Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020), pages 1–10
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Towards an Ontology Based on Hallig-Wartburg's *Begriffssystem* for Historical Linguistic Linked Data

Sabine Tittel, Frances Gillis-Webber, Alessandro A. Nannini
Heidelberg Academy of Sciences and Humanities, University of Cape Town, University of Vienna
Heidelberg, Germany, Cape Town, South Africa, Vienna, Austria
sabine.tittel@urz.uni-heidelberg.de, fran@fynbosch.com, alessandro.alfredo.nannini@univie.ac.at

Abstract

To empower end users in searching for historical linguistic content with a performance that far exceeds the research functions offered by websites of, e.g., historical dictionaries, is undoubtedly a major advantage of (Linguistic) Linked Open Data ([L]LOD). An important aim of lexicography is to enable a language-independent, onomasiological approach, and the modelling of linguistic resources following the LOD paradigm facilitates the semantic mapping to ontologies making this approach possible. Hallig-Wartburg's *Begriffssystem* (HW) is a well-known extra-linguistic conceptual system used as an onomasiological framework by many historical lexicographical and lexicological works. Published in 1952, HW has meanwhile been digitised. With proprietary XML data as the starting point, our goal is the transformation of HW into Linked Open Data in order to facilitate its use by linguistic resources modelled as LOD. In this paper, we describe the particularities of the HW conceptual model and the method of converting HW: We discuss two approaches, (i) the representation of HW in RDF using SKOS, the SKOS thesaurus extension, and XKOS, and (ii) the creation of a lightweight ontology expressed in OWL, based on the RDF/SKOS model. The outcome is illustrated with use cases of medieval Gascon, and Italian.

Keywords: Historical Linguistics, Linked Open Data, Ontology Authoring

1. Introduction

As the most solid grounding of the Semantic Web, the Linked Data (LD) paradigm is used to represent and interlink structured data on the web. The standard proposed by the W3C for representing LD (LOD respectively, with 'O' symbolising open access) is the graph data model *Resource Description Framework* (RDF) that represents data in the form of triples with subject, predicate, and object, each identified through URIs that are accessible via HTTP (Cyganiak et al., 2014). There are many advantages to representing linguistic resources in RDF, and applying LD principles to them, such as structural and conceptual interoperability, uniform access through standard Web protocols, and resource integration and federation (Chiarcos et al., 2013). Representing dictionary data as Linguistic Linked Open Data (LLOD) is a very promising approach, especially as it allows for interoperability among different lexicographic resources through the use of common vocabularies that have emerged for the modelling of linguistic data. The OntoLex-*lemon* vocabulary (Cimiano et al., 2016) has been established as the *de facto* standard RDF data model for LLOD; it provides the framework for the representation of language data such as lexical entries, their written representations, and their meanings. The data modelled with OntoLex-*lemon* can easily be integrated by linking to external resources, such as ontologies for linguistic annotations (e.g., LexInfo[1]), and extra-linguistic information, such as place names (e.g., TGN[2]). We point out that the typical scenario of (historical) linguistic research is characterised by poor data accessibility through searching for *words* and their formal representations across resources of different languages and language stages. This scenario hampers semantic driven research of the *meanings* of the words, particularly for historical language data with non-standardised word spelling. To facilitate access independent from the *words* and their formal representations, the data modelling must, hence, also be enriched by semantic mapping (of entries, senses, concepts) to appropriate ontologies that depict the 'real world' (DBpedia[3], AGROVOC[4], AAT[5], etc.). The use of an external extra-linguistic ontology as a cross-mapping hub for linguistic resources, especially for historical resources, is able to overcome the typical, word-form driven research scenario. This is facilitated by OntoLex-*lemon* and its "principle of semantics by reference in the sense that the semantics of a lexical entry is expressed by reference to an individual, class or property defined in an ontology" (Cimiano et al., 2016, 2.1). One such ontology—in the philosophical meaning of the term—is the so-called Hallig-Wartburg (HW), first published in 1952 (^{2}1963): *Begriffssystem als Grundlage für die Lexikographie* (Hallig and von Wartburg, 1963). In this paper, we focus on the use of HW by linguistic resources and on its transition from a printed book to an LOD resource in order to facilitate its use by linguistic resources on the Semantic Web.

The remainder of the paper is structured as follows: In section 2., we describe the role of HW for linguistic resources of historical language stages that have been or intend to be modelled as LOD. In section 3., we discuss an attempt to convert HW from the original book, via an XML digitisation, into an LOD resource that can be used for semantic mapping. In light of the requirements of the LOD paradigm, we first evaluate a thesaurus-like RDF/SKOS model in section 3.1.; in section 3.2., we discuss its further conversion to an ontological model, and we show its practical application with the use case of data from two historical

[1] `https://lexinfo.net/` [12-02-2020].
[2] `https://www.getty.edu/research/tools/vocabularies/tgn/index.html` [12-02-2020].
[3] `https://wiki.dbpedia.org/` [12-02-2020].
[4] `http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/` [13-02-2020].
[5] `https://www.getty.edu/research/tools/vocabularies/aat/` [12-02-2020].

dictionaries, DAG and LEI, in section 4. Our approach reveals difficulties and shortcomings both with respect to a re-engineering of the ontological model and to the conceptual scheme of HW itself, which we discuss in section 5.

2. Onomasiological Lexicography and the use of Hallig-Wartburg's *Begriffssystem*

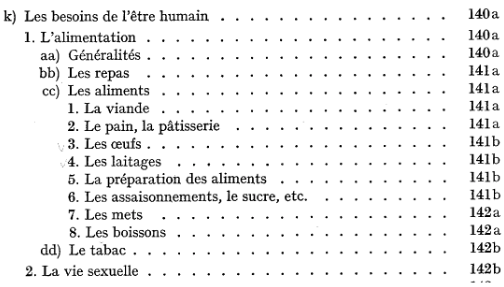
Traditional lexicography either follows a semasiological approach in presenting dictionary data, i.e., the data is ordered by the *words*, or an onomasiological approach, i.e., the data is ordered by the *meaning* of the words. For an onomasiological approach, a thesaurus-like categorisation of the world is needed as a structuring means. Resources referred to as thesauri include the *Historical Thesaurus of the Oxford English Dictionary* (HTOED) (Kay, 2009), Roget's *Thesaurus of English words and phrases* (first edition London 1852, Davidson (2002)), and Dornseiff's *Der deutsche Wortschatz nach Sachgruppen* (Dornseiff, 1934). Possibly the best-known example of a thesaurus-like categorisation of the world used within Romance philology and the reference work of the discipline is Hallig-Wartburg.

2.1. Structure of Hallig-Wartburg

Hallig's and Wartburg's *Begriffssystem*—German for 'system of concepts'—is a conceptual scheme in that it is a controlled vocabulary with a hierarchically structured set of concepts. At first glance, it seems to be a thesaurus-like resource. However, ISO 25964 defines a thesaurus as a "controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms", a term being a "word or phrase used to label a concept" and a concept being a "unit of thought" (International Organization for Standardization, 2011). The terms come from the vocabulary of one or several natural language(s) meaning that they are lexicalised in that language and typically expressed with equivalence relationships (synonyms, quasi-synonyms or antonyms) in the thesaurus (Kless et al., 2012a; Kless et al., 2012b); cp. also Helou et al. (2014) on ontology entities expressed in natural language by associating them with terms. The lexicalisation of the labelling terms is the decisive factor for the classification of HW as not compliant with ISO 25964. HW does not provide lexicalised terms in a natural language. HW, unlike thesauri such as HTOED, Roget, and also the thesaurus-like, lexical database WordNet (Fellbaum, 1998), does not spring from a list of words of a natural language (the 'terms'), e.g., of a (semasiologically structured) dictionary, word list or similar source. Instead, it is meant to be a resource for the use of, e.g., onomasiologically structured dictionaries: It is an *extra-linguistic* reference system of the real world reflecting the model of thought of a 'talented average person' (HW 12), independent from language and with an a priori character ("ein empirisches, aus sprachlichen Allgemeinbegriffen bestehendes, [...] auf phänomenologischer Grundlage beruhenden Gliederungsprinzipien gestaltetes außersprachliches Bezugssystem", ib. 21). HW contains approx. 1675 non-lexicalised concepts ordered in a nine-level hierarchy.

It is clear that a concept must be communicated by a sign, and, indeed, the HW concepts are denoted by *words* of the French language. However, these words are only vehicles and, thus, arbitrary: HW makes it explicit that the words, e.g., 'La mer', are mere symbols of the concepts and not to be misunderstood as lexemes of the French lexicon (ib. 16; 72). This can be illustrated by, e.g., *périodique* (periodical) and *quotidien* (daily) that are both sub-concepts of the concept of *fois* (time [occasion]), not of 'period' and 'day', respectively (ib. 17). As a consequence, concepts may occur several times (with cross-references), e.g., 'fishing' both as an occupation and a sport (ib. 73). The authors of HW were aware of possible misunderstandings and point out that a particular identification of the emblematic character of the French words, e.g., through square brackets, would have been useful but that they refrained from this for the sake of readability (ib.).

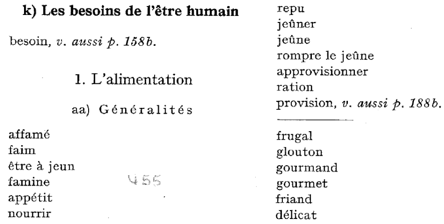
The concepts of the upper six levels of the hierarchy are denoted by French non-lexicalised categories, e.g., 'L'univers', 'Le ciel et l'atmosphère', and 'Le ciel et les corps célestes', and, additionally, the concepts are identified by a system of capital letters between A and C, followed by Roman numerals, Arabic lower case letters, etc.: 'A', 'A I', 'B II h', etc. This six-level hierarchy forms the 'Plan' with 524 concepts, the outline with the logical abstraction of concepts representing broader, conceptual fields, cf. HW 101–112 (Fig. 1).



k) Les besoins de l'être humain	140a
1. L'alimentation	140a
aa) Généralités	140a
bb) Les repas	141a
cc) Les aliments	141a
1. La viande	141a
2. Le pain, la pâtisserie	141a
3. Les œufs	141b
4. Les laitages	141b
5. La préparation des aliments	141b
6. Les assaisonnements, le sucre, etc.	141b
7. Les mets	142a
8. Les boissons	142a
dd) Le tabac	142b
2. La vie sexuelle	142b

Figure 1: The 'Plan' (extract), HW 103.

In HW 113–229, the six conceptual levels of the 'Plan' are then further extended by another, up to three-level hierarchy of approx. 1,150 finer-grained concepts for "lexicography proper as represented by the 'words' classified in its application" (Orr, cited by HW 20, footnote 4), which we will refer to as 'Application' in the following (Fig. 2). These concepts are not consecutively numbered.



k) Les besoins de l'être humain

besoin, *v. aussi p. 158b.*

1. L'alimentation

aa) Généralités

affamé
faim
être à jeun
famine
appétit
nourrir
repu
jeûner
jeûne
rompre le jeûne
approvisionner
ration
provision, *v. aussi p. 188b.*

frugal
glouton
gourmand
gourmet
friand
délicat

Figure 2: Finer-grained 'Application' (extract), HW 141.

Thesauri (and this applies to a conceptual scheme such as

HW as well) establish hierarchical relationships and associative relationships between concepts. The hierarchical relationships can be generic, a whole-part relation, and a concept-instance relation; the associative relationships exist between hierarchically unrelated but semantically or conceptually related concepts (Kless et al., 2012a, 135f.). HW contains hierarchical (both generic and whole-part relations) and also associative relationships between the concepts (HW 18); neither cyclic hierarchical relationships nor orphans. HW prioritises the hierarchical over the associative classification but deliberately prefers the latter in cases where an association seems more 'natural' (ib.), particularly in fields where the concepts are closely connected to specialised domains, such as house building and hunting. With this approach to classification, HW wants to take account of the fact that every language has its own peculiar interpenetration of systematics and non-systematics, which is reflected in the linguistic interpretation of the world (ib.) E.g., the concept 'construire' (to construct) is neither hierarchically allocated to 'L'action' (B II h 3) [together with 'faire' (to make) and 'créer' (to create)], nor to 'L'espace' (space, C I e) [together with 'assembler' (to assemble)]. Instead, it is associated to the concept of house building, i.e., 'La construction' (B III b 7 bb, sub 'L'habitation, la maison'). The concept 'miette' (crumb) is logically a sub-concept of 'morceau' (part, sub-concept of C I d 'Le nombre et la quantité') but associated to the concept 'Le pain, la pâtisserie' (bread, patisserie, B I k 1 cc 2), and 'saumure' (brine) is a concept associated to 'La viande' (meat, B I k 1 cc 1). An example for a hierarchical, whole-part relation is the relation of the concept 'les narines' (nostrils) to its superordinate concept 'Le corps et les membres' (the body and its parts).

The concepts and their classification reveal problematic congruencies, wrong hierarchisation, and inconsistencies[6]:

1. On levels 1-6, we find the identical concept 'Généralités' 27 times, semantically disambiguated through its place in the hierarchy, e.g., as a sub-concept of 'Les arbres'; these concepts can be suppressed since one could simply refer to the respective superordinate concept. On levels 7-9, 'esp.' (abbreviating *espèces*, sub-species, e.g., of the apple) occurs.
2. On levels 8 and 9, we find the string 'etc.' as a concept denomination.
3. On levels 7-9, some concepts are followed by references to homonymic concept denominations (printed in italics, separated by a comma), e.g., 'port, *v. aussi p. 197a*'.
4. On levels 7-9, some concept denominations are specified through German definitions. In some cases, this aims at the semantic disambiguation of homonymic concept denominations within the same superordinate concept, e.g., 'beau-père "Schwiegervater"' (father-in-law) / 'beau-père "Stiefvater"' (stepfather).
5. C II a 17 'La phonétique' is on the same hierarchy level as C II a 18 'La linguistique' but should be a sub-concept of the latter.
6. We find 'alchimie' falsely classified under A II e 'Les métaux' which is a sub-concept of the top concept A 'L'Univers'. However, this top concept should contain only sub-concepts related to organic and inorganic nature, and not to human activities (HW 89).
7. Similarly, under A IV 'Les animaux' we find 'Les animaux fabuleux' (fabulous beasts) and its sub-concepts 'phénix' (phoenix) and 'dragon' (dragon), concepts that cannot be separated from human conception and should, thus, rather be associated to B II e 'L'imagination'.
8. A classification inconsistency is the presence of the sub-concept 'Le tabac' (tobacco, B I k 1 dd) under 'Les aliments' (food, B I k 1), as if tobacco were food.

[6]Naturally, concepts that reflect the zeitgeist of the time of HW's creation, e.g. 'Les costumes nationaux et pittoresques', are to be found as well.

2.2. Lexicographical and Lexicological Resources using Hallig-Wartburg

HW has been chosen by numerous lexicographical and lexicological works as a means of semantic structure. The most comprehensive *Französisches Etymologisches Wörterbuch* (FEW) (von Wartburg, since 1922) is a dictionary of the Galloromance languages and dialects covering the period from the middle ages until today, structured by the alphabetical order of the etyma of the treated word families. The words of unknown or uncertain origin are treated in vol. 21–23 where they are grouped onomasiologically, ordered by the HW concepts. The HW concepts form the structural backbone of the dictionaries *Dictionnaire onomasiologique de l'ancien occitan* (DAO) (Baldinger, 1975 to 2005) and the *Dictionnaire onomasiologique de l'ancien gascon* (DAG) (Baldinger, since 1975): both follow HW to structure the editing and publishing of the dictionary entries (Glessgen and Tittel, 2018, 805). Semantic criteria are used in the *Lessico Etimologico Italiano* (LEI) (Pfister, since 1979) to build the structure of very complex articles, as in the FEW 21–23 (Tancke, 1997, 466); in these cases, the lexicographical sections are ordered by semantic categories (in Italian language) that closely recall those of HW. Recently, the online edition of the *Dictionnaire de l'occitan médiéval* (DOM) (Stempel, 1996 to 2013) started evaluating the introduction of HW concepts to align the entries to those of DAG*él*.[7] The *Dictionnaire étymologique de l'ancien français* (DEAF) (Baldinger, since 1971) follows a semasiological approach but inherits HW categories when it refers to entries of FEW 21–23. The *Mittelhochdeutsche Begriffsdatenbank* (MHDBDB[8]) creates an onomasiological database for Middle High German, building on HW (Hinkelmanns, 2019): the HW categorisation has been further developed with the application on the lexis of Middle High German *Frauendienst* by Ulrich von Lichtenstein (1255) and of *Lanzelet* by Ulrich von Zatzikhoven (after 1193) (Schmidt, 1980; Schmidt, 1988; Schmidt, 1993). Also, many onomasiologically structured lexicological studies on medieval until 16th century French, Italian, Spanish, Gascon and Occitan resources (literary texts, architecture, Bible, etc.), use HW concepts, e.g., Bevans (1941) on the Old French vocabu-

[7]Personal communication by Maria Selig, DOM.
[8]`http://mhdbdb.sbg.ac.at/` [06-02-2020].

lary of Champagne[9], Keller (1953) on the vocabulary used by Wace (* approx. 1110 – † after 1174), de Man (1956) on the Brabant language in archival sources 1300-1550, etc. (Baldinger, 1959, 1091f.).

2.3. Hallig-Wartburg in Linked Open Data resources

As a contribution to the emerging linguistic LOD cloud and to expand the inadequately represented historical linguistic resources, efforts to model these lexicographic resources as Linked Data have been initiated: The FEW is currently digitally available as bitmap images[10] but a digitisation by means of XML is underway (Renders, 2015), and Renders (2019) announces a study on how to model etymological data of the FEW as LOD. For the electronic version of the LEI, LEI-Digitale (Prifti, 2019), the LEI editors carry out feasibility studies on LOD modelling and semantic mapping to HW or to a taxonomy based on HW (Nannini, in progress). Tittel and Chiarcos (2018) created a RDF data model for the electronic version of the DEAF (DEAF*él*) and Tittel (in progress) for DAG*él*, the electronic complement to the DAG (Glessgen, since 2014). The relaunch of the MHDBDB (planned for 2020) will include an RDF version of the data (Hinkelmanns, 2019).

3. From the *Begriffssystem* to an Ontology

The representation of HW in RDF, and SKOS or as an ontology, achieves compatibility with other Semantic Web technologies and is thought to facilitate interoperability across linguistic resources applying HW as their onomasiological framework. This helps to establish the word-form- and language-independent access to these resources: a pivotal motivation to model them as LOD and to include references to the HW concepts. A potential reuse both of HW and of the linguistic resources using HW is also thought to be promoted by the fact that the HW RDF graph is easy to be referenced by other bigger, more comprehensive and more detailed LOD resources, independent from a natural language. Also, recall one of the main principles of the LD paradigm: to provide useful information (in RDF) that is returned when navigating to a URI, i.e., provide dereferenceable URIs.

However, the native format of the HW is a book publication which, thus, needs to be converted into a format compliant with the LOD paradigm. For the digital editing of the DAG*él*, the 524 numbered concepts (the ‘Plan’, Fig. 1) of HW (second edition 1963) have been digitised in 2014 using DAG’s dictionary writing system (Glessgen and Tittel, 2018). The finer-grained approx. 1,150 concepts of the ‘Application’ (Fig. 2) were excluded from the digitisation because the DAG*él* uses only the concepts of the ‘Plan’ as its framework. As a first step towards an RDF graph based on HW, we exported the data as XML from the DAG*él*’s database. The XML structure is based on rows with a single XML element `field` and one attribute with two possible contents, as shown in List. 1. Alas, it does not contain information that can easily be exploited for a future hierarchical representation of the category levels, as visualised in Fig. 3.

```
<?xml version="1.0"?>
<resultset
 xmlns:xsi="http://www.w3.org/2001/
 XMLSchema-instance">
<row>
 <field name="identifier">B I k 1 cc 1</field>
 <field name="concept">La viande</field>
</row>
<row>
 <field name="identifier">B I k 1 cc 2</field>
 <field name="concept">Le pain, la pâtisserie</field>
</row>
</resultset>
```

Listing 1: Extract of XML data.

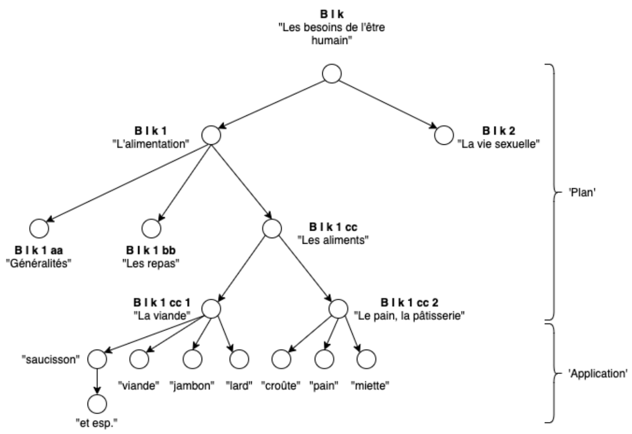


Figure 3: Hallig-Wartburg concept hierarchy.

3.1. Hallig-Wartburg in RDF and SKOS

HW is represented in a standard format of a Knowledge Organisation System (KOS), a system to represent classification schemes, thesauri, taxonomies and similar structures. The W3C has defined the *Simple Knowledge Organization System* (SKOS) which provides a data model and vocabulary for expressing KOSs in RDF (Miles and Bechhofer, 2009). Two types of semantic relations are distinguished by SKOS: hierarchical and associative. The hierarchical relation is typically represented by the ‘narrower’ and the ‘broader’ property, an associative relation is indicated by the use of ‘related’. However, the specific nature of concept relations cannot be expressed. The ISO 25964 SKOS extension (Miles and Brickley, 2004) distinguishes finer-grained semantic relations between the concepts and aims at providing better interoperability between SKOS and the thesaurus standard. It is ideal to explicitly express hierarchical generic and whole-part relationships through the SKOS-Thes properties ‘broaderGeneric’ and ‘broaderPartitive’ respectively. The associative relation expressing a partitive relationship between concepts can be expressed through the more specific property ‘relatedPartOf’. *Extended Knowledge Organization System* (XKOS) was developed to extend SKOS for statistical classification, and one of its features, comparable to the SKOS-thesaurus extension, is to refine SKOS’ semantic properties (Cyganiak

[9]Draws on the *Questionnaire*, the dialectal recordings made by Rudolf Hallig as a preparation for HW (Christmann and Böckle, 1983, 398).

[10]`https://apps.atilf.fr/lecteurFEW/` [accessed 05-02-2020].

et al., 2017). XKOS is a public working draft of a potential specification and therefore we chose to use the SKOS-thesaurus extension to express the semantic relations, although the properties of the latter are still classified as ‘unstable’. Nevertheless, XKOS offers possibilities to define the classification levels of a KOS which we deem valuable for our approach. The representation of HW’s hierarchical and associative relationships is thus straightforward. However, the respective relations are not explicitly expressed in the original source, and a representation in SKOS must comprise a manual assessment of the relations.

We converted the XML data into RDF and SKOS (including extensions), applying the following rules:

1. Since the HW is concept-based according to ISO 25964, all HW concepts can be represented as SKOS concepts.
2. To define the hierarchy levels and their respective members, we include XKOS ‘ClassificationLevel’.
3. We define the three concepts of the top level, A ‘L’univers’, B ‘L’homme’, and C ‘L’homme et l’univers’, as top concepts of the concept scheme (List. 2, l. 7).
4. We utilize the content of XML <field name="concept"> as the concept denomination: to emphasize the symbolic character of the denomination by capitalising all characters, eliminating French accents and replacing spaces, punctuation marks, and apostrophes with an underscore, e.g., `L_HOMME_ET_L_UNIVERS`.
5. We also utilize said content to add a SKOS ‘scopeNote’ providing information about the scope of the concept. Aiming at removing possible ambiguity or misunderstanding of the non-lexicalised information (erroneously as ‘terms’) we deem a scope note the accurate ‘translation’ of the information given in HW.
6. In SKOS, preferred and alternative lexical labels can be used for “generating or creating human-readable representations of a knowledge organization system” (Miles and Bechhofer, 2009); it is consistent with SKOS to assign (multiple) alternative lexical label(s) but no preferred lexical label to a resource. SKOS does not specify whether a resource with none of the two lexical labels is consistent with the SKOS data model, however, it is said to be advised to include a lexical label “in order to generate an optimum human-readable display” (ib.). Considering this advice and the *de facto* missing terms in HW that could naturally become lexical labels, we propose to misuse the concept denominations: We allocate an additional function to the French words used as arbitrary symbols by Hallig and Wartburg interpreting them as ‘terms’ expressed through `skos:altLabel`, e.g., “Les besoins de l’être humain”. This design decision aims to compensate for the missing terms but refrains from declaring preferred labels.
7. For backwards compatibility, we preserve the consecutive numbers of the upper six levels as contained in XML <field name="identifier">, using the SKOS ‘notation’ property; we define the string literal by a particular HW specific identification scheme <hwIdentificationScheme>.
8. We eliminate concepts denominated by ‘etc.’, assuming that the linguistic resources using HW as a reference do not classify lexemes under a concept ‘etc.’ (approved by the editorial team of the DAG*él*).
9. Hierarchical generic relations are expressed through `skos-thes:broaderGeneric`, e.g., the relation between ‘La viande’ and ‘Les aliments’ (List. 2, l. 40), hierarchical whole-part relations through `skos-thes:broaderPartitive`, e.g., the relation between ‘les narines’ and ‘Le corps et les membres’ (List. 2, l. 51), and associative relations through `skos-thes:relatedPartOf`, e.g., the relation between ‘miette’ and ‘Le pain, la pâtisserie’ (List. 2, l. 46). To enable navigation from the top concept level down into the hierarchy, we include the SKOS ‘narrower’ property (l. 27; 41).
10. We distinguish homonymic concepts within the same superordinate concept, that are, thus, not disambiguated by their respective, different superordinate concepts, as follows: We add a number to the concept denomination and preserve the German definitions that are used for the semantic disambiguation as a SKOS ‘editorialNote’, e.g., sub B III a 1 aa 3 (‘La parenté’), ‘beau-père’: `:BEAU_PERE_1 skos:scopeNote "beau-père"@fr skos:editorialNote "Schwiegervater"@de` and `:BEAU_PERE_2 skos:scopeNote "beau-père"@fr skos:editorialNote "Stiefvater"@de`. We chose `editorialNote` over the ostensibly obvious SKOS property `definition` to be able to use the latter for a further knowledge enrichment with accurate genus-differentiae sense definitions.
11. We eliminate references to pages with homonymic concepts assuming that this information won’t be of value for semantic integration.

The result is shown in List. 2, the data is provided in Turtle syntax (Prud’hommeaux and Carothers, 2014).[11]

```
1 @prefix     : <http://example.org/hallig-wartburg#> .
2
3 :HW a skos:ConceptScheme ;
4   skos:prefLabel "HW classification scheme"@en ;
5   xkos:numberOfLevels 9 ;
6   xkos:levels ( :HW_Level1 ... :HW_Level9 ) ;
7   skos:hasTopConcept :L_HOMME , :L_UNIVERS , ... .
8
9 :hwIdentificationScheme a rdfs:Datatype ;
10   rdfs:comment "HW concept identification scheme" ;
11   owl:oneOf (
12     "B"^^xsd:string
13     "B I k 1 cc 1"^^xsd:string
14     "B I k 1 cc 2"^^xsd:string ... ) .
15 :HW_Level1 a xkos:ClassificationLevel ;
16   xkos:depth 1 ;
17   skos:member :L_UNIVERS , :L_HOMME ,
18   :L_HOMME_ET_L_UNIVERS .
19 :L_HOMME a skos:Concept ;
20   skos:altLabel "L'homme"@fr ;
21   skos:scopeNote "L'homme"@fr ;
22   skos:notation "B"^^:hwIdentificationScheme;
23   skos:inScheme :HW ;
24   skos:topConceptOf :HW ;

27
```

[11] For the sake of brevity, we suppress (lines of) code that do not add substantial value, and standard namespaces are assumed defined the usual way, also in List. 3, 5 and 6.

```
  skos:narrower :L_HOMME_ETRE_PHYSIQUE .
:L_HOMME_ETRE_PHYSIQUE a skos:Concept ;
  skos:altLabel "L'homme, être physique"@fr ;
  skos:scopeNote "L'homme, être physique"@fr ;
  skos:notation "B I"^^:hwIdentificationScheme ;
  skos:inScheme :HW ;
  skos-thes:broaderGeneric :L_HOMME ;
  skos:narrower :LE_SEXE , :LA_RACE , ... .
:LA_VIANDE a skos:Concept ;
  skos:altLabel "La viande"@fr ;
  skos:scopeNote "La viande"@fr ;
  skos:notation "B I k 1 cc 1"^^:hwIdentificationScheme ;
  skos:inScheme :HW ;
  skos-thes:broaderGeneric :LES_ALIMENTS ;
  skos:narrower :VIANDE , :JAMBON , :LARD ... .
:MIETTE a skos:Concept ;
  skos:altLabel "miette"@fr ;
  skos:scopeNote "miette"@fr ;
  skos:inScheme :HW ;
  skos-thes:relatedPartOf :LE_PAIN_LA_PATISSERIE .
:LES_NARINES a skos:Concept ;
  skos:altLabel "les narines"@fr ;
  skos:scopeNote "les narines"@fr ;
  skos:inScheme :HW ;
  skos-thes:broaderPartitive :LE_CORPS_ET_LES_MEMBRES .
```

Listing 2: Extract of RDF data.

We have considered including the Lemon-tree vocabulary into the modelling. Lemon-tree has specifically been designed to model lexicographical thesaurus-like resources as LD, bridging SKOS and the OntoLex-*lemon* vocabulary (Stolk, 2019). Yet, for the modelling of HW, following the examples given by Lemon-tree, only SKOS and XKOS would be used, hence the advantage would not be obvious.[12] The MHDBDB has created a SKOS model of the onomasiological framework (extending HW) that structures the data.[13] However, its design differs significantly from the result of our attempt: The model excludes both the original HW identifiers and the French concept denominations. Instead, concept denominations have been translated to German and English, and they are treated as lexical terms, expressed through the SKOS property 'prefLabel'. The model expresses the relationships solely as hierarchical generic through SKOS 'broader' (not using the inverse relation 'narrower', resulting in the fact that a navigation from a top level down is not possible). In any case, it has become clear that an LOD compliant model of HW presents a desideratum in the discipline of historical linguistic data.

3.2. Towards an Ontological Model

The HW RDF/SKOS model is compliant with the LOD paradigm but it is a representation close to the book published in 1953. With the means of a KOS, it lacks of conceptual abstraction, nuanced semantic relations, and information integration for interoperability (cp. Soergel et al. (2006)). The Web Ontology Language (OWL) (Bechhofer et al., 2004) is a popular W3C recommended format to express ontologies, offering an alternative means for porting KOSs to the Semantic Web. The next step is, thus, to construct an ontological model of the HW in OWL on the basis of the RDF/SKOS model. This will allow for more expressivity and descriptiveness than offered by SKOS relations, also preparing for future extension. The result will be a lightweight ontology, i.e., an RDF document serialised in OWL, its benefit over the RDF/SKOS model being better interoperability and the potential for a extra-linguistic cross-mapping hub for the (historical) linguistic resources using HW concepts as their onomasiological architecture: A lightweight ontology based on HW provides a possibility for resources such as DAG*él*, LEI, DEAF, and MHDBDB to create instances of the HW classes.

The HW concepts meet the requirement of reflecting universal categories and the SKOS concepts (instances in SKOS) can thus be represented as classes in OWL (cp. Baker et al. (2013, 38); Kless et al. (2012b, 406-409)). This is a viable approach for creating an ontology in OWL Full but its result of course does not have inferencing qualities. Adding the expressive capabilities to allow for reasoning over the ontological model requires a re-engineering of the SKOS model into a formal ontology expressed with OWL DL, which we will discuss shortly in section 5.

The syntactic conversion from the SKOS model into OWL Full is not straightforward. The fact that thesauri-like KOSs express concept relations through basically two kinds of relationships only (hierarchical and associative) makes them underspecified from the perspective of an ontological model (Kless et al., 2012b). At the same time, the aligning of specific relationships in a thesaurus to relationships in an ontological model is not obvious and lacks of corresponding relata, in particular, associative relationships rarely find their matches (ib. 412). In this paper, we demonstrate the approach of adopting the relationships expressed by SKOS and its thesaurus extension (ib. 422): The conversion of the concepts ordered hierarchically by the generic relation into class/sub-class relations (expressed by means of RDFS 'subClassOf') (Brickley and Guha, 2014) is obvious; `skos-thes:broaderPartitive` will be preserved for the hierarchical whole-part relationship, and `skos-thes:relatedPartOf` for the associative relationship. The lexical label can be expressed through RDFS 'label', the SKOS properties 'scopeNote' and 'notation' will be preserved. We conducted a small study representing sample data of HW as an ontological model, see List. 3.

```
<rdf:RDF xmlns="https://example.org/hallig-wartburg-
 ontology#">

<owl:Ontology rdf:about="https://example.org/hallig-
 wartburg-ontology#">
 <dct:title xml:lang="en">Hallig-Wartburg Ontology
 </dct:title>
 <vann:preferredNamespacePrefix>hw
 </vann:preferredNamespacePrefix>
 <dct:description xml:lang="en">Ontology based on ...
 </dct:description>
 <owl:versionInfo rdf:datatype="http://www.w3.org/
 2001/XMLSchema#string">1.0.0
 </owl:versionInfo>
</owl:Ontology>

<!-- datatype properties -->
<owl:DatatypeProperty rdf:about="https://lod.academy/
 hw-onto/ns/hw#hwIdentificationScheme">
 <rdfs:label xml:lang="en">HW Identification Scheme
 </rdfs:label>
 <rdfs:range>
   <rdfs:Datatype>
    <owl:oneOf>...</owl:oneOf>
   </rdfs:Datatype>
 </rdfs:range>
</owl:DatatypeProperty>
```

[12] A linguistic resource could, however, use Lemon-trees's object property `isSenseInConcept` to relate a "lexical sense to a concept that captures its meaning to some extent (that is, partially or even fully)" (Stolk, 2019).

[13] We thank Peter Hinkelmanns, MHDBDB, for making the model available to us and for sharing thoughts on how to model HW in SKOS.

```
<!--  classes  -->
<owl:Class rdf:about="https://example.org/hallig-
wartburg-ontology#LA_VIANDE">
 <skos:scopeNote xml:lang="fr">La viande</skos:scopeNote>
 <skos:notation rdf:datatype="https://lod.academy/
 hw-onto/ns/hw#hwIdentificationScheme">
 B I k 1 cc 1</skos:notation>
 <rdfs:label xml:lang="fr">La viande</rdfs:label>
 <rdfs:subClassOf rdf:resource="https://example.org/
 hallig-wartburg-ontology#LES_ALIMENTS"/>
</owl:Class>
<owl:Class rdf:about="https://example.org/hallig-
wartburg-ontology#MIETTE">
 <skos:scopeNote xml:lang="fr">miette</skos:scopeNote>
 <rdfs:label xml:lang="fr">miette</rdfs:label>
 <rdfs:subClassOf rdf:resource="https://example.org/
 hallig-wartburg-ontology#HWCat"/>
 <skos-thes:relatedPartOf rdf:resource="https://example.
 org/hallig-wartburg-ontology#LE_PAIN_LA_PATISSERIE"/>
</owl:Class>
<owl:Class rdf:about="https://example.org/hallig-wartburg-
 ontology#LES_NARINES">
 <skos:scopeNote xml:lang="fr">les narines</skos:scopeNote>
 <rdfs:label xml:lang="fr">les narines</rdfs:label>
 <rdfs:subClassOf rdf:resource="https://example.org/
 hallig-wartburg-ontology#HWCat"/>
 <skos-thes:broaderPartitive rdf:resource="https://example.
 org/hallig-wartburg-ontology#LE_CORPS_ET_LES_MEMBRES"/>
</owl:Class>
</rdf:RDF>
```

Listing 3: Extract of OWL ontology (RDF/XML syntax).

4. Practical Application

With the use cases of Old Gascon *bacon* (ham), entry of DAG*él*, and of Italian *cantuccino* (a twice-baked almond biscuit), entry of LEI, we demonstrate how—through the interlinking of linguistic resources via the OntoLex-*lemon* vocabulary—the integration of a reference to a concept of the HW ontology can be integrated into an LOD resource.

Old Gascon *bacon*. The conversion of DAG*él* dictionary entries into RDF is an automated process, broadly similar to the conversion of DEAF (Tittel and Chiarcos, 2018). To automatically insert a mapping of a sense definition to the correct HW concept is straightforward, given that a reference from each sense to HW is part of the XML resource data, as shown in List. 4.

```
<m:definition>viande de porc sal&#xE9;e afin de
  la conserver</m:definition>
<m:cat-onomas cat="B I k cc 1">B I k 1 cc 1 /
  La viande</m:cat-onomas>
```

Listing 4: XML resource data of a DAG*él* entry (extract).

The content of the XML element `<cat-onomas>` can be transformed into `hw:LA_VIANDE`, to which we can refer through OntoLex-*lemon*'s object property `isConceptOf`, as shown in List. 5, l. 14.

```
1  @prefix dag: <http://dag.adw.uni-heidelberg.de/
2               lemme/> .
3  @prefix hw:  <http://example.org/hallig-wartburg-
4               ontology#> .
5
6  dag:bacon a ontolex:LexicalEntry ;
7   ontolex:sense dag:bacon_sense ;
8   ontolex:evokes dag:bacon_lexConcept ;
9   ontolex:canonicalForm dag:bacon_form .
10 dag:bacon_form a ontolex:Form ;
11  ontolex:writtenRep "bacon"@oc-x-40000006 .
12
13 dag:bacon_lexConcept a ontolex:LexicalConcept ;
14  ontolex:isConceptOf hw:LA_VIANDE ;
15  ontolex:definition "viande de porc salée afin de la
16  conserver"@fr ;
17  ontolex:lexicalizedSense dag:bacon_sense .
```

Listing 5: Minimal example of DAG*él* data (RDF/Turtle).

We point out that a finer-grained concept for the Old Gascon lexeme *bacon* is available, i.e., JAMBON (ham). However, DAG*él* only uses the numbered concepts of HW's 'Plan' (Fig. 1) and thus refers to the super-concept `LA_VIANDE`. As a consequence, a manual post-processing should include replacing `LA_VIANDE` by `JAMBON`. Please note that, in List. 5, l. 11, we use the language tag for Old Gascon `oc-x-40000006`, a shortened form that expands to `oc-x-02q35735-241050--1500` using the Web application for generating and decoding language tags at `https://londisizwe.org/language-tags/` [07-02-2020].[14]

Italian *cantuccino*. The digitisation of the LEI and its modelling as LOD is still work in progress. We can, however, show a manually created example of entry *cantuccino* (LEI 10,1458,32) in List. 6.

```
@prefix     lei: <http://www.lei-digitale.org/> .

lei:cantuccino a ontolex:LexicalEntry ;
 ontolex:sense lei:cantuccino_sense ;
 ontolex:evokes lei:cantuccino_lexConcept ;
 ontolex:canonicalForm lei:cantuccino_form .
lei:cantuccino_form a ontolex:Form ;
 ontolex:writtenRep "cantuccino"@it .

lei:cantuccino_lexConcept a ontolex:LexicalConcept ;
 ontolex:isConceptOf hw:LE_PAIN_LA_PATISSERIE ;
 ontolex:definition "un pezzetto, un ritaglio di pane
 dolce mandorlato"@it ;
 ontolex:lexicalizedSense lei:cantuccino_sense .
```

Listing 6: Minimal example of LEI data (RDF/Turtle).

HW ontology as cross-mapping hub. The integration of references to the HW ontology is a model to be followed by other resources, where word-sense units refer to the same HW concepts, thus, installing the HW lightweight ontology as a cross-mapping hub and an access point to semantic-driven, language- and word-form independent research. E.g., a database search for the string 'pâtisserie' within the sense definitions of all DEAF*él* entries produces 46 results: *friolete* f. "pâtisserie légère", *fromagie* f. "pâtisserie faite de fromage et d'œufs", etc. In DAG*él*, we find the lexeme *habanhas* m. "pâtisserie semi-sucrée à base de fèves".[15] A mapping of these lexemes to the corresponding HW concept LE_PAIN_LA_PATISSERIE could thus be integrated into the LOD versions of DEAF and DAG in an automated way, leading, in this example, to a semantically driven, extra-linguistic cross-linking of LEI, DAG, and DEAF.

5. Discussion and Future Work

In this paper, we have argued that the modelling of HW as an LOD resource is an important step towards resource integration and cross-language accessibility of historical linguistic resources. The lightweight ontology based on HW provides a model for external resources, facilitating references for semantic mapping. However, moving from

[14] ISO 639 does not provide a language code for Old Gascon and we thus follow the pattern to create a unique and decodable language tag described by Gillis-Webber and Tittel (2020).

[15] A search for the HW concept 'B I k 1 cc 2' produces 21 lexemes but is less precise, leading also to lexemes denoting flour, sieving flour, etc.

the RDF/SKOS format towards an ontology should include adding knowledge that enriches the model through additional concepts, relationships, terms, and descriptive metadata. This means adding labels in other languages, and scholastic genus–differentia definitions to help grasp the concepts, e.g., LA_VIANDE: "flesh of animals (including fishes and birds and snails) used as food" (useful resources, i.e., dictionaries, WordNet, etc., for this task need to be evaluated considering conceptualisation incongruences and translation problems [cp. Bizzoni et al. (2014) on the Ancient Greek WordNet]; a cooperation with MHDBDB seems promising in this regard). As a first step, we have published the identification scheme used in Hallig-Wartburg (as shown in List. 3), available at `https://lod.academy/hw-onto/ns/hw#`.

Re-engineering the Model into a Formal Ontology. To enable reasoning over the HW ontology (that is not possible with the OWL Full model demonstrated above) and to introduce more expressive semantic relations for this purpose requires the SKOS model to be re-engineered into a formal ontology. The disjointness condition between OWL classes and individuals (the SKOS concepts) must hold true for OWL DL, thus, any SKOS and SKOS-THES relations will need to be removed. However, to align the relationships expressed through SKOS / SKOS-THES properties with OWL DL is clearly not obvious (Keet and Artale, 2008; Kless et al., 2012b; Baker et al., 2013; Adams et al., 2015). It involves finding equivalences for hierarchical whole-part (spatial, structural, etc.) relationships, associative relationships (e.g., action and action instrument / results / participant / target / etc. (Kless et al., 2012b, 422f.), and coining custom relation properties for relating nuanced same-level and cross-level relations. Using the re-engineering of the AGROVOC thesaurus as an example (Baker et al., 2019), the cost-benefit ratio of a presumably very time-consuming task must be considered. We thus identify a feasibility analysis of (i) re-assessing the relationships expressed in the original HW resource, (ii) making them explicit and (iii) expressing them through relations in OWL as future work.

Insufficient Scope and Granularity of HW concepts. HW shows significant shortcomings that hamper an accurate semantic mapping, reducing its relevance as an extra-linguistic cross-mapping hub. The scope and granularity of HW's categories do not suffice when modelling the lexical units of an entire language: HW is little appropriate for the mapping of the so-called small words (e.g., pronouns, articles). The differentiation is inadequate: HW is primarily geared to general language and lacks any kind of technical precision, e.g., in fields like 'L'astronomie' and 'La biologie' that are reduced to one single concept, respectively.

Insufficient Possibilities for Depicting Historical Life. Regional and cultural imprints through time go hand in hand with semantic shift. The HW, like other extra-linguistic conceptualisations of the world such as DBpedia, depicts modern reality. To map Old Italian *àghila* to HW 'aigle' or DBpedia 'Eagle'[16] is straightforward. However, with language change and semantic shift, many problems arise that make the semantic mapping from a lexeme in a (medieval) historical linguistic resource to an entity of a conceptual model of the modern world difficult: (i) *things* (abstract or real) denoted by medieval words do not exist anymore, (ii) *words* are extinct and, thus, the concepts denoted by them are hard to identify in a modern world ontology, e.g., Old French *jaonoi* m. "gorse-covered terrain", DEAF J 398,30, (iii) *meanings* of words are extinct, and their modern equivalence is not obvious, e.g., Old French *jambe* f. ("leg", and also:) "post that serves as a support (for a door lintel, a mantelpiece, a vault, etc.)", DEAF J 94,15, and (iv) *meanings* have undergone semantic shift and the underlying concept is clearly different from the one of symbolized by the modern corresponding word. E.g., the veine was considered a sort of blood vessel that transports the 'nourishing blood' from the liver to each part of the body, and the sperm designated both the male and the female generative cell, etc.[17] Hence, a mapping to the modern concepts of 'veine' and 'sperm' is not possible without causing semantic discrepancies. We refer to this circumstance as the *Historical Semantic Gap*. Khan et al. (2014) address the issue of modelling semantic shift with extending the OntoLex-*lemon* vocabulary by adding a time interval to capture different concepts of one lexeme through time. This approach is a major enhancement from the point of view of historical linguistics. However, it does not solve the problem of semantic mapping to an extra-linguistic conceptual model where the historical concept is not represented.

To stabilise HW's role as an onomasiological reference system for historical (linguistic) resources, it must be elaborated in two ways: The net of concepts must be refined and concepts with historically appropriate content must be added. We call the latter process the *historicisation* of HW. To prepare for a future extension towards historicised content, we foresee a class `HistCat` and a symmetric (object) property `hasModernCounterpart`, cf. List. 7.

```
<owl:SymmetricProperty rdf:about="https://example.org/
 hallig-wartburg-ontology#hasModernCounterpart">
 <rdfs:label xml:lang="en">has modern counterpart
 </rdfs:label>
</owl:SymmetricProperty>

<owl:Class rdf:about="https://example.org/
 hallig-wartburg-ontology#HistCat">
 <rdfs:label xml:lang="en">historicised concept
 </rdfs:label>
</owl:Class>
```

Listing 7: Added property and class to HW ontology.

HW presents few categories that mirror the specification of historical times: Only four concepts include the notion of 'ancient', e.g. 'Les armes anciennes' (early weapons, next to 'Les armes modernes') and 'Les bâtiments de guerre anciens' (early warships, next to 'Les bâtiments de guerre modernes'). With the added class and object property, e.g., the class `LES_ARMES_ANCIENNES` can be defined a subclass of `HistCat` and refer to `LES_ARMES_MODERNES` through the property `hasModernCounterpart`. This would, thus, support the use of HW as an onomasiological framework by both historical and modern resources.

[16] `http://dbpedia.org/page/Eagle` [10-02-2020].

[17] DEAF*pré* VEINE[1], `https://deaf-server.adw.uni-heidelberg.de/lemme/veine1`; ESPERME `.../lemme/esperme` [25-02-2020].

6. Acknowledgements

The work of Frances Gillis-Webber was financially supported by Hasso Plattner Institute for Digital Engineering.

7. Bibliographical References

Adams, D., Jansen, L., and Milton, S. (2015). A content-focused method for re-engineering thesauri into semantically adequate ontologies. *Semantic Web*, 09.

Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., and Summers, E. (2013). Key choices in the design of Simple Knowledge Organization System (SKOS). *Journal of Web Semantics*, 20:35 – 49.

Baker, T., Whitehead, B., Musker, R., and Keizer, J. (2019). Global agricultural concept space: lightweight semantics for pragmatic interoperability. *npj Science of Food*, 3, 12.

Baldinger, K. (1959). *s.v.* Romanistik. *Deutsche Literaturzeitung*, 80:1090–1093.

Baldinger, K. (1975 to 2005). *Dictionnaire onomasiologique de l'ancien occitan – DAO* (fondé par Kurt Baldinger, rédigé par Inge Popelar, puis Bernard Henschel, puis Nicoline Hörsch/Winkler et Tiana Shabafrouz). Niemeyer [Heidelberger Akademie der Wissenschaften / Kommission für das Altokzitanische und Altgaskognische Wörterbuch], Tübingen.

Baldinger, K. (since 1971). *Dictionnaire étymologique de l'ancien français – DEAF*. Presses de L'Université Laval / Niemeyer / De Gruyter, Québec/Tübingen/Berlin. [continued by Frankwalt Möhren, and Thomas Städtler; DEAF*él*: https://deaf-server.adw.uni-heidelberg.de].

Baldinger, K. (since 1975). *Dictionnaire onomasiologique de l'ancien gascon – DAG* (fondé par Kurt Baldinger, dirigé par Inge Popelar, puis Nicoline Hörsch / Winkler et Tiana Shabafrouz, sous la direction de Jean-Pierre Chambon, puis Martin Glessgen). De Gruyter [Heidelberger Akademie der Wissenschaften / Kommission für das Altokzitanische und Altgaskognische Wörterbuch], Tübingen / Berlin.

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., and Stein, L. A. (2004). OWL Web Ontology Language. Reference. W3C Recommendation 10 February 2004. URL: https://www.w3.org/TR/2004/REC-owl-ref-20040210/ [accessed: 09-02-2020].

Bevans, C. (1941). *The Old French vocabulary of Champagne. A descriptive study based on localized and dated documents*. University of Chicago Libraries, Chicago.

Bizzoni, Y., Boschetti, F., Diakoff, H., Gratta, R. D., Monachini, M., and Crane, G. (2014). The Making of Ancient Greek WordNet. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1140–1147, Reykjavik, Iceland. European Language Resources Association (ELRA).

Brickley, D. and Guha, R. (2014). RDF Schema 1.1. W3C Recommendation 25 February 2014. URL: https://www.w3.org/TR/rdf-schema/ [accessed: 13-02-2020].

Chiarcos, C., McCrae, J., Cimiano, P., and Fellbaum, C. (2013). Towards Open Data for Linguistics: Lexical Linked Data. In Alessandro Oltramari, et al., editors, *New Trends of Research in Ontologies and Lexical Resources: Ideas, Projects, Systems*, pages 7–25. Springer, Berlin, Heidelberg.

Christmann, H. and Böckle, K. (1983). Bespr. von Schwake, Der Wortschatz des *Cliges*. *Zeitschrift für romanische Philologie*, 99:397–403.

Cimiano, P., McCrae, J. P., and Buitelaar, P. (2016). Lexicon Model for Ontologies: Community Report, 10 May 2016. Final Community Group Report 10 May 2016. URL: https://www.w3.org/2016/05/ontolex/ [accessed: 10-02-2020].

Cyganiak, R., Wood, D., and Lanthaler, M. (2014). RDF 1.1. concepts and abstract syntax: W3C recommendation 25 February 2014. URL: https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/ [11-02-2020].

Cyganiak, R., Gillman, D., Grim, R., Jaques, Y., and Thomas, W. (2017). An SKOS extension for representing statistical classifications, ed. F. Cotton, Unofficial Draft, 1 January 2017. URL: http://www.ddialliance.org/Specification/XKOS/1.0/OWL/xkos.html [accessed: 07-02-2020].

Davidson, G. (2002). *Roget's thesaurus of English words and phrases*. (150th anniversary edition). Penguin Books, London.

de Man, L. (1956). Bijdrage tot een systematisch glossarium van de Brabantse oorkondentaal. Leuvens Archief van circa 1300 to 1550. *Deel*, I.

Dornseiff, F. (1934). *Der deutsche Wortschatz nach Sachgruppen*. de Gruyter, Berlin.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Gillis-Webber, F. and Tittel, S. (2020). A Framework for Shared Agreement of Language Tags beyond ISO 639. In *Proceedings of LREC 2020 [accepted paper]*, N.N.

Glessgen, M. and Tittel, S. (2018). Le *Dictionnaire électronique de l'ancien gascon* (DAG*él*). In Roberto Antonelli, et al., editors, *Atti del XXVIII Congresso internazionale di linguistica e filologia romanza (Roma, 18-23 luglio 2016)*, volume 1, pages 805–818. Société de Linguistique Romane / Éditions de linguistique et de philologie ELiPi, Bibliothèque de Linguistique Romane 15,1.

Glessgen, M. (since 2014). Dictionnaire de l'ancien gascon – DAGél. (en collaboration avec Sabine Tittel) URL: https://dag.adw.uni-heidelberg.de/ [accessed: 02-02-2020].

Hallig, R. and von Wartburg, W. (1963). *Begriffssystem als Grundlage für die Lexikographie / Système raisonné des concepts pour servir de base à la lexicographie*. Akademie-Verlag, Berlin. [first edition 1952].

Helou, M., Jarrar, M., Palmonari, M., and Fellbaum, C. (2014). Towards building lexical ontology via cross-language matching. In *GWC 2014: Proceedings of the 7th Global Wordnet Conference*, pages 346–354.

Hinkelmanns, P. (2019). Mittelhochdeutsche Lexikographie und Semantic Web. Die Anbindung der 'Mittelhochdeutschen Begriffsdatenbank' an Linked Open Data. *Das Mittelalter*, 24(1):129–141.

International Organization for Standardization. (2011). International Standard ISO 25964-1:2011, Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval, Part 2: Interoperability with other vocabularies. URL: https://www.iso.org/standard/53657.html.

Kay, C. (2009). *Historical thesaurus of the Oxford English dictionary*. Oxford University Press, Oxford.

Keet, C. and Artale, A. (2008). Representing and reasoning over a taxonomy of part-whole relations. *Applied Ontology*, 3:91–110.

Keller, H.-E. (1953). *Etude descriptive sur le vocabulaire de Wace*. Akad. der Wiss. Berlin, Veröffentl. Inst. für Rom. Spr.wiss. 7, Berlin.

Khan, F., Frontini, F., and Boschetti, F. (2014). Using lemon to Model Lexical Semantic Shift in Diachronic Lexical Resources. In *Proceedings of the 3rd Workshop on Linked Data in Linguistics: Multilingual Knowledge Resources and Natural Language*, Reykjavik, Iceland.

Kless, D., Jansen, L., Lindenthal, J., and Wiebensohn, J. (2012a). A method for re-engineering a thesaurus into an ontology. *Frontiers in Artificial Intelligence and Applications*, DOI 10.3233/978-1-61499-084-0-133:133–146.

Kless, D., Milton, S., and Kazmierczak, E. (2012b). Relationships and Relata in Ontologies and Thesauri: Differences and Similarities. *Applied Ontology*, 7:401–428, 11.

Miles, A. and Bechhofer, S. (2009). SKOS Simple Knowledge Organization System reference: W3C recommendation 18 August 2009. URL: https://www.w3.org/TR/2009/REC-skos-reference-20090818/ [accessed: 07-02-2020].

Miles, A. and Brickley, D. (2004). SKOS Extensions Vocabulary Specification. URL: www.w3.org/2004/02/skos/extensions/spec/2004-10-18.html [20-02-2020].

Nannini, A. (in progress). La mappatura semantica del Lessico Etimologico Italiano (LEI). Doctoral thesis.

Pfister, M. (since 1979). *Lessico Etimologico Italiano – LEI*. Reichert, Wiesbaden. [2001– together with W. Schweickard, 2018– W. Schweickard together with E. Prifti].

Prifti, E. (2019). Lo stato della digitalizzazione del LEI. Un resoconto. In Lino Leonardi et al., editors, *Italiano antico, italiano plurale. Testi e lessico del Medioevo nel mondo digitale*, page [in print]. N.N., Firenze.

Prud'hommeaux, E. and Carothers, G. (2014). RDF 1.1 Turtle: Terse RDF Triple Language. W3C Recommendation, 25 February 2014. URL: http://www.w3.org/TR/turtle/ [accessed: 07-02-2020].

Renders, P. (2015). *L'informatisation du Französisches Etymologisches Wörterbuch. Modélisation d'un discours étymologique*. ELIPHI, Strasbourg.

Renders, P. (2019). Integrating the Etymological Dimension into the Onto-Lex Lemon Model: A Case of Study. In *Electronic lexicography in the 21st century (eLEX 2019). Book of Abstracts*, pages 71–72.

Schmidt, K. (1980). *Begriffsglossare und Indices zu Ulrich von Lichtenstein*. Indices zur deutschen Literatur 14/15. Kraus International Publications, München.

Schmidt, K. (1988). Der Beitrag der Begriffsorientierten Lexikographie zur systematischen Erfassung von Sprachwandel und das Begriffswörterbuch zur Mhd. Epik. In Wolfgang Bachofer, editor, *Mittelhochdeutsches Wörterbuch in der Diskussion. Symposion zur mittelhochdeutschen Lexikographie, Hamburg, Oktober 1985*, pages 35–49, Tübingen. Niemeyer.

Schmidt, K. (1993). *Begriffsglossar und Index zu Ulrich von Zatzikhoven Lanzelet*. Indices zur deutschen Literatur 25. Niemeyer, Tübingen.

Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., and Katz, S. (2006). Reengineering Thesauri for New Applications: the AGROVOC Example. *Journal of Digital Information*, 4(4).

Stempel, W.-D. (1996 to 2013). *Dictionnaire de l'occitan médiévale – DOM*. Niemeyer / De Gruyter, Tübingen/Berlin. [continued by Maria Selig; electr. version: http://www. dom-en-ligne.de].

Stolk, S. (2019). Lemon-tree. Document 30 March 2019. Latest editor's draft. URL: https://ssstolk.github.io/onto/lemon-tree/index.html [accessed: 07-02-2020].

Tancke, G. (1997). Note per un avviamento al Lessico Etimologico Italiano (LEI). In Günter Holtus, et al., editors, *Italica et Romanica. Festschrift für Max Pfister zum 65. Geburtstag*, pages 457–487. Niemeyer, Tübingen.

Tittel, S. and Chiarcos, C. (2018). Historical Lexicography of Old French and Linked Open Data: Transforming the Resources of the *Dictionnaire étymologique de l'ancien français* with OntoLex-Lemon. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). GLOBALEX Workshop (GLOBALEX-2018), Miyazaki, Japan, 2018*, pages 58–66, Paris (ELRA).

Tittel, S. (in progress). Integration von historischer lexikalischer Semantik und Ontologien in den Digital Humanities. Habilitation thesis.

von Wartburg, W. (since 1922). *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes – FEW*. ATILF. [continued by O. Jänicke, C.T. Gossen, J.-P. Chambon, J.-P. Chauveau, and Yan Greub].