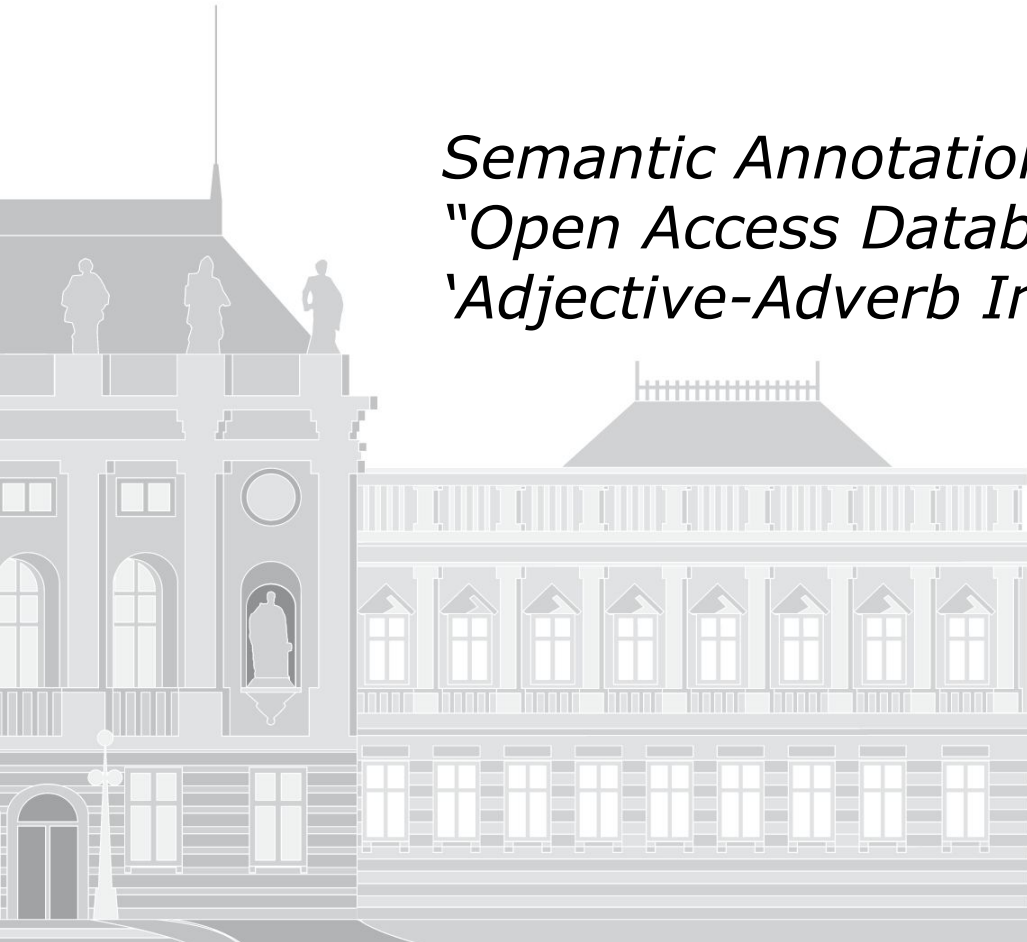


*Semantic Annotation in the Project
"Open Access Database
'Adjective-Adverb Interfaces' in Romance"*

Christopher Pollin,
Gerlinde Schneider,
Katharina Gerhalter,
Martin Hummel



Open Access Database

"Adjective-Adverb Interfaces in Romance"

- *Open Research Data Pilot*, Austrian Science Fund
- September 2017 to December 2019
- *PI*: Martin Hummel (Institute for Romance Studies)
- *Data acquisition*: Katharina Gerhalter (Institute for Romance Studies)
- *Data modelling*: Gerlinde Schneider and Christopher Pollin (Centre for Information Modelling)

<https://adjective-adverb.uni-graz.at/de/forschen/projekte/open-access-database-2017-2019/>

Research group

Investigates relations between
the word classes of **adjective** and **adverb** in Romance languages

- Research data as an output from several projects and publications
- Complex linguistic annotations
- Annotation model is developed further for new requirements
- Degree and emphasis of the annotation varies
- Multilingual data

Objectives

- Possibilities and challenges of open linguistic research data
- Comprehensive database for the diverse data of the research group
- Querying across corpora and languages
- Open access to linguistically annotated data in a reasonable way
 - via standardized formats and interfaces
- Long-term availability and preservation of the data

In this talk:

Using semantic technologies to reach these aims

Adjective-Adverb Interfaces

~ Adjectives with adverbial function

Adjective-Adverbs

*ver **claro***

Inflected Adverbs

***altos** subieran los fumos*

Discourse markers

cierto

Adverbial prepositional phrases

*de **seguro***

Mostly in substandard language and regional varieties

Annotation of AA-Interfaces

Syntactic information (eg. relative word order)

Morphosyntactic information (eg. word class)

Semantic information (eg. semantic target)

→ Adjective-Adverbs

+ entities that relate to the AA

Verb; Subject of the AA construction; Preposition + Article/ + Possessive

Annotated Corpora

- French: *Dictionnaire Historique de l'Adjectif-Adverbe* (dicoadverbe)
 - > 13.000 examples, 11th - 20th century
- Spanish: Reading corpus for *Sintaxis Histórica de la Lengua Española* (2014, Company Company) - Martin Hummel “Los adjetivos adverbiales”
 - > 1.200 examples, 13th - 21st century
- Spanish: Corpus on diachrony of Spanish
 - > 2.200 examples, 13th - 21st century

- (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)
- (2) [...] tan [a::alto::altos::apvmln] [v::subir::subieran::i] [s::los fumos::mp]
e los sacrificios

Categories for Adverb Annotation

place	main category	selectable options	code	examples
1.	Morphosyntactic Structure	adjective	a	volar [a::alto::alto::auvnl̩n]
		derived <i>-mente</i>	m	[a::exactamente::exacto::munnsn] cuatro metros
		derived <i>-ește</i>	e	iară [a::românește::român::euvnm̩n]
		derived <i>-ly</i>	l	to run [a::quickly::quick::luvnm̩n]
		derived <i>-iș</i>	i	furiș
		derived <i>-ul</i>	u	dreptul
2.	Inflection <i>(inaudible-audible for French; specific categories for other languages)</i>	uninflected /m.sg.	u	volar [a::alto::alto::auvnl̩n]
		audible inflection	a	vivre [a::saine::sain::aasn̩m̩n]
		inaudible inflection	i	je m'en vais [a::seule::seul::aisnm̩n]
		Feminine Singular	f	con [a::sola::solo::afnnsn] una palabra de tu boca
		Neuter Singular	n	just in case an [a::example::example::anvnm̩n]
		Masculine Plural	p	los [a::firmes::firme::apn̩nm̩n] enamorados
		Feminine Plural	x	las aves vuelan [a::altas::alto::axvnl̩n]
		Neuter Plural	z	just in case an [a::example::example::azvnm̩n]

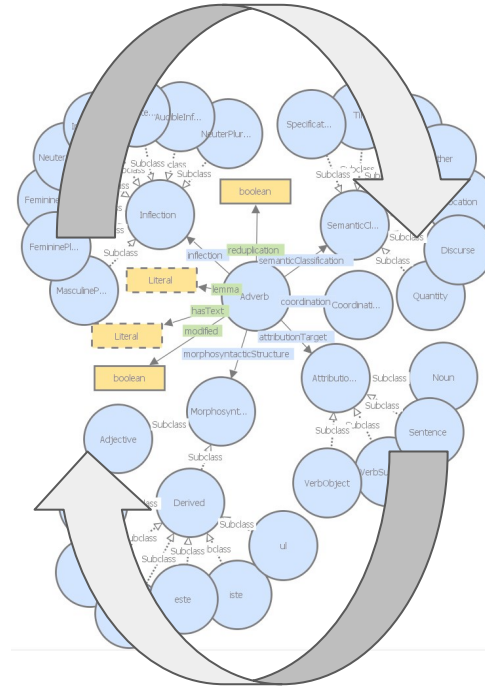
Related work

- Linguistic Linked Open Data Cloud (LLOD)
 - Ontologies of Linguistic Annotations (OLiA)
[Chiarcos et al., 2016]
 - NLP Interchange Format (NIF)
[Hellmann et al., 2013]
- Standardized URI schemas, REST interfaces,
RDF, RDF/OWL-based ontologies

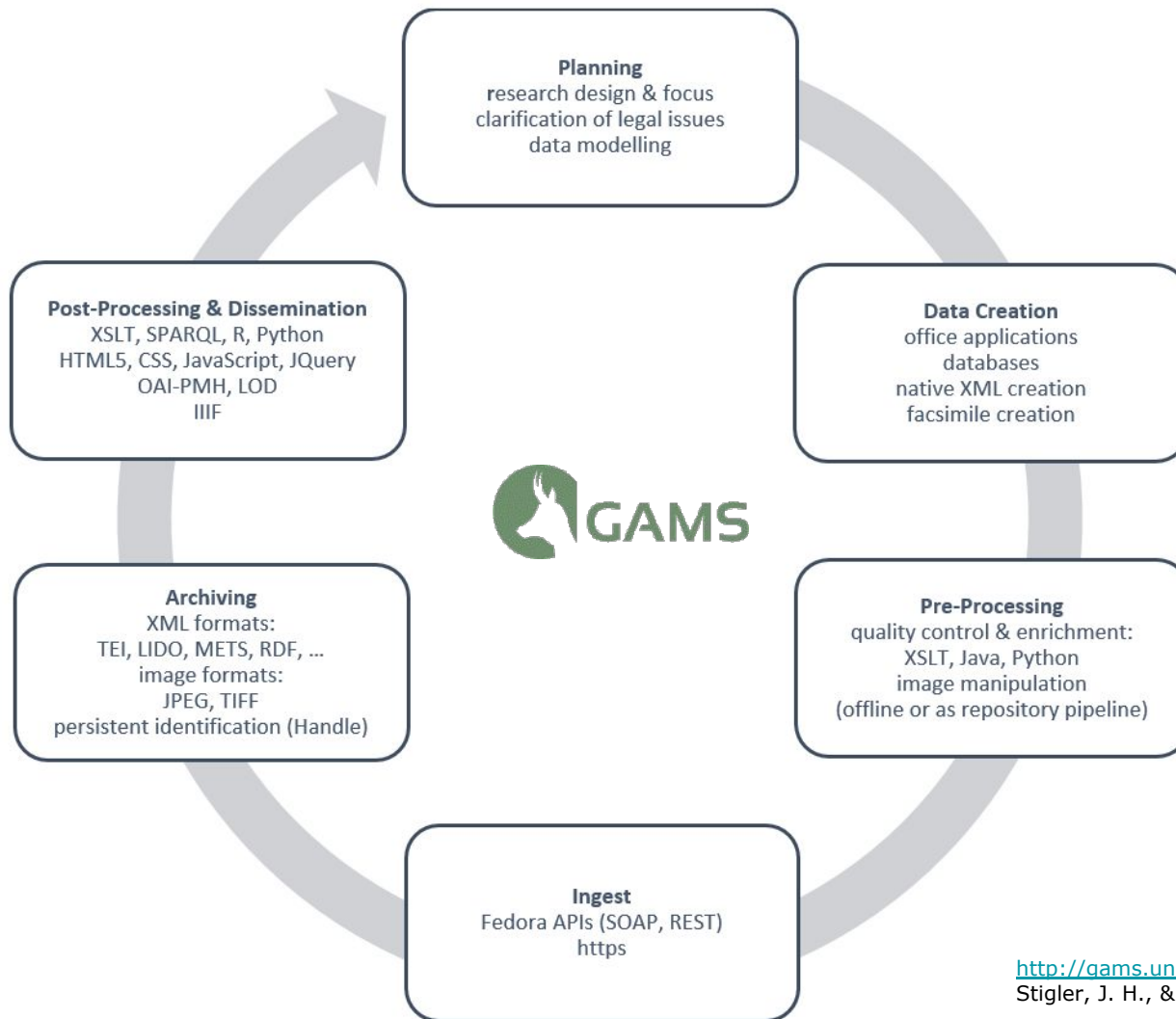
AAIF-Ontology

"a formal, explicit specification of a shared conceptualization"

[Brost, 1995]







WORD to (not the best) TEI



tan

[a::alto::altos::apvmln]

[v::subir::subieran::i]

[s::los fumos::mp]

e los sacrificios

- | | | |
|----|----------------------------|------------------|
| 1. | Morphosyntactic structure: | adjective |
| 2. | Inflection: | masculine plural |
| 3. | Attribution target | verb |
| 4. | Modified | yes |
| 5. | Semantic Classification | location |
| 6. | Reduplication | no |

<s>e dize mestre Pedro que este pujamiento
dell aguaque fuera tanto en alto porque

<phr type="syntagm">tan

<w type="adverb" lemma="alto"
function="apvmln">altos</w>

<w lemma="subir" function="i"
type="verb">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 suziedad d'aquellos fumos ell
aire.

</s>

RDF

```
<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>
```

```
<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>
```

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

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

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

SPARQL

```
SELECT ?Adverb_text ?Adverb_lemma ?Verb_text ?Verb_lemma ?Entry_text  
{
```

```
  #get SH3 corpus, text and XML
```

```
  ?Entry      gams:isMemberOfCollection <https://gams.uni-graz.at/o:aaif.sh3>;  
              aaif:phrase ?Phrase;  
              gams:textualContent ?Entry_text;  
              gams:XMLContent ?XMLContent.
```

```
  #get Adverb
```

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

```
  #get Verb
```

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

```
  #further criterias for the adverb
```

```
  ?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>.
```

```
}
```


Conclusion

- Long-term preservation: self-describing data and model
- Domain-specific ontology: flexible - interoperable - transparent
- Linked Open Data
- Word → TEI → RDF
- Search interface

Challenges

- Overlapping structures and different levels of annotation
- Keeping sequence of text

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. (2017). Adjectives with adverbial functions in Romance. In M. Hummel & S. Valera (eds), *Adjective adverb interfaces in Romance*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 13-46.
- 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), 207-216.
- Yi, M. (2008). *Topic Maps-based Ontology and Semantic Web*. Saarbrücken: Dr. Müller.