceedings of the First International Workshop on Linked Science 2011*, pages 53–58, Bonn, Germany.

Nordhoff, S. (2012). Linked Data for Linguistic Diversity Research: Glottolog/Langdoc and ASJP Online. In Christian Chiarcos, et al., editors, *Linked Data in Linguistics: Representing and Connecting Language Data and Language Metadata*, pages 191–200. Springer, Berlin, Heidelberg.