Semantic Annotation in the Project
"Open Access Database 'Adjective-Adverb Interfaces' in Romance"

Christopher Pollin, Gerlinde Schneider, Katharina Gerhalter, Martin Hummel

Centre for Information Modeling & Institute for Romance Studies, University of Graz

Elisabethstraße 59/III, 8010 Graz, Merangasse 70, 8010 Graz

{christopher.pollin, gerlinde.schneider, katharina.gerhalter, martin.hummel}@uni-graz.at

Abstract

This paper describes the creation, the annotation process and the model of the Open Access Database 'Adjective-Adverb Interfaces in Romance' (AAIF) project, with its approach to the creation of a domain-specific ontology. In order to make research data accessible, interoperable, extensible, and transferable, data is annotated in TEI/XML, formalized and enriched with RDF and its conceptual data model is stored in and published via the GAMS digital repository. This produces semantically-enriched, annotated multilingual research data that allows retrieval across heterogeneous corpora. The annotation model expressed in the ontology is offered for further reuse.

Keywords: annotated data, open access, semantic enrichment, ontology based, RDF, TEI, GAMS

1. Introduction

Annotation has always played a crucial role in humanities textual scholarship as well as in linguistic research; increasing with the development of digital methods and tools. For this reason, research data in these areas very often consist of annotated text in various form. The taxonomy TaDiRAH[1] describes the digital research practice of annotating as the '*activity of making information about a digital object explicit by adding, e.g., comments, metadata or keywords* [...]'. Schöch (2013) distinguishes between two types of data in the context of research in the humanities: big data and smart data. The former is unstructured, implicit, large in volume, and varied in form. The latter is semi-structured or structured, explicit, small in scale and of limited heterogeneity. According to these criteria, annotated linguistic corpora are smart data.

The data the project *Open Access Database 'Adjective-Adverb Interfaces in Romance'* (AAIF)[2] deals with are complex linguistic annotations. The project aims to survey the possibilities and challenges of open data and open access with regard to linguistic research data. The project focuses on the interoperability and accessibility of data, with particular respect to reusability in the sense of the FAIR[3] Data Principles. Topics discussed by this paper include data creation, annotation, data preservation and publication process by means of the GAMS[4] repository and accessibility via a search interface. These aspects are tied together by semantic technologies, using an ontology-based approach that is relevant to other domains of digital data. In the following, we want to investigate the application of semantic technologies to meet the challenges described above.

2. Project and Challenges

Funding authority policy, as well as a re-thinking in research communities, has led to a situation where more and more richly annotated research data is becoming openly accessible and integrable. AAIF, a project within the *Austrian Science Fund* programme *Open Research Data Pilot,* focuses on how to publish linguistically annotated data to make it reusable within and outside of the domain while making the underlying annotation model available.

The project builds upon the work of the Research group on *Interfaces of Adjective and Adverb in Romance*.[5] In the course of the project, different corpora, each annotated with respect to the complex relations between the word classes of adjective and adverb in Romance languages, are going to be integrated to one comprehensive database. This will enable querying across corpora and languages and thus allow for cross-linguistic generalizations. The expandability of the system for new data has to be considered during the whole process.

As the corpora were compiled and annotated in response to diverse, very specific research questions within the domain, the degree and emphasis of the annotation varies. Adjective-adverb phrases can have a very flat annotation, where for example, only one adverb and verb are marked and lemmatized; others are very extensively annotated with semantic and morphosyntactic information. Additionally, the applied annotation model has been developed further over time. With more diverse research questions and a deeper understanding of the field, some categories were added and changed. All this results in data that is annotated very heterogeneously and will remain so in the future. These issues significantly complicate the endeavor of integrating all data into one database while concurrently preserving the rich annotation each corpus holds.

---

[1] Taxonomy of Digital Research Activities in the Humanities, http://tadirah.dariah.eu/vocab

[2] https://adjective-adverb.uni-graz.at/en/research/projects/open-access-database

[3] https://www.force11.org/group/fairgroup/fairprinciples

[4] http://gams.uni-graz.at

[5] https://adjective-adverb.uni-graz.at

Other challenges are the multilingual character of the data and providing a search interface for a broad variety of selections and combination of categories.

3. Related Work

There have been considerable efforts to increase interoperability across linguistic resources and between NLP tools using semantic technologies. Establishing a *Linguistic Linked Open Data Cloud* (LLOD)[6] as a means for sharing these resources was an important step in this endeavour. Of particular note is the development of an OWL/DL-based reference model to formalize the mapping between annotation models within the framework of the Ontologies of Linguistic Annotations (Chiarcos et al., 2016).[7]

The OLiA ontologies serve as a top-level knowledge base for annotation terminology for linguistic phenomena and provide a detailed terminological reference model. They were developed as part of an infrastructure for the sustainable maintenance of linguistic resources; their primary fields of application include the formalization of annotation schemes and concept-based querying over heterogeneously annotated corpora (Chiarcos & Sukhareva, 2015). Hellmann et al. (2013) propose the NLP Interchange Format (NIF)[8], a framework that uses the richness of Linked Data technologies to foster interoperability between NLP tools, resources and annotation. NIF uses standardized URI schemas, REST interfaces, and RDF/OWL-based ontologies to connect heterogeneous but interoperable applications and resources across the web.

4. Data and Annotation

Adjective-adverb interfaces are specified as linguistic phenomena related to Romance adjectives with adverbial functions. For example, the use of adjective-adverbs in Spanish such as *volar alto* 'to fly high' or *ver claro* 'to see clear', discourse markers such as *cierto* 'true' and adverbial prepositional phrases like *de seguro* 'certainly', *en serio* 'seriously', *a malas* 'badly, in bad terms'.

To classify the various functions and meanings of adjective-adverbs an in-depth morphosyntactic as well as syntactic and semantic classification is used and reflected in the annotation of the respective data. Manual lemmatization is used to unify orthographic variation as well as inflected forms and enhance search mechanisms.

Research and data collection focus on historical and present-language records of adjective-adverb interfaces.

The corpus of the *Dictionnaire historique de l'adjectif-adverbe* (Hummel and Gazdik in preparation) has been available as a database since 2005 and contains 13569 entries (619101 word tokens) from the 11th to the 20st century. The corpus was compiled from examples located in the *Frantext*[9] Corpus and the *Corpus of the Dictionnaire du Moyen Français*[10]*s*. Annotation focuses on the verb phrase *verb + adjective-adverb* (e.g. *voir grand* 'to think big') which has been tagged according to the following criteria: syntax (relative order of verb and adverb), verb data (lemma, reflexive or non-reflexive use of verb, coordination), and attribute data (lemma, morphological form, number of syllables, coordination and extension). Around 4500 examples from present-day internet communication will be added to this database (=*IC*).

The corpus used for the *Sintáxis Histórica III*-chapter "Adjetivos Adverbiales" (=*SH3*) (Hummel, 2014) contains 1262 entries (92887 word tokens) with examples from the 13th to 21st century. Examples were collected by reading whole texts to locate appearances of adjective-adverb interfaces. Annotation takes into account a wider spectrum of adjective-adverbs, including prepositional phrases and inflected adverbs, such as in the following example. The adjective-adverb *alto* shows plural-concordance with the subject of the sentence *los fumos*, although it still modifies the verb *subir*:

(1)   [...] este pujamiento dell agua que fuera tanto en alto porque tan *altos* subieran los fumos de los sacrificios que los de Caím fizieran a los ídolos (1252-1284; Alfonso X; General Estoria. Primera Parte; p. 55, SH3)

A second corpus on the diachrony of Spanish compiled from records discovered in the *Corpus del Diccionario Histórico* (=*CDH*)[11] contains 2284 entries (82538 word tokens) from the 13th to 21st century. Focus is set on modal adverbs such as *ver claro* 'to see clearly', illustrated in the following example:

(2)   Cuando usted habla de la política del Ejército, hay algo que no veo *claro*. (1967, Viñas, David; Los hombres de a caballo; CDH)

Both corpora are tagged according to syntax (position), morphology (e.g. inflection), and semantics (e.g. semantic target).

More corpora are in preparation to be added to the database during the project, such as a corpus based on Latin American Spanish examples from 16th to 19th century discovered in the *Corpus Diacrónico y Diatópico del Español de América* (=*CORDIAM)*. Additionally, data from individual investigations on limited sets of adjective-adverbs, such as Spanish *justo* and *cabal* (Gerhalter, 2018), and linguistic data concerning the usage of adjective-adverbs in southern Italian varieties (Ledgeway, 2017) as well as data especially for the Old Romanian period and the 19th century will be added to the database.

---

[6] http://linguistic-lod.org

[7] http://purl.org/olia

[8] http://persistence.uni-leipzig.org/nlp2rdf

[9] http://www.frantext.fr

[10] http://www.atilf.fr/dmf/; version DMF1, 2003

[11] http://www.rae.es/recursos/banco-de-datos/cdh

5. Annotation model

In the course of numerous projects, the research group has developed further the annotation model used for the analysis of adjective-adverb interfaces. The model is elaborated and has proven useful for its application as well as for extensibility. It is broken down to the concepts and features considered necessary to support research in the domain.

The corpora already in the database are available in the form of TEI/XML. Although its implementation is a legacy from earlier projects and poses some weaknesses, the model expressed in TEI is useful to make annotations on a syntactic level. The core of the analysis is represented by an entry which consists of a sentence *<tei:s>* in which the context relevant to the research question is separately marked as a phrase *<tei:phr type="syntagm">*. Within these, all relevant components are encoded as *<tei:w>*, with respective annotation in the attributes. The morphosyntactic and semantic information annotating the respective tokens are collected as a character sequence in the attribute *function*.

The following XML snippet shows a comprehensive annotation of the abovementioned Example 1, taken from the *SH3* corpus:

```
<s>e dize maestre Pedro que este pujamiento dell agua
        que fuera tanto en alto porque
  <phr type="syntagm">tan
    <w type="adverb" lemma="alto"
        function="apvmln">altos</w>
    <w type="verb" lemma="subir"
        function="i">subieran</w>
    <w type="subject" function="mp">los fumos</w>
  de los sacrificios</phr>
que los de Caím fizieran a los ídolos, e que se lavasse de
la suziedat d'aquellos fumos ell aire.</s>
```

Examining the annotation of the adverb in the example, the attribute *function* contains six character positions (*apvmln*) that provide information about (1) *Morphosyntactic structure*, taking into account the formal structure of the adverbial (Adjective-adverb, in this case), (2) *Inflection* (Plural, for the plural suffix in *altos*), (3) Attribution target for the syntactic scope of the adverbial (Verb, which is *subieran*), (4) *Modification* of the adverb (Modified, by *tan*), (5) *Semantic classification* (Location) and (6) Reduplication of the same type, such as *claro, claro* (No, in this case). Position (0) indicates the case of *Coordination* such as *claro y alto* and is not set in this case. Example categories and features of the annotation model are shown on the case of adverbs in Figure 1.

The following entry from the corpus of the *Dictionnaire historique de l'adjectif-adverbe* illustrates the heterogeneity of the data when it comes to annotation:

(3)   Deffendez vous et parler *droit*, En verité n'a point de honte, Deshonneur vous en adviendroit, De faulx rappors ne tenez compte. (1432, Regnier, Jean; Les Fortunes et adversitez)

```
 <s><interpGrp><interp xml:lang="fr">parler
directement, avec franchise, honnêteté
</interp></interpGrp>
    <phr type="syntagm">Deffendez vous et
      <w type="verb" lemma="parler">parlez</w>
      <w type="adverb" n="syllables:1" lemma="droit"
        function="auvn">droit</w>
    </phr> En verité n'a point de honte, Deshonneur vous
en adviendroit, De faulx rappors ne tenez compte.</s>
```

Compared to Example 1b the subject of the phrase is not tagged at all, while the adverb holds additional information on its number of syllables. The adverb is only tagged for its (1) *Morphosyntactic structure* (Adjective), (2) *Inflection* (Uninflected), (3) *Attribution target* (Verb) and (4) *Modification* (Not modified). Positions (5) *Semantic classification* and (6) *Reduplication* are not considered. Information on the meaning of the phrase is added in a separate *<tei:interp>* tag.

6. Long-term data preservation and publication

GAMS serves as digital infrastructure for storing and publishing the data. It is an asset management system for the humanities and serves the purpose of administration, publication and long-term preservation of digital resources (Steiner/Stigler, 2018). It is based on the open source repository software FEDORA-Commons. Using specific content models and services for varying data streams, data can be stored and disseminated for public use. The data can be accessed as HTML, as archival data in XML, and via various APIs. In the course of the project, an object model for linguistic corpus data will be implemented offering corpus relevant methods, formats and programming interfaces such as those defined by the CLARIN ERIC.[12] At the moment the corpus data is available and archived in the introduced TEI/XML format, but is prepared to be obtainable as TCF[13] data and in the TreeTagger[14] format.

GAMS implements a disseminator for RDF data via the triplestore Blazegraph[15]. The AAIF project uses this disseminator to provide access to the corpora via RDF. A GAMS-specific object for ontologies is in use representing the annotation model. An object that allows SPARQL queries and fulltext search can be used as a basis for more in-depth retrieval functionalities.

---

[12] https://www.clarin.eu

[13] https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/The_TCF_Format

[14] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

[15] https://www.blazegraph.com

1.	Morphosyntactic Structure	adjective	a
		derived -*mente*	m
		derived -*este*	e
		derived -*ly*	l
		derived -*iş*	i
		derived -*ul*	u
2.	Inflection *(inaudible-audible for French; specific categories for other languages)*	uninflected /m.sg.	u
		audible inflection	a
		inaudible inflection	i
		Feminine Singular	f
		Neuter Singular	n
		Masculine Plural	p
		Feminine Plural	x
		Neuter Plural	z
3.	Attribution target	Verb	v
		Verb + Subject	s
		Verb + Object	o
		Sentence	S
		Adverb	A
		Adjective	a
		noun or syntagma without verb-reference	n
		other	O
4.	Modified	no	n
		yes *(muy, tan, très...)*	m
5.	Semantic classification	Manner	m
		Quantity/Intensity	q
		Time	t
		Location	l
		Discourse	d
		Specification	s
		Undefined/other	u
6.	Reduplication	no	n
		yes, repeated	r
extra	Coordination	*two coordinated Adverbs, tag both*	C

Figure 1: AAIF Annotation model for the category of Adverbs

7. Ontology-based Approach

The fundamental concept behind the Semantic Web is to have self-descriptive, interconnected and structured data. In this sense, ontologies are machine-readable knowledge representations for defining standardized and conceptual data models for data integration, information retrieval or common knowledge bases. Breitmann et al. (2007) summarize an ontology as "*a formal, explicit specification of a shared conceptualization*", clearly building upon the notion of Gruber (1993) to develop ontologies for knowledge-sharing purposes.

A domain-specific ontology, seen as a conceptual data model, is flexible, reusable and interoperable on the one hand, and promotes reuse and a shared understanding for users, on the other (Yi, 2008). These advantages offered by ontologies suit well to the challenges of the current project. To provide a conceptual model for the semantic data, the established annotation model is updated and expressed in a domain-specific RDFs ontology. Figure 2 shows a work-in-progress version of the modeling.[16]

The ontology, developed in an iterative and collaborative process, serves to make the data retrievable and to operate a parameterizable query interface. Furthermore, it is used as a reference model within the research group as well as for other research on adjective-adverb interfaces. Linking to the OLiA reference model will provide an essential basis for interoperability.
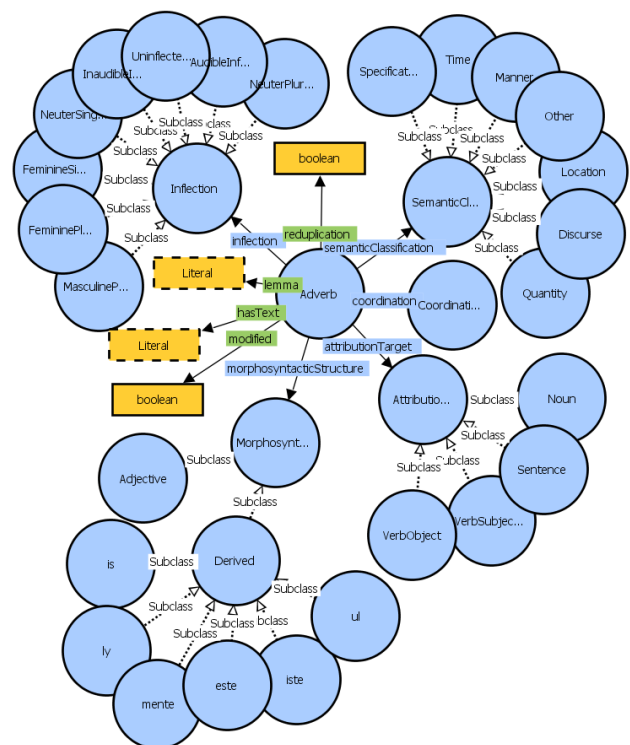

Figure 2: Representation of aaif:Adverb in the AAIF-ontology

Pollin and Vogeler (2017) describe an ingest scenario for the GAMS repository for the semantic enrichment of TEI/XML entities and structures with RDF data, based on a domain-specific RDFs data model. RDFs, seen as conceptual data model, provides interoperability through linking to top-level-ontologies. Based on this workflow, the information from the TEI data is extracted, assigned URIs and transformed into an RDF representation.[17] Using RDF allows for explication of syntactic as well as morphological structures. The following example shows the identical sentence with its annotation in RDF/XML:

```xml
<aaif:Entry rdf:about="#Entry-274">
 <aaif:phrase  rdf:resource="#Entry-274-Phrase-1"/>
 <gams:XMLContent rdf:parseType="XMLLiteral">
  <phr type="syntagm">tan
   <w type="adverb" lemma="alto" function="apvmln">
   altos</w> <w type="verb" lemma="subir"
    function="i">subieran</w>
   <w type="subject" function="mp">los fumos</w>
   de los  sacrificios
  </phr>
 </gams:XMLContent>
</aaif:Entry>
```

The class aaif:Entry represents the documentary reference including the context of every aaif:Phrase. For storing the fulltext index and the TEI/XML fragment of every phrase, the project-independent data properties gams:textualContent and gams:XMLContent are used.[18] The connections to all listed phrases are represented with the object property aaif:phrase. This allows to reference multiple aaif:Phrase to one aaif:Entry.

```xml
<aaif:Phrase rdf:about="#Entry-274-Phrase-1">
 <aaif:subject
    rdf:resource="#Entry-274-Phrase-1-Subject-1"/>
 <aaif:verb rdf:resource="#Entry-274-Phrase-1-Verb-1"/>
 <aaif:adverb
    rdf:resource="#Entry-274-Phrase-1-Adverb-1"/>
</aaif:Phrase>
```

Every aaif:Phrase gathers at least one aaif:Adverb and can optionally link to a aaif:Subject, aaif:Verb or a aaif:Preposition. Data properties like aaif:text or aaif:lemma are used to store the actual string and the normalized lemma.

```xml
<aaif:Subject rdf:about="#Entry-274-Phrase-1-Subject-1">
 <aaif:text>los fumos</aaif:text>
 <aaif:genus rdf:resource="/o:aaif.ontology#Masculine"/>
 <aaif:numerus rdf:resource="/o:aaif.ontology#Plural"/>
</aaif:Subject>
```

The grammatical case and grammatical number of the subject *los fumos* are described using aaif:genus and aaif:numerus. These object properties point to concepts in the domain-specific AAIF-ontology.

```xml
<aaif:Verb rdf:about="#Entry-274-Phrase-1-Verb-1">
 <aaif:text>subieran</aaif:text>
 <aaif:lemma>subir</aaif:lemma>
 <aaif:syntacticConstruction
    rdf:resource="/o:aaif.ontology#Intransitive"/>
</aaif:Verb>
```

In this example of a aaif:Verb, the syntactic construction is set to intransitive.

```xml
<aaif:Adverb rdf:about="#Entry-274-Phrase-1-Adverb-1">
 <aaif:text>altos</aaif:text>
 <aaif:lemma>alto</aaif:lemma>
```

```xml
 <aaif:morphosyntacticStructure
    rdf:resource="/o:aaif.ontology#Adjective"/>
 <aaif:inflection
    rdf:resource="/o:aaif.ontology#MasculinePlural"/>
 <aaif:attributionTarget
    rdf:resource="/o:aaif.ontology#Verb"/>
 <aaif:modified>true</aaif:modified>
 <aaif:semanticClassification
    rdf:resource="/o:aaif.ontology#Location"/>
 <aaif:reduplication>false</aaif:reduplication>
</aaif:Adverb>
```

The annotation of the adverb that was implicitly represented by a character sequence in the TEI model is now made explicit by RDF triples:
(1) aaif:morphosyntacticStructure: *#Adjective*,
(2) aaif:inflection: *#MasculinePlural*,
(3) aaif:attributionTarget: *#Verb*,
(4) aaif:modified: true ,
(5) aaif:semanticClassification *#Location*,
(6) aaif:reduplication: false.

8.   Data retrieval

The following SPARQL query is used to analyze the occurrence of adverb inflection. According to standard grammar, adverbs are uninflected. However, in the older texts adverbs often appear in an inflected form, as in the example *altos subieran los fumos.* This phenomenon is very interesting from a linguistic point of view, as it is not in line with the norm. It also shows very well the interfaces between the word classes of adjectives, which in Romance languages are systematically inflected, and adverbs, which are usually uninflected.

It queries all adverbs that have the morphosyntactic structure *Adjective*, a *Masculine Plural* or *Feminine Plural* inflection and the attribution target *Verb*.

```sparql
PREFIX bds:   <http://www.bigdata.com/rdf/search#>
PREFIX gams:  <https://gams.uni-graz.at/o:gams-ontology/#>
PREFIX aaif:  <https://gams.uni-graz.at/o:aaif.ontology#>

SELECT ?Adverb_text ?Adverb_lemma ?Verb_text
?Verb_lemma ?Entry_text
{
  ?Entry gams:isMemberOfCollection
        <https://gams.uni-graz.at/o:aaif.sh3>;
     aaif:phrase ?Phrase;
     gams:textualContent ?Entry_text;
     gams:XMLContent ?XMLContent.

 ?Phrase aaif:adverb ?Adverb.
 ?Adverb aaif:text ?Adverb_text;
         aaif:lemma ?Adverb_lemma.

 OPTIONAL{?Phrase aaif:verb/aaif:text ?Verb_text.
          ?Phrase aaif:verb/aaif:lemma ?Verb_lemma.}

 ?Adverb aaif:morphosyntacticStructure
   <https://gams.uni-graz.at/o:aaif.ontology#Adjective>.

{?Adverb aaif:inflection
   <https://gams.uni-graz.at/o:aaif.ontology#MasculinePlural>.}
```

```
  UNION {?Adverb aaif:inflection
    <https://gams.uni-graz.at/o:aaif.ontology#FemininePlural>.}

?Adverb  aaif:attributionTarget
    <https://gams.uni-graz.at/o:aaif.ontology#Verb>.
}
```

The search result returns a list of adverbs, verbs and their entries.[19] Such queries can help to work on comparable research questions and can be run on multiple corpora with different languages and heterogenous initial data. We made the experience that Blazegraph has a good performance, when it comes to such queries.

9.   Conclusion

Within the AAIF research group, appearances of adjective-adverb interfaces in Romance languages are annotated regarding their morphosyntactic and semantic structures. This results in "smart data" with different levels of expressiveness. The data comprises multilingual corpora including several Spanish and French Corpora, in addition to other Romance languages (Italian, Romanian and Portuguese) which will be integrated. All annotations are derived from a common data model, which serves as basis for the current TEI/XML markup. Data is published and made openly accessible through the GAMS repository. Projects in the context of Linked Linguistic Data employ semantic technologies to achieve interoperability between linguistic resources and tools; here, semantic enrichment is used to integrate the datasets of the AAIF project and make them retrievable and discoverable across corpora. A domain-specific ontology is being developed to serve as a conceptual model for the analysis of adjective-adverb relations.

The annotation model developed by AAIF can be used as a reference for similar research questions. As the model is flexible and reusable, different kinds of corpora can be annotated according to it and integrated into the database. The result is a comprehensive resource that makes it possible to explore the domain—as well as other aspects of Romance linguistics—in greater depth. Having a dataset holding very accurate annotation on the semantic relationship of certain syntactic entities may also facilitate automatic language processing.

Future project goals may include linking to lexicographic databases and enriching the data on the level of word meaning.

10.   References

Breitmann K. et al. (2007). Semantic web: concepts, technologies and applications. London: Springer Science & Business Media.

Chiarcos, C. & Sukhareva, M. (2015). Olia–ontologies of linguistic annotation. Semantic Web, 6. Jg., Nr. 4, pp. 379-386.

Chiarcos, C., Fäth, C. & Sukhareva, M. (2016). Developing and Using the Ontologies of Linguistic Annotation (2006-2016). In J. P. McCrae et al. (Eds.), Proceedings of the LREC Workshop "LDL 2016 - 5th Workshop on Linked Data in Linguistics".

Gerhalter, K. 2018. Paradigmas y polifuncionalidad. La diacronía de *preciso / precisamente*, *justo / justamente*, *exacto / exactamente* y *cabal / cabalmente*. PhD diss., University of Graz.

Gruber, T. R. (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In International Journal Human-Computer Studies 43, pp 907-928.

Hellmann, S. et al. (2013). Integrating NLP using Linked Data. In Proceeding ISWC '13 Proceedings of the 12th International Semantic Web Conference - Part II. New York: Springer, pp. 98-113. doi:10.1007/978-3-642-41338-4_7

Hummel, M. (2014). Los adjetivos adverbiales. In C. Company Company (Ed.), Sintaxis histórica de la lengua española, Part III. México: Universidad Nacional Autónoma de México/Fondo de Cultura Económica, pp. 613–731.

Ledgeway, A. (2017). Parameters in Romance adverb agreement. In Hummel M./Valera S. (eds.): Adjective Adverb Interfaces in Romance, pp. 47-80.

Pollin, C. & Vogeler, G. (2017). Semantically Enriched Historical Data. Drawing on the Example of the Digital Edition of the "Urfehdebucher der Stadt Basel". In Proceedings of the Second Workshop on Humanities in the Semantic Web co-located with 16th International Semantic Web Conference. Vienna. CEUR Workshop Proceedings pp. 27-32.

Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the Humanities. In Journal of Digital Humanities, 2. Jg., Nr. 3, pp. 2-13.

Stigler, J. H. & Steiner, E. (2018). Gams – An Infrastructure for the Long-term Preservation and Publication of Research Data from the Humanities. Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, 71(1), pp. 207-216.

Yi, M. (2008). Topic Maps-based Ontology and Semantic Web. Saarbrücken: Dr. Müller.

---

[19] http://glossa.uni-graz.at/archive/objects/query:aaif.getsh3/methods/sdef:Query/get, prototype, 04.07.2018