Displaying Language Diversity on the European Dictionary Portal - COST ENeL - Case Study on Colours and Emotions and their Cultural References

Amelie Dorn¹, Alina Villalva², Voula Giouli³, Wiebke Blanck⁴, Kira Kovalenko⁵, Eveline Wandl-Vogt¹

Austrian Academy of Sciences, Austrian Centre for Digital Humanities¹; Universidade de Lisboa²; Institute for Language and Speech Processing, Athen RIC³; Göttingen Academy of Sciences, Deutsches Wörterbuch von J.Grimm und W. Grimm⁴; Russian Academy of Sciences, Institute for Linguistic Studies⁵
e-mail: Amelie.Dorn@oeaw.ac.at; AlinaVillalva@letras.ulisboa.pt; voula@ilsp.gr; wiebke.blanck@gmx.net; kira.kovalenko@gmail.com; Eveline.Wandl-Vogt@oeaw.ac.at

Abstract

In this paper we present a case study on colour and emotion terms and their cultural references in the framework of the COST European Network of e-Lexicography (ENeL), working towards Pan-European lexicography. We take an initial use case of red in connection with emotions (anger) and look at its roots across different European languages, including Russian. Our data model offers the possibility of connecting these fields in the context of digital lexicography using markup for etymological information with description standards like ONTOLEX or TEI. This is particularly relevant for using and displaying such data on the European Dictionary Portal, potentially offering access to detailed diachronic and synchronic lexicographic knowledge across a variety of languages.

Keywords: linguistic diversity; digital lexicography; European roots

1 Introduction: COST ENeL, WG4 and the Task Group “European Roots: Colour and Emotion Terms”

In keeping up with today’s digital age, the field of lexicography has seen the need of including new and innovative methods for linking a variety of different lexical sources, from mono- and multi-lingual to standard and non-standard works. Within the Cost Action of the European Network of e-Lexicography (ENeL)¹, it has been the main goal (see working group (WG) 4)² to work towards Pan-European lexicography, aiming to display single dictionary sources on the European Dictionary Portal³. This paper takes the use case of red in the field of colour and emotion terms, which are both areas investigated in the task group European roots⁴ within the COST ENeL framework⁵. Colours,

¹ website: www.elexicography.eu [last accessed 06/11/2015]
² MOU: http://www.cost.eu/COST_Actions/isch/IS1305 [last accessed 30/10/2015]
⁴ COST ENeL WG4: European roots https://sites.google.com/a/campus.ul.pt/european-roots/home [last accessed 21/04/2016]
⁵ see also Colour and colour naming: crosslinguistic approaches. International conference, University of
and red in particular, and their cultural associations have been well documented (cf. Deutscher, 2010; Berlin & Kay, 1969; Gage, 1993; Villalva et al, 2014), as has the connection between these two fields. During a COST ENEl WG 4 meeting, some related comments were raised by members of the European roots group which can be summarised as follows:

a. “Red” (and any other colour) involves different (and sometimes similar) idioms, phrases and ideas in many European languages.”

b. “The words that could be explored in this context should be words of emotions that also show connections to colours (i.e., firstly, “red””).

Exploring the connection between colour words and words of emotions was one of the starting points in the European roots group as members of this group had been working on that question previously. The question was raised how to represent such a structure in the context of digital lexicography. In this paper we take an initial use case of red in connection with emotions (anger) and look at its roots across different European languages with the potential of connecting these fields in the context of digital lexicography.

2 The Data

The collection of data on specific colour and emotion words has been implemented within the COST ENEl European roots task group, with the aim of designing and testing an electronic dictionary prototype for the root of simple words and their derivatives. The model entries for the Portuguese word roots rai- (raiva/anger) and vermelh- (vermelho/red) have been collected and listed according to a set of criteria. Being studied from an etymological point of view, cognates will be identified in different languages and the spreading of each etymon in European languages will be traced. To-date equivalents of red and anger have been gathered across a number of European languages (Portuguese, Galician, Spanish, Catalan, Italian, French, Romanian, Greek, German and Polish) and it is envisaged to extend this corpus towards further languages. In the study we envisage to also include non-standard linguistic data (from the Dictionary/Database of Bavarian dialects in Austria (WBO/DBÖ), and here we also include examples of Russian.


11 Datenbank der Bairischen Mundarten in Österreich @ electronically mapped. Accessed at: http://wboe.oeaw.ac.at/ [20/04/2016]
3 Data Model for European Colour and Emotion Terms

The development of recent digital means, together with new and innovative methods of connecting data, has enabled the integration of various sources, making linked linguistic resources available online. The data model applied in this study aims to show similarities and/or differences across other European languages, offering detailed lexicographic knowledge right from the word root with key etymological information. Although the fields of colour and emotion words collected in the ROOTS project are not connected, establishing a link between these two fields seems valid as many idioms, phrases and metaphors on emotions are also related to colour terms.

3.1 Background

A number of lexical resources propose models for representing emotion concepts. WordNet–Affect (Strapparava, 2004), an extension of WordNet Domains, comprises a subset of synsets that are suitable for representing affective concepts correlated with affective words. The affective concepts representing emotional states are individuated by synsets marked with the a–label emotion. A–labels for concepts representing moods, situations eliciting emotions, or emotional responses are also provided for.

From another perspective, OntoEmotion, (Francisco, Gervás & Peinado, 2007) uses semantic web technologies to represent a taxonomic structure of emotional categories, on the basis of psychological models. The ontology takes into account basic emotions and specific emotional categories, hierarchically organized. It comprises two classes: Emotion (i.e. the superclass of all emotional concepts), and Word (i.e. language-specific lexicalizations of emotions, that is, emotion-denoting words in English and Spanish).

Our proposal is based on the approach taken for representing the semantic field of emotions in the conceptual lexicon of Modern Greek (Fotopoulou & Giouli, 2015). This lexicon capitalises on the notion of lexical fields, along with the Saussurean notion of SIGN and its two facets, namely, the SIGNIFIER and the SIGNIFIED as the building blocks (main classes) of a linguistic ontology. Top-level classes under SIGNIFIED class roughly correspond to semantic fields/domains; words are mapped onto domain-specific concepts in the ontology and they are interlinked via a rich network of relations.

3.2 The data model

The data model then offers the possibility of comparative cognitive linguistic analysis, markup for etymological information using a description standard like ONTOLEX (Chiacros, forthcoming), TEI13 (Schopper, Bowers & Wandl-Vogt, 2015) and possibly TBX, consequently linking various sources. This is particularly relevant for using and displaying such data on online dictionary portals, such as the European Dictionary Portal, where our data model potentially offers access to detailed diachronic and synchronic lexicographic knowledge across a variety of languages. The synergy of etymological, contextual and cross-language sources makes our suggested model particularly attractive, compared to others often drawing on electronic editions of single-language dictionaries or otherwise collected digital corpora.


13 www.slideshare.net/JackBowers1 Accessed [28/04/2016]
The infrastructure under development assumes the form of a linguistic ontology and provides descriptions beyond roots, at various layers of linguistic analysis. Thus, the comparative study of data across languages and/or semantic fields and from a synchronic/diachronic perspective is deemed feasible. The model comprises lexical-semantic objects (classes) and Data Categories (DCs) in the form of feature-value pairs that are suitable for representing the semantic fields emotion and colours. More specifically, the semantic modelling of emotion-related terms was based on a psychological – and in this sense, language-independent – model of emotions\textsuperscript{14} (Plutchik, 2001), appropriately extended on the basis of corpus evidence (Giouli, Fotopoulou, Mouka & Saridakis, 2013). Emotion-denoting words and phrases are mapped to one or more emotion concepts defined; generic semantic relations (i.e. hyponymy, hyperonymy, holonymy, etc.) and meaningful relations that are specific to the semantic field are also employed (i.e. feelsEmotion, triggersEmotion, etc.). Moreover, in order to better account for emotion-related troponyms, other feature-value pairs have also been adopted, and emotion is described as a multidimensional concept. In this way, feature Intensity with possible values: “low”, “medium”, “high”, and “uncertain” is employed to distinguish words/phrases denoting emotion of the same type, yet of different intensity (i.e., fear and terror in English). Finally, the semantic orientation of emotion is also encoded by using the feature Polarity with possible values “positive”, “negative” and “neutral”.

Our model, therefore, provides the means for mapping forms (lexical entries in their synchronic view or etymological roots and variants) to concepts. The model can be easily extended and customised once other senses or properties need to be defined; lexicalizations that correspond to obsolete senses or senses that are not already available in the current scheme can, thus, be captured.

4 The Pilot Study: Red and its European Cultural Connections

Following the initial idea raised at the COST meeting (see 1. above), we look at red in this pilot paper and explore both its European cultural connections and its link to the emotion anger. Red is particularly interesting since its cultural associations are well documented (Deutscher, 2010; Berlin & Kay, 1969; Gage, 1993; Vasilievich, Kuznetsova & Mischenko, 2005; Ciolaneanu, 2015). It is associated both with strong positive (love, vitality, passion) and negative (aggression, anger, danger) connotations. From an etymological perspective and from the data-modelling point of view, this is highly interesting to show, across different languages, which etymological term \textit{red} developed from, which concepts it is related to, and how the colour and related hues (e.g., crimson, vermilion) developed over time and how we perceive them today. Derived words, phrases and idioms, e.g. in relation to emotions, provide further cultural as well as linguistic insights. Table 1 shows example phrases with \textit{red} across different languages connected with the concept of the emotion \textit{anger}. Blank fields indicate the use of other constructions or expressions.

<table>
<thead>
<tr>
<th>Portuguese</th>
<th>Spanish</th>
<th>Italian</th>
<th>French</th>
<th>Romanian</th>
<th>German</th>
<th>Russian</th>
<th>Greek</th>
</tr>
</thead>
<tbody>
<tr>
<td>ver rojo</td>
<td>vedere</td>
<td>voir rouge</td>
<td>a vedea roșu in fața ochilor</td>
<td>rot sehen</td>
<td>κοκκινίζω από το κακό μου [kokinizo apo to kako mu]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rosso</td>
<td>rouge</td>
<td>rouge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>comme un chiffon rouge</td>
<td>ein rotes Tuch sein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Example phrases across different languages. Row 1: “to see red” (= to get angry). Row 2: “to be like a red rag (to a bull)” (= to make someone angry)

“Spanish (rojo), French (rouge), German (rot) and Italian (rosso) take the Indo-European root ‘h₁reudʰ’, (see Latin ruber/rubus) for red, but Portuguese (vermelho) favoured the Latin noun vermiculus as the unmarked designation for red” (Villalva, Silvestre & Pacheco (2014), p.1).

In Modern Greek, adjectives of diverse etymological origin (Indo-European and ancient Greek) lexicalize the notion ‘red’, namely, ροδίνος [rodinos], ροδαλός [rodalos], ερυθρός [erythros] and κόκκινος [kokinos]. These adjectives and their derivatives (nouns and verbs) are quite often found to occur in expressions that are conceptually related to emotions. For example, the adjective ροδίνος (= light red, as in the colour of roses) occurs in the phrase τα βλέπω ολα ρόδινα / themPL.ACC.PRON see1SG all red (= lit. to see everything rose-red, to be optimistic) bearing a positive connotation; on the contrary, words etymologically related to the adjective κόκκινος ([kokinos], red) are found to occur in idioms that denote a negative emotion (anger or shame) of a high intensity, i.e. in the expression κοκκινίζω από το κακό μου / become-red1SG from the bad my1SG.GEN.PRON (= lit. to get red from rage, become irritated or very angry – see also Table 1 above).

In many Slavic languages, also exist adjectives with the root ‘h₁reudʰ’. In Russian it is used only in some dialects, but not in the literary language. Idioms connected with red (красный) and anger include действовать/ подействовать, как красная тряпка на быка ‘to be like a red flag to a bull’. Non-idiomatic collocations, with red connected to different emotions in Russian include красный от стыда 'red with shame', красный от смущения 'red with embarrassment', красный от гнева 'red with anger', and similar expressions with the verb краснеть / покраснеть 'turn red' (derived from красный 'red'): краснеть / покраснеть от стыда, смущения, гнева 'turn red with shame, embarrassment, anger'. In German ‘rot vor Wut’ denotes the colour of someone’s face when they are very angry. As for hues of red (e.g., crimson) in Russian not derived from красный (red), but also used in phrases and idioms related to anger and rage, include багровый (crimson; sometimes translated purple\(^{15}\)) and багроветь 'turn crimson'. In idioms of comparison the context determines

\(^{15}\) — Извольте принять! — крикнул он, весь багровый от гнева, — я не обязан стоять над вами; а то после скажете, что не получили. Сосчитайте. Достоевский Ф.М. "Подросток" (National corpus of the Russian language [http://www.ruscorpora.ru]) — “Kindly take this!” he shouted, purple with rage. “It’s not for me to stand over you, but if I don’t you’ll be saying afterwards you haven’t had the money. Count it.” Fyodor Dostoyevsky, A Raw Youth ([https://ebooks.adelaide.edu.au/d/dostoyevsky/d72r/complete.html](https://ebooks.adelaide.edu.au/d/dostoyevsky/d72r/complete.html)). [accessed: 7/11/2015]
whether there is a connection with emotions in Russian, e.g. красный как рак (as red as a crawfish),
when a person is ‘as red as a crawfish’ because of anger or heat. In Portuguese, ficar vermelho como um pimentão ‘to turn red like a chili pepper’ (Spanish: ponerse rojo como un tomate ‘to turn red like a tomato’), can be used in relation to shame (English: to blush), but also for anger.
Furthermore, the word red may even be absent if a typically red entity is mentioned, as in Portuguese parecia um tomate (‘he looked like a tomato’) having the same meaning (shame or anger), as ficou uma lagosta (‘he turned a lobster’) meaning someone’s skin is red because of sunburn.

By means of the proposed Root dictionary and digital data modelling, detailed linked linguistic information can be made available in the lexicographic context, providing specific etymological and also diachronic information across languages, linking various linguistic fields.

5 The European Dictionary Portal: Displaying Red and Anger

The European Dictionary portal offers the possibility of searching interactively for words across a variety of mono-lingual and bilingual dictionaries. The integration of the proposed Root dictionary offers an opportunity to interlink languages on the historical concept level, which was relevant for the naming of words. It might give ideas on cultural connections beyond our recent nation areas. Furthermore, it allows to discuss historic and semantic connections and their change of concepts in time. Finally, the data modelling is supporting an open resource with the aim to offer some kind of “new encyclopaedias” where words are connected to and embedded in their real world context.

6 References


