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


→ 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

- 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

element: *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

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)

\[\text{:Rotwein\_lex}\]

\[
\text{rdf:type ontolex:LexicalEntry ;}
\text{lexinfo:partOfSpeech lexinfo:noun ;}
\text{rdf:_1 :Rot\_comp ;}
\text{rdf:_2 :wein\_comp ;}
\text{decomp:constituent :Rot\_comp ;}
\text{decomp:constituent :wein\_comp ;}
\text{decomp:subterm :Wein\_lex ;}
\text{decomp:subterm :rot\_lex ;}
\text{ontolex:denotes <http://www.oeaw.ac.at/acdh/compound#https://www.wikidata.org/wiki/Q1827> .}
\]

\[\text{:Wein\_lex}\]

\[
\text{rdf:type ontolex:LexicalEntry ;}
\text{lexinfo:partOfSpeech lexinfo:noun .}
\]

\[\text{:rot\_lex}\]

\[
\text{rdf:type ontolex:LexicalEntry ;}
\text{lexinfo:partOfSpeech lexinfo:adjective .}
\]
Encoding language specific data in Ontolex
Conclusions

Colours and colour terminology

• Linguistically and culturally diverse
  • Challenge in dealing with heterogeneous data → DH tools offering possibilities 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 possibilities 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 possibilities of linking and connecting different sources
- Challenge for a lexicographic modelling → including cultural concepts; comparing languages structurally
Thank you!
References


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