Connecting European Cultural Diversity: a Case Study on Colour Terms

Challenges in a digital lexicographic context

Amelie Dorn, Eveline Wandl-Vogt & Thierry Declerck
Austrian Centre for Digital Humanities (ACDH)
Austrian Academy of Sciences (ÖAW)

Outline

- Introduction
- Colours and Culture across Europe
- Language Diversity and Colour Terms
- Challenges for a Digital Lexicographic Output
- Heterogeneous Data in a Digital Humanities Context
- Conclusion and summary

Introduction

Colours and colour terminology

→ widely studied across different disciplines:

linguistics, psychology, physics, anthropology, natural sciences

→ an important part of many fields of life

→ highly inter/trans-disciplinary research area

Introduction

Colours and colour terminology

→ relativist vs universalist approach (cf. Berlin & Kay, 1969, Wierzbicka, 2006)

→ areas of interest in the literature: colour semantics and perception (cf. Berlin & Kay, 1969; Deutscher, 2010; Gage, 1993; Wierzbicka, 2006)

→ a language‘s colour system may depend on general perception and people‘s experiences

Colours and Culture across Europe

Colour terms and cultural associations (emotions, superstitions, traditions)

→ manifested in phrases, idioms, collocations

→ naming orientated on the immediate nature, environment and everyday life

→ possibly subject to change over time

Language diversity and colour terms

Colours in a European cross-cultural context

Colour terms within the COST ENeL framework

→ conference on colour naming (Lisbon 2015)

→ *European roots* (collecting words concerning *colours* and *emotions*)

→ Study on idioms concerning *colours and emotions – „red“ and „anger“*

(Blanck, 2015 at eLex ; Dorn, Villalva, Giouli, Blanck, Kovalenko & Wandl-Vogt, submitted)

Language diversity and colour terms

Colours in a European cross-cultural context

→ idioms concerning *colours and emotions:* similarities + differences

e.g. English *red with anger*
German *rot vor Wut*
French *rouge de colère*
Portuguese *verde de raiva*

Language diversity and colour terms

Colours in a European cross-cultural context

diversity = challenge for electronic lexicographic output

- how should linguistic/cultural diversity be addressed?
- translation – which language?
- dealing with heterogeneous data?
- how to display and model the concept behind the idioms across languages?

Challenges for a lexicographic output

Colours in a European cross-cultural context

Modelling the concepts behind the idioms across languages?

English	*red with anger*
	[red referring to colour of the face]
German	*rot vor Wut*
	[red referring to colour of the face]
Portuguese	*verde de raiva*
	[green referring to colour of …]

Heterogeneous data in a DH context

→ DH: linking, disciplines, sources and resources of various fields

→ offering tools for dealing with heterogeneous data

→ linking lexicographic information/output with other areas

example: *Red wine* → *vino tinto, vi negre, vino rosso*

Heterogeneous data in a DH context: Colour of wine, colour of grape

→ **Red wine** is a type of wine made from dark-coloured (black) grape varieties {@en}

→ **Rotwein** ist ein aus blauen Weintrauben hergestellter Wein {@de}

→ Un **vin rouge** est obtenu par la fermentation du moût de raisins noirs {@fr}

→ El **vino tinto** es un tipo de vino procedente mayormente de mostos de uvas tintas {@es}

→ El **vi negre** és aquell procedent de mostos de raïm negre {@ca}

→ Il **vino rosso** è prodotto unicamente con uve a bacca rossa {@it}

Heterogeneous data in a DH context: Use of Knowledge Sources -- Wikidata

→ A URI for the *concept* “Red Wine”: https://www.wikidata.org/wiki/Q1827

→ Labels in many languages linked to this source (represented in JSON):

→ {"item":{"type":"uri","value":"http://www.wikidata.org/entity/**Q1827**"},"itemLabel":{"xml:lang":"ca","type":"literal","value":"Vi negre"}}

→ {"item":{"type":"uri","value":"http://www.wikidata.org/entity/**Q1827**"},"itemLabel":{"xml:lang":"de","type":"literal","value":"Rotwein"}}

→ {"item":{"type":"uri","value":"http://www.wikidata.org/entity/**Q1827**"},"itemLabel":{"xml:lang":"en","type":"literal","value":"red wine"}}

→ …

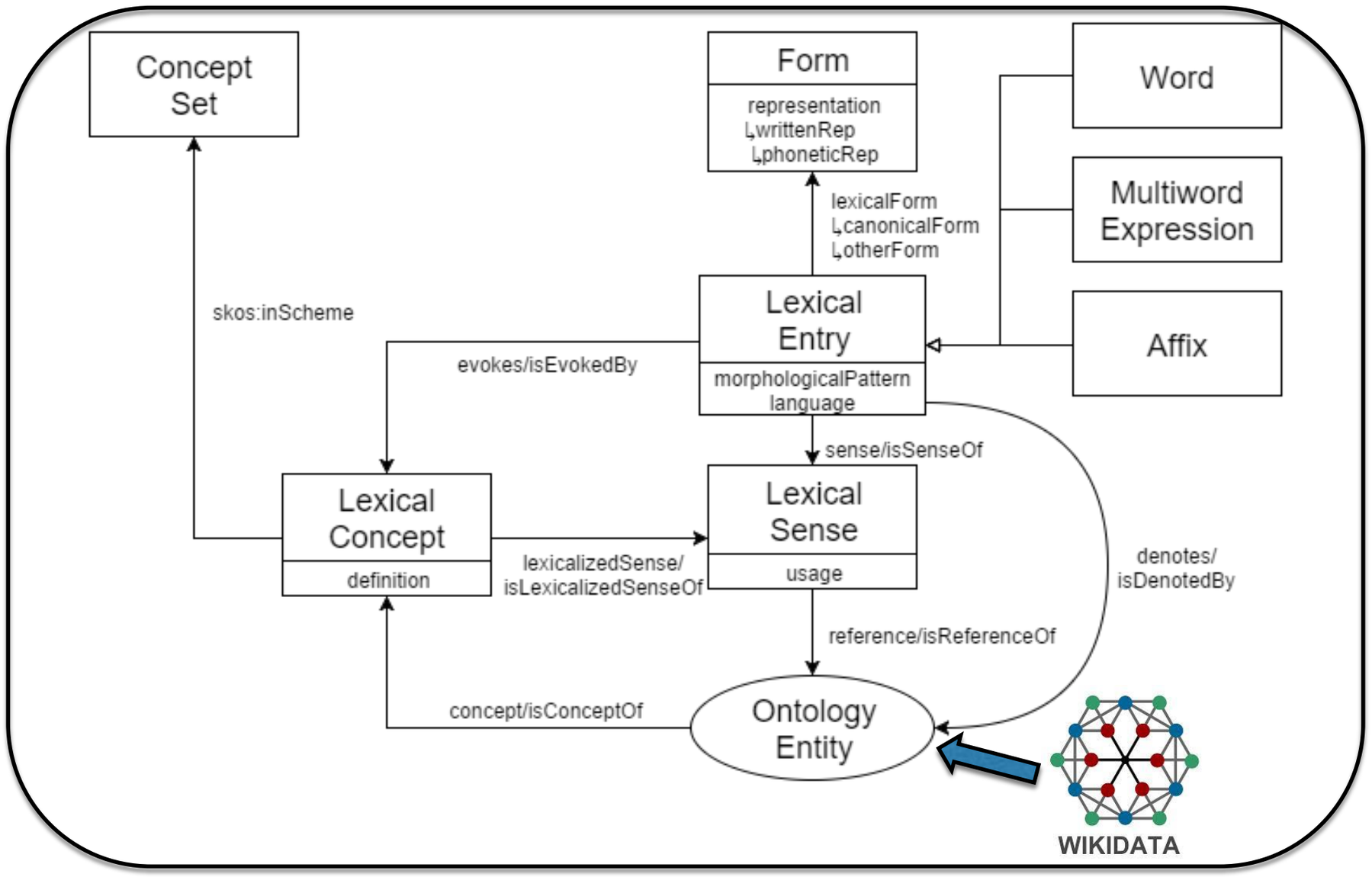
→ But no further lexical information given.

Heterogeneous data in a DH context: Combining Wikidata and Ontolex

→ Ontolex is a specification of the *lexicon model for ontologies* (lemon) as resulting from the work of the W3C Ontology Lexicon Community Group: https://www.w3.org/community/ontolex/wiki/Final_Model_Specification.

Heterogeneous data in a DH context: Combining Wikidata and Ontolex

Ontolex



Graph: courtesy of John McCrae

Encoding lexical data in Ontolex

→ **Rotwein** ist ein aus blauen Weintrauben hergestellter Wein {@de}

→ „Rotwein“ is a compound word, a Noun, consisting of an Adjective and a Noun. The Adjective is in first position and the Noun in the second one. The Adjective refers to a colour, and the Noun to an alcoholic beverage.

Encoding lexical data in Ontolex (simplified)

:Rotwein_lex

rdf:type ontolex:LexicalEntry ;

lexinfo:partOfSpeech lexinfo:noun ;

rdf:_1 :Rot_comp ;

rdf:_2 :wein_comp ;

decomp:constituent :Rot_comp ;

decomp:constituent :wein_comp ;

decomp:subterm :Wein_lex ;

decomp:subterm :rot_lex ;

ontolex:denotes <http://www.oeaw.ac.at/acdh/compound#https://www.wikidata.org/wiki/Q1827> .

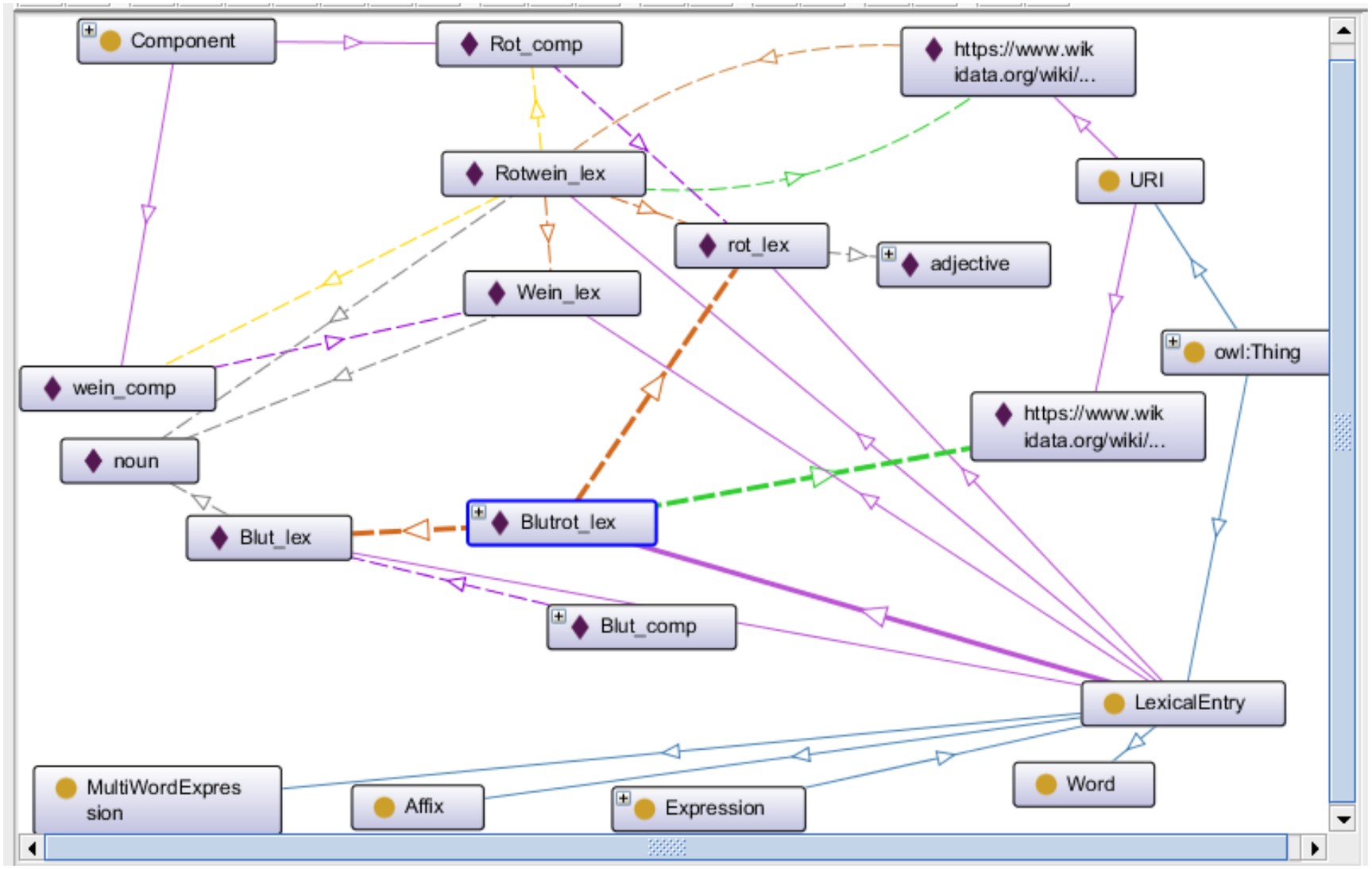
```
:Wein_lex
   rdf:type ontolex:LexicalEntry ;
   lexinfo:partOfSpeech lexinfo:noun .
```

```
:rot_lex
  rdf:type ontolex:LexicalEntry ;
  lexinfo:partOfSpeech lexinfo:adjective .
```

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

austrian centre for digital humanities

Encoding language specific data in Ontolex



Conclusions

Colours and colour terminology

- Linguistically and culturally diverse
- Challenge in dealing with heterogeneous data → DH tools offering possibilites of linking and connecting different sources
- Challenge for a lexicographic output → connecting concepts; displaying structure of languages

Conclusions

Colours and colour terminology

- Linguistically and culturally diverse
- Challenge in dealing with heterogeneous data → DH tools offering possibilites of linking and connecting different sources
- Challenge for a lexicographic output → connecting concepts; displaying structure of languages

Conclusions

Colours and colour terminology

- Linguistically and culturally diverse
- Challenge in dealing with highly variable data → DH tools offering possibilites of linking and connecting different sources
- Challenge for a lexicographic modelling → including cultural concepts; comparing languages structurally

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

austrian centre for digital humanities

Thank you!

References

Berlin, B. & Kay, P. (1969) *Basic colors terms: their universality and evolution*. University of University of California Press.

Deutscher, G. (2010) G. (2010) G. (2010) *Through the Language Glass. How Words Colour Your World.* London: William Heinemann.

Gage, J. (1993) *Colour and Culture: Practice Meaning from Antiquity to Abstraction.* London: Thames and Hudson.

Wierzbicka, A. 2006. The semantics of colour: A new paradigm, in CP Biggam, CJ Kay (eds), Progress in Colour Studies. John Benjamins Publishing Company, Amsterdam, The Netherlands, pp. 1-24.

Presentation by Blanck, W. 2015 at eLex: Congress of Electronic lexicography in the 21st century: Linking lexical data in the digital age. Herstmonceux, UK, 11-13 August 2015. https://elex.link/elex2015/

Dorn, Villalva, Giouli, Blanck, Kovalenko & Wandl-Vogt, (2016) *Displaying Language Diversity on the European Dictionary Portal - COST ENeL - Case Study on Colours and Emotions and their Cultural References.* Proceedings of Euralex 2016, Tbilisi, Georgia.

Ontolex graph slide 14: courtesy of John P. McCrae
Graph slide 17: Thierry Declerck