

**5th Workshop on Linked Data in Linguistics:
Managing, Building and Using Linked Language Resources.
Portorož, Slovenia, 24th May 2016. Co-located with LREC 2016**

Representation of Polarity Information of Elements of German Compound Words

Thierry Declerck

Universität des Saarlandes Fachrichtung 4.7
Allgemeine Linguistik

&

DFKI GmbH, Language Technology Lab



Starting point: SentiMerge

SentiMerge, a resource that encodes polarity information for German words on the basis of integration processes performed on four pre-existing polarity lexicons for German (Clematide and Klenner, 2010; Remus et al. 2010; Waltinger, 2010 and Klenner et al., 2012).

Entry	POS	Polarity Value	Confidence	Translation
arbeitslos	AJ	-0.968	14.527	jobless
Freihalten	V	0.777	7.966	to keep free
Goldhochzeit	N	0.628	5.823	golden wedding anniversary
Rotsperre	N	-0.628	5.823	red card suspension
Abdeckblech	N	0.0	7.966	cover plate

Examples of SentiMerge entries, all compound words

SentiMerge is described in: Emerson, G and Declerck, T. (2014). SentiMerge: Combining Sentiment Lexicons in a Bayesian Framework. In *Proceedings of the 2014 Workshop on Lexical and Grammatical Resources for Language Processing*. Dublin, Ireland

- To represent the SentiMerge data in a Linguistic Linked Open Data compliant format
- Use of modules of the *lexicon model for ontologies* (lemon), resulting from the W3C Ontology-Lexica Community Group, for the encoding of lexical information:
 - Ontology-lexicon interface (*ontolex*) – core model
 - Decomposition module (*decomp*)
- Use of the MARL model for the encoding of polarity information
- Integration of *ontolex-decomp* and MARL

Ontology-lexicon interface (*ontolex*)

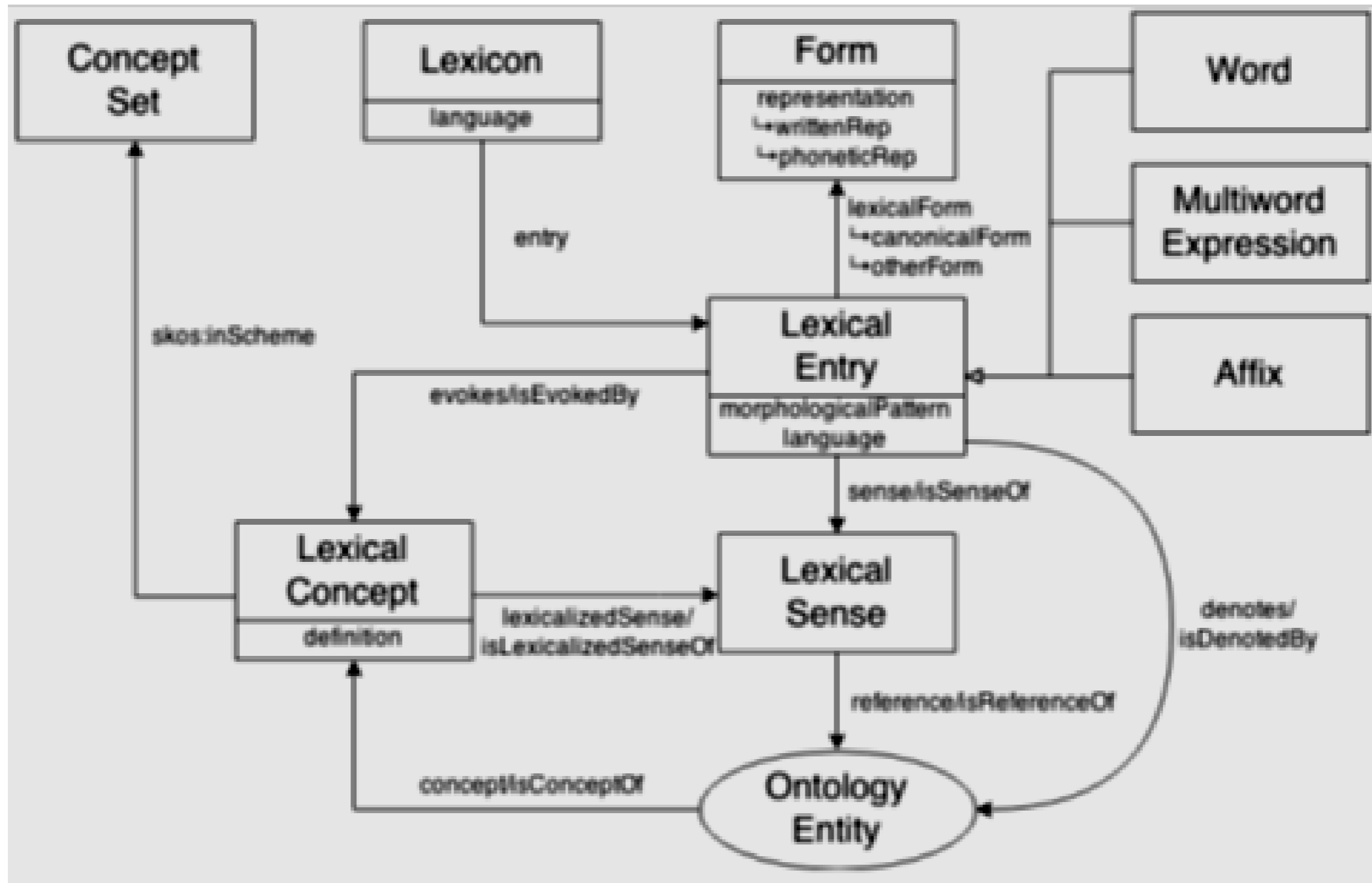


Figure created by John P. McCrae for the W3C Ontology-Lexica Community Group

Decomposition module (*decomp*)

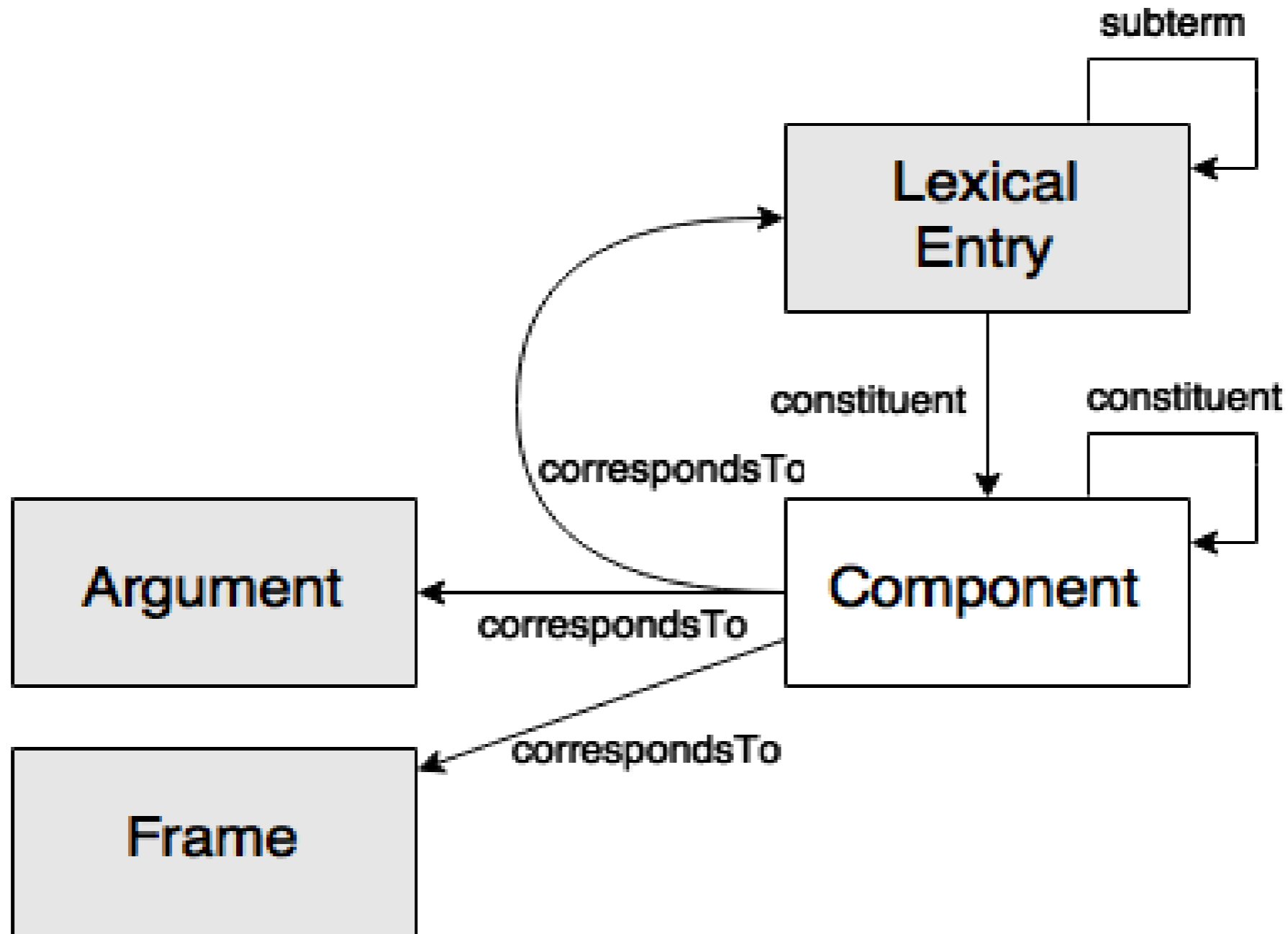


Figure created by John P. McCrae for the W3C Ontology-Lexica Community Group

Example of a compound in *ontolex* and *decomp*



Representation of the compound

:Rotsperre_lex

rdf:type ontolex:MultiwordExpression ;

lexinfo:partOfSpeech lexinfo:noun ;

rdf:_1 :Rot_comp ;

rdf:_2 :sperre_comp ;

decomp:constituent :Rot_comp ;

decomp:constituent :sperre_comp ;

decomp:subterm :Sperre_lex ;

decomp:subterm :rot_lex ;

ontolex:denotes

<<https://www.wikidata.org/wiki/Q1827>

> .

Representation of the components

:Rot_comp

rdf:type decomp:Component

decomp:correspondsTo :rot_lex .

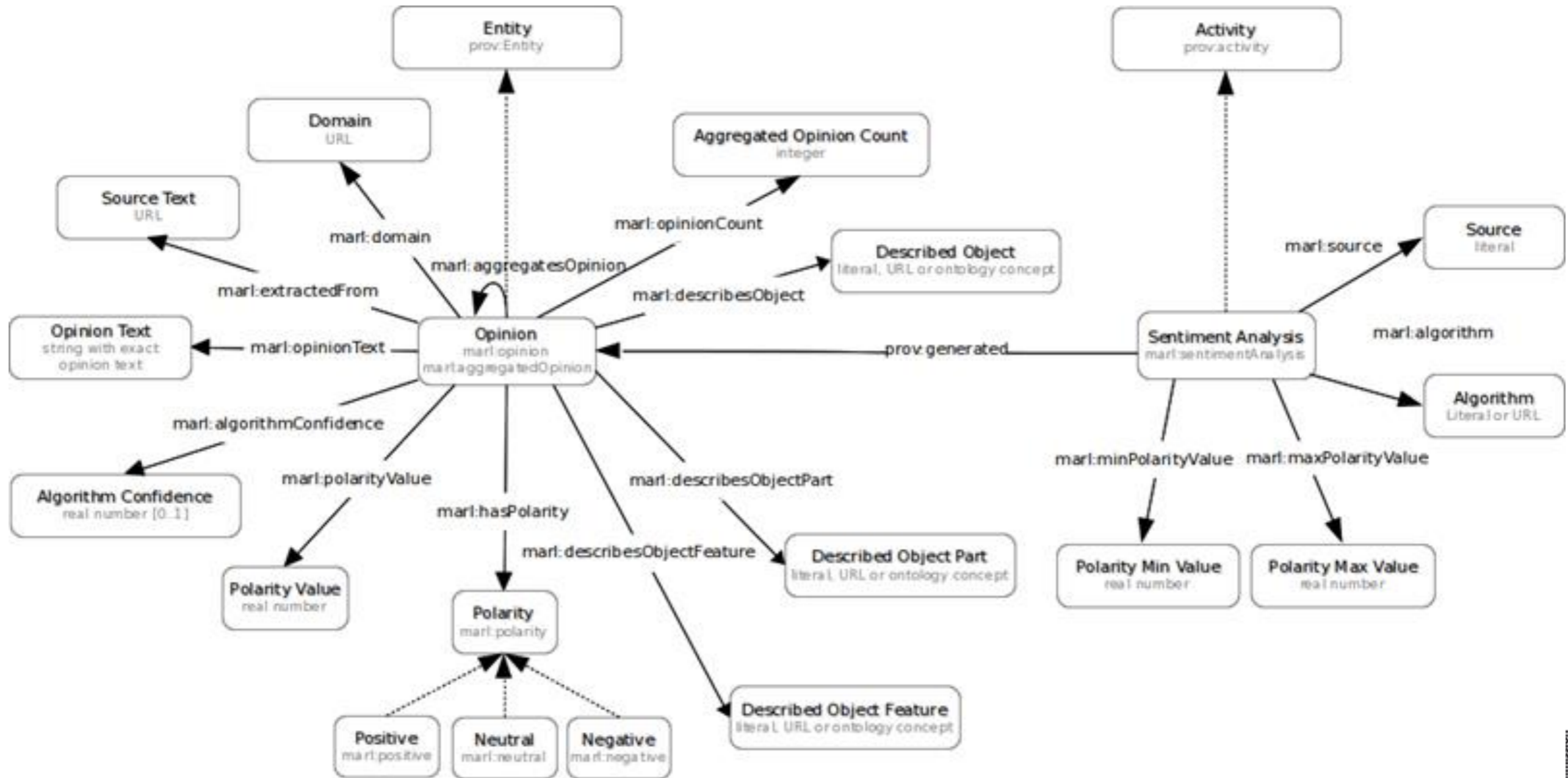
:sperre_comp

rdf:type decomp:Component ;

decomp:correspondsTo :Sperre_lex .



The MARL Model



<http://www.gsi.dit.upm.es/ontologies/marl>

Integrating the MARL information as part of an ontolex:sense, as in (Buitelaar et al., 2013)



:rotsperre_sense

rdf:type ontolex:LexicalSense ;

op:assessedBy :SentiMerge ;

op:hasPolarity op:Negative ;

op:maxPolarityValue "1.0"^^xsd:double ;

op:minPolarityValue "-1.0"^^xsd:double ;

op:polarityValue "-0.628"^^xsd:double ;

rdfs:label "Sense for the German word \"Rotsperre\" @en ;

ontolex:isSenseOf :Rotsperre_lex ;

ontolex:reference

<<http://de.dbpedia.org/resource/Wettkampfsperre>> .



Senses of the Component „Sperre“



A first sense for „Sperre“

:sperre_sense1

rdf:type ontolex:LexicalSense ;

op:assessedBy :TD ;

op:hasPolarity op:Neutral ;

op:maxPolarityValue "1.0"^^xsd:double ;

op:minPolarityValue "-1.0"^^xsd:double ;

op:polarityValue "0.0"^^xsd:double ;

rdfs:label "A sense for the German word
"Sperre\""@en ;

ontolex:isSenseOf :Sperre_lex ;

ontolex:reference
<<http://de.dbpedia.org/resource/Lock>> .

A second sense for „Sperre“ (same as for the compound)

:sperre_sense2

rdf:type ontolex:LexicalSense ;

op:assessedBy :SentiMerge ;

op:hasPolarity op:Negative ;

op:maxPolarityValue "1.0"^^xsd:double ;

op:minPolarityValue "-1.0"^^xsd:double ;

op:polarityValue "1.0"^^xsd:double ;

rdfs:label "A sense for the German word
"Sperre\""@en ;

ontolex:isSenseOf :Sperre_lex ;

ontolex:reference
<<http://de.dbpedia.org/resource/Wettkampfsperre>> .



Our Proposal: Associating „senses“ to components of a compound



```
:sperre1_comp a          decomp:Component ;
```

```
    decomp:correspondsTo :Sperre_lex ;
```

```
    ontolex:sense        :sperre_sense1 .
```

```
:sperre2_comp a          decomp:Component ;
```

```
    decomp:correspondsTo :Sperre_lex ;
```

```
    ontolex:sense        :sperre_sense2 .
```

- Still need to add information about the position of the component in the compound in order to more appropriately link a component to a sense
- Still the need to add information about the `decomp:Frame` and `decomp:Arguments` in order to more appropriately link a component to a sense



- We presented on-going work on the LLOD compliant encoding of the SentiMerge lexicon, with a focus on compound words and on the possibility to associate senses and polarity to components of compounds.
- For this we presented an integration of the ontalex and decomp modules of lemon with the MARL model.
- Current work on making use of the position of components in a compound and of arguments information related to components.

Thanks for your attention!



- Questions?
- Positive (**0.968**, 14.527) comments?



- **PHEME -- Computing Veracity: the Fourth Challenge of Big Data -- is supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 611233**
 - <http://www.pHEME.eu/>
- **FREME – Open Framework Of E-Services for Multilingual and Semantic Enrichment of Digital Content – is supported by the Horizon 2020 Framework Programme of the European Union under Grant Agreement Number 644771**
 - <http://www.freme-project.eu/>

References



Buitelaar, P., Arcan, M., Iglesias, C.A., Sánchez, J.F. and Strapparava, C. (2013). Linguistic Linked Data for Sentiment Analysis. In *Proceedings of the 2nd Workshop on Linked Data in Linguistics (LDL 2013): Representing and linking lexicons, terminologies and other language data*. Collocated with the Conference on Generative Approaches to the Lexicon, Pisa, Italy.

Clematide, S, Klenner, M. (2010). Evaluation and extension of a polarity lexicon for German. In *Proceedings of the Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*. Held in conjunction to ECAI 2010, Lisbon, Portugal.

Clematide, S., Gindl, S., Klenner, M., Petrakis, S., Remus, R., Ruppenhofer, J., Waltinger, U. and Wiegand, M. (2012). MLSA - A Multi-layered Reference Corpus for German Sentiment Analysis. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey.

Declerck, T. and Lendvai, P. (2015). Towards the Representation of Hashtags in Linguistic Linked Open Data Format. In *Proceedings of the Second Workshop on Natural Language Processing and Linked Open Data*. Hissar, Bulgaria.

Emerson, G and Declerck, T. (2014). SentiMerge: Combining Sentiment Lexicons in a Bayesian Framework. In *Proceedings of the 2014 Workshop on Lexical and Grammatical Resources for Language Processing*. Dublin, Ireland.

Francopoulo, G., George, M., Calzolari, N., Monachini, M., Bel, N., Pet, M. and Soria, C. (2006). Lexical Markup Framework (LMF). In *Proceedings of the fifth international conference on Language Resources and Evaluation*.

Klenner, M., Clematide, S., Petrakis, S. and Luder, M. (2012). "Compositional syntax-based phrase-level polarity annotation for German". In *Proceedings of the 10th International Workshop on Treebanks and Linguistic Theories (TLT 2012)*, Heidelberg, Germany.

Krieger, H.-U. and Declerck, T. (2014). TMO - The Federated Ontology of the TrendMiner Project. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2014)*

McCrae, J.-P., Aguado-de-Cea, G., Buitelaar, P., Cimiano, P., Declerck, P., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D. and Wunner, T. (2012). Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation*, 46(4), pp. 701-719.

Remus, R., Quasthoff, U. and Heyer, G. (2010). SentiWS - a Publicly Available German-language Resource for Sentiment Analysis. In *Proceedings of the 7th International Language Resources and Evaluation (LREC'10)*.

Waltinger, U. (2010). Sentiment Analysis Reloaded: A Comparative Study On Sentiment Polarity Identification Combining Machine Learning And Subjectivity Features". In *Proceedings of the 6th International Conference on Web Information Systems and Technologies (WEBIST '10)*.

Westerski, A. and Sánchez-Rada, J.F. (2013). Marl Ontology Specification, V1.0 May 2013. Available at <http://www.gsi.dit.upm.es/ontologies/marl>

