

2. Full report on research activities and results

1. Background

1.1. Relevant previous research

Our aim was to take the data of the colour terms of the contemporary Slavonic languages, and induce inheritance hierarchies which could be interpreted as representing the historical development of the language. To do this we required a suitable formalism. This was provided by DATR, a typed lexical knowledge representation language developed to express default inference, which the researchers have used in previous lexical representation work.

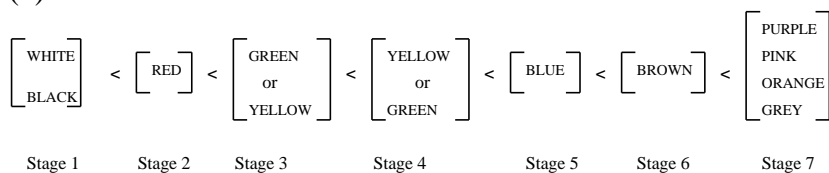
The Slavonic colour system is of special interest in that one of its members, Russian, has moved beyond the eleven basic colour terms of Berlin and Kay's theory. Corbett and Davies worked together to establish the basic status of the 'twelfth' term *goluboj* 'light blue' (Corbett and Davies 1997). Much of their work has also concentrated on field methods for establishing Berlin and Kay basic colour terms. An important research resource we used is the *Slavonic Languages* monograph (Comrie and Corbett 1993), where each of the language chapters dedicates a section to colour terms. This has allowed for a comparison to be made between the findings of our informant work and the standard assumptions about the core colour vocabulary of the Slavonic languages. The Slavonic languages are relatively close genetically, and have proved rich ground for historical research.

1.2. Theoretical bases and hypotheses

We followed the simple principle that language phenomena, in this case lexical knowledge, remain by default unchanged. Any changes which occur are expressed as overrides to the default situation, or in the case of subsequent borrowings or other innovations, as additions to the existing knowledge base. Putative basic terms of contemporary languages belonging to the same language family were elicited from a recognized psycholinguistic test. These were used to induce a default inheritance hierarchy, representing an hypothesis about the historical development of the language family. Using this default framework we investigated the history of the colour terms of Slavonic.

We were also interested in the well-known typological constraints on colour lexicons provided by Berlin and Kay (1969), shown in (1).

(1)



The implicational hierarchy in (1) can be seen as a constraint on the evolution of basic colour terms in a language. In the lexicalization of basic colour categories, languages evolve through Stages 1 to 7, and (1) shows the order in which the categories are lexicalized. Hence a Stage 5 language with a BLUE term must have emerged from a Stage 4 language which lacked a BLUE but had terms for WHITE, BLACK, RED, YELLOW and GREEN. This language may in turn evolve to Stage 6 where a basic term for BROWN will be added. Note that at Stage 7 there is no predicted ordering with respect to the lexicalization of PURPLE, PINK, ORANGE and GREY.

The typological constraints were important as confirmation for our default reconstruction model. It turned out that our model showed Proto-Slavonic as a Stage 4 language and not Stage 5, as suggested in Priestly (1987). Moreover the model pinpointed areas where Slavonic adds interesting variation to the typology.

2. Objectives

2.1. Descriptive objective

Our descriptive objective was to be in a position to give an account of the Slavonic colour lexicon. We have gathered valuable data about the Contemporary Slavonic languages for a synchronic account. Using descriptive resources such as dictionaries and specific works on the history of colour terms in different Slavonic languages, we have compiled detailed datasets of East, West and South Slavonic for a diachronic account (see Deliverables). Our model has supplied ‘missing’ data using its default inheritance mechanism.

2.2. Typological objective

Colour terms constitute a well defined area for cross-linguistic comparison, allowing for the establishment of linguistic universals, for example those outlined in Berlin and Kay’s (1969) basic colour terms typology. A well-known problem case for the typology, which was commented on in the original Berlin and Kay monograph, is the Russian colour system. Given that the Russian colour system is unusual, we explored the possibility that Slavonic may present further challenges to the Basic Colour Terms theory. We discovered two main areas where Slavonic is at variance with the standard theory, namely the BLUE category in East Slavonic, and the evolution of the GREY category throughout Slavonic.

Using a recognized psycholinguistic test, the list-task, we produced evidence that both Ukrainian and Belarusian have two basic terms for BLUE, just as Russian (Hippisley 2001). The first term is inherited from East Slavonic *sinij* denoting ‘(dark) blue’, and is the term found in Russian. The second is a borrowing from Polish *blekitny* which in both Ukrainian and Belarusian means ‘light blue’. Though Russian has clearly influenced the development of the Ukrainian and Belarusian colour systems in this way, the Russian basic term itself *goluboj* has not been borrowed. It appears that a category, in this case the second BLUE, may be borrowed separately from the basic term that denotes it. Thus the evolution of categories operate independently of the shifts in basic status of the terms themselves. While a second BLUE category challenges the original Berlin and Kay theory, it can be accommodated in later revisions, notably Kay and McDaniel (1978), and Kay, Berlin, Maffi and Merrifield

(1997). The six elemental colours WHITE, BLACK, RED, GREEN, YELLOW, BLUE combine to form compound colours. The development of other compounds other than the original four cited in Berlin and Kay are therefore theoretically possible, such that the evolutionary process may be taken beyond the basic eleven categories. Thus LIGHT BLUE is taken to be the combination of the elementals BLUE and WHITE, just as PINK is the combination of RED and WHITE (Kay and McDaniel (1978: 640)).

The category GREY was revealed to be of typological interest. One of the results of the default reconstruction model was the placement of terms for GREY at nodes in the hierarchy representing the historical split between West Slavonic, East Slavonic and South Slavonic. Three main terms were established for the contemporary languages, elicited by the list-task. These are derived from either Proto-Slavonic *si-v* or the Early-Proto-Slavonic set *xoi̯-r* and *xoi̯-d*. Thus *si-v* is found for basic GREY in all the South Slavonic languages (e.g. Macedonian *siv*, Serbo-Croat *siv*), but *xoi̯-r* for all of East Slavonic (Ukrainian *sir*, Belarusian *šër* and Russian *ser*) and some of West Slavonic (Polish *szar*, Upper Sorbian *šer* and Lower Sorbian *šer*) and *xoi̯-d* for the rest for West Slavonic (Czech *šed* and Slovak *šed*). This split across the Slavonic languages for GREY is indicative of an innovative category. Again it challenges the standard theory, but accommodation is made for the GREY category in Berlin and Kay (1969: 45) where it is provided with a special ‘wildcard’ status, allowing it to appear anywhere after Stage 4. Given that the next split in our model concerns terms for BLUE it would appear that in Slavonic GREY emerges at a stage between Stage 4 (GREEN) and Stage 5 (BLUE).

2.3 *Theoretical objective*

The theoretical objective was to induce a default inheritance hierarchy from synchronic data combined with a typological constraint to provide a model for the historical development of a language family. The underlying notion of defaults would be used to capture the principle that language phenomena remain largely unchanged; any overrides would mark an innovation. We developed a model of lexical change in one well defined area of the lexicon, colour terms, that has taken place in a single language family, Slavonic. Incorporating techniques from the field of Artificial Intelligence, we made use of two kinds of reasoning, induction and default reasoning, thereby capturing reconstruction on the one hand, and lexical development on the other. We used overrides to pinpoint areas of obsolescence, innovation and semantic shift that have taken place in the Slavonic colour lexicon. The model was typologically informed, incorporating Berlin and Kay’s claims about lexical change in colour terms. It was also psycholinguistically informed, operating on data that is the result of an established psycholinguistic test to elicit basic colour terms. Finally the model was formalised, allowing it to be computationally tested.

2.4 *Methodological objective*

The methodological objective was to use data about contemporary languages belonging to a single family to induce a default hierarchy that could be interpreted in terms of the historical development of that language family. This was achieved by selecting appropriate computing tools. DATR is a computable lexical knowledge

representation language which defines networks by links typed by attribute paths through which information is inherited (Gazdar 1990: 1). Due to its principle of default inference, the networks are specifically non-monotonic, or default. A generalization locating tool was then used, Gazdar's 'cin-din', which takes DATR representations as input which it converts into DATR-expressed default inheritance hierarchies (Hippisley and Gazdar 2000: 129). Gathering the data met another methodological objective: determining what constitutes a *basic* colour term. A reasonable index of basicness is psychological salience, which can be measured by the list task. To gather synchronic data on the basic terms of Slavonic we carried out the list-task on speakers of 11 Slavonic languages. Additional tests were used for Sorbian speakers.

3. Methods

A first step was to gather synchronic data on basic colour terms. This was then used to drive the default inheritance model, which in turn would provide a context for investigating diachronic data, as well as the typological constraints laid down by Berlin and Kay.

3.1. Establishing a synchronic account

To establish the Slavonic basic colour terms in a synchronic context we consulted the *Slavonic Languages* monograph where authors of the language chapters have produced lists of basic colour terms, primarily using dictionaries and textual frequency counts as guides. Later in the project, we decided to extend the method. We tested the psychological salience of colour terms amongst native speakers of the language, as a guide to their basic status, in accord with Berlin and Kay. We carried out the list task amongst informants of each of the Slavonic languages. Informants are required to list as many colour terms as they can think of within a short space of time. Frequency of terms and their position on the list are measures of their salience. In each case questionnaires with instructions written in the native language were given to informants. In most cases, native speakers conducted the test. The results were then compared with the dictionary and text based findings. Basic terms show strong agreement across informants.

We also decided to direct special attention to the two endangered Slavonic languages Upper Sorbian and Lower Sorbian. A field trip was organized and a number of salience test carried out. As well as the list task, informants performed the 'colour naming' task and the 'best example task'. In the naming task, colour tiles representative of the colour space have to be named by an informant. The salient terms are marked out by high frequency of occurrence and high degree of consensus in the tiles they denote. In the 'best example' task, informants declare which tile is the best example of a supposed salient term (see Hippisley and Davies 2001).

3.2. Establishing a diachronic account

Default inheritance allows the structuring of information such that general information is encoded near the top of an inheritance hierarchy, while more specific information is located lower in the hierarchy. Information is retrieved from the bottom

of the hierarchy upwards, with more specific information overriding more general. An obvious application is to historical reconstruction of a language family. Shared linguistic facts and, by implication, a shared history, are captured by abstracting to a common node from which daughter language nodes inherit. Since the inheritance employed is default, inherited information may be overridden as well as augmented, thus allowing a language's unique identity to emerge. Figure 1 shows a default inheritance hierarchy of Slavonic. Generalizations about the Slavonic languages appear at abstraction nodes: East Slavonic will contain facts generalizable for Russian, Ukrainian and Belarusian. Facts general to Slavonic as a whole can be stated at the root node COMMON_SLAVONIC. The individual languages are represented as leaf nodes, inheriting generalizations about Slavonic, as well as containing language specific material.

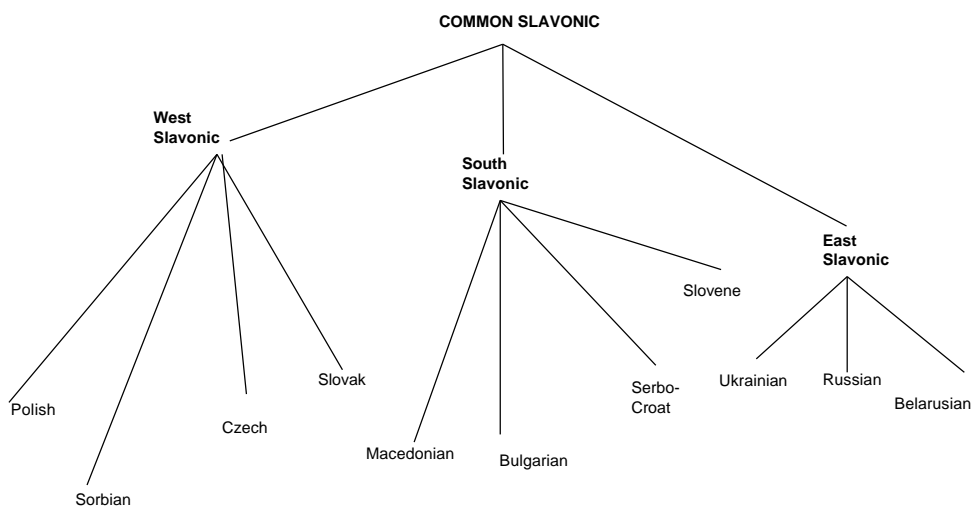


Figure 1: Slavonic as an inheritance tree

The inheritance hierarchy in Figure 1 makes a number of claims. It states that there are different kinds of facts. For example, there are facts which are true only of Polish, and they will be recorded at the Polish leaf node. There are also facts general to West Slavonic, but not East and West Slavonic. They will be situated at the West Slavonic node. Finally there are generalizations which hold for Slavonic as a whole, and are placed at the root node COMMON_SLAVONIC. The question is how such a default hierarchy relates to historical reconstruction. The claim is that the organization of defaults and overrides is related to historical development. The highest level defaults, those stored at the root node COMMON_SLAVONIC, are facts about the reconstructed ancestor language Proto-Slavonic. Lower level defaults, such as those at West Slavonic, mark a West Slavonic innovation, etc. Finally facts at the Polish node mark a distinctly Polish innovation. Our important claim is that 'by default' hierarchies stating generalizations about contemporary Slavonic are a manifestation of the historical development of Slavonic. In other words, the shape of a hierarchy representing contemporary Slavonic shares by default the shape of a hierarchy representing the historical Slavonic family tree.

The method involved two major processes, the first where inheritance hierarchies were induced from contemporary data, and the second where the hierarchies are 'left

to run' to deliver theorems about the history of the language family. In the first process, an East Slavonic hierarchy was induced using basic colour term data from Russian, Belarusian and Ukrainian. Data from Polish, Czech, Slovak were used to induce a West Slavonic hierarchy, and similarly data from Bulgarian, Macedonian, Serbo-Croat and Slovene induced a South Slavonic hierarchy. In Figure 2 we have the induced East Slavonic hierarchy.

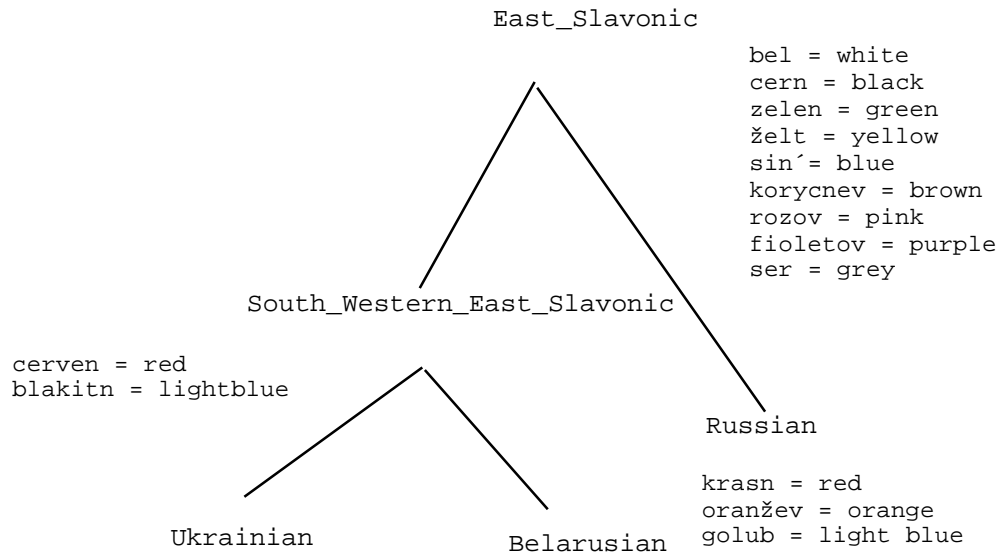


Figure 2. The induced East Slavonic hierarchy.

The data induce a hierarchy containing a root node, which we can label as `East_Slavonic`, from which the daughter languages inherit. An additional abstraction node has been generated, one which generalizes over Ukrainian and Belarusian. How does the default induction relate to the historical situation? In fact it rather elegantly captures the historic division of East Slavonic into the North-Eastern dialect (Russian) and the South-Western dialect (Belarusian and Ukrainian), as discussed for example in Schenker (1993: 116). The hierarchy shows how the basic colour terms are largely shared across the three languages. The differences separate Russian from the other two. Once we abstract away from phonological differences, the generalization locating tool could make no further distinctions for basic colour term between Ukrainian and Belarusian. As an historical hypothesis, it claims that East Slavonic split into two groups, one of which draws Ukrainian and Belarusian together. The root nodes (higher level nodes) of each hierarchy were then assembled to induce a Slavonic languages hierarchy, shown in Figure 3.

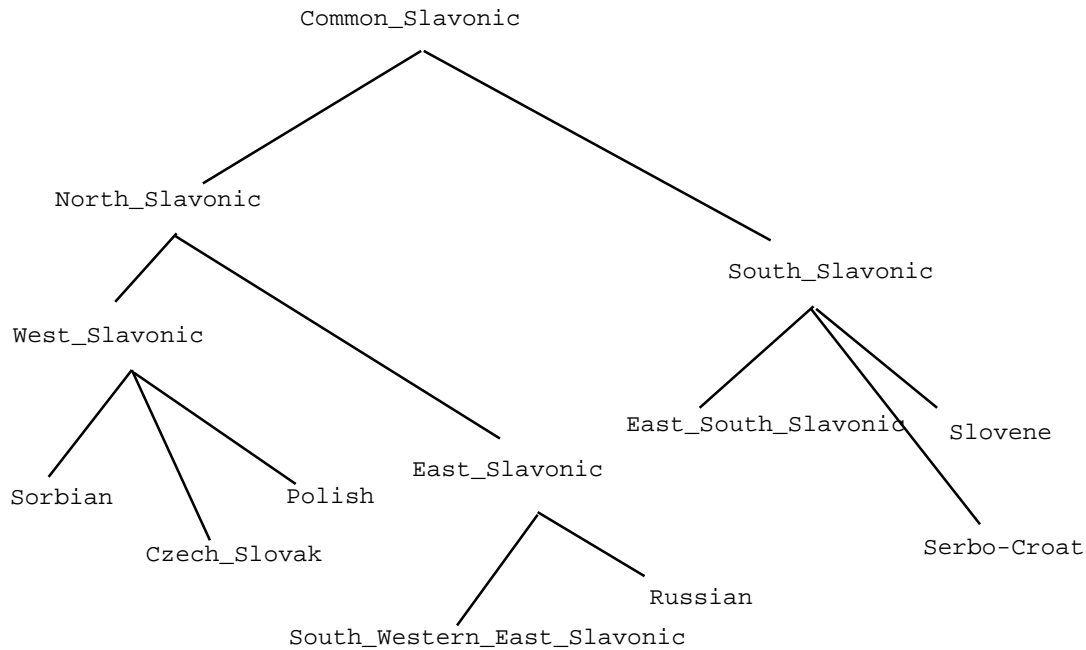


Figure 1. The major structure of the induced Slavonic languages hierarchy (some languages omitted for reasons of clarity)

The second major process is to treat these induced hierarchies as historical trees. The facts stated at the root node of the hierarchy will be inherited down the hierarchy by default. This is to be interpreted as historical inheritance. Where inheritance has to be overridden for any reason the historical interpretation is obsolescence, semantic shift or lexical innovation. The default-override mechanism has two important consequences for an account of colour terms, one a consequence of inheritance and the other of overriding. For inheritance where there are ‘gaps’ at nodes representing a language at a given historic period these will be ‘filled in’ by default, and in this way we can recover missing information concerning the lexical history. Where basic terms are overridden this will point to a shift to non-basic status due to lexical replacement; and where non-basic terms are overridden, this will mark a shift to basic status.

4. Results

4.1 Synchronic account

We compared the inventories of basic colour terms deduced from dictionaries and other literary sources summarised in *Slavonic Languages* with that produced from the findings of the list task. The results give a general picture of the basic colour terms of the contemporary standard languages of Slavonic, and are given in the following tables.

4.1.1 East Slavonic

Table 1: Basic colour terms in East Slavonic.

<i>Colour category</i>	<i>Russian</i>	<i>Ukrainian</i>	<i>Belarusian</i>
<i>Elemental</i>			
WHITE	bel	bil	bel
BLACK	cern	corn	corn
RED	krasn	cervon	cyrvon
GREEN	zelen	zelen	zjalën
YELLOW	želt	žovt	žou t
BLUE	sin´	syn´	sin´
<i>Compound</i>			
BROWN	korycnev	korycnjav	karycnev
PINK	rozov	rožev	ružov
PURPLE	fioletov	fioletov	fijaletov
ORANGE	oranžev	-	-
GREY	ser	sir	še^r/ser
LIGHT BLUE	golub	blakytn	blakitn

All three East Slavonic languages have the additional LIGHT BLUE category. Two of the three (Ukrainian and Belarusian) lack an ORANGE category.

4.1.2 West Slavonic

Table 2: Basic colour terms in West Slavonic.

<i>Colour category</i>	<i>CZECH-SLOVAK</i>		<i>SORBIAN</i>		<i>LECHITIC</i>
	<i>Czech</i>	<i>Slovak</i>	<i>Lower Sorbian</i>	<i>Upper Sorbian</i>	<i>Polish</i>
<i>Elemental</i>					
WHITE	bíl	biel	bel	bel	bial
BLACK	cern	ciern	carn	corn	czarn
RED	cerven	cerven	cerwjen	cerwjen	czerwon
GREEN	zelen	zelen	zelen	zelen	zielon
YELLOW	žlut	žlt	žolt	žolt	zólt
BLUE	modr	modr	modr	módr	niebiesk
<i>Compound</i>					
BROWN	hned	hned	brun	brun	brazow
PINK	ružov	ružov	rožow	róžow	rózow
PURPLE	fialov	fialov	lylow	fijalkow	fioletow
ORANGE	oranžov	oranžov	-	-	pomaranzczow
GREY	šed	šed	šery	šer	szar
LIGHT BLUE	-	-	-	-	-

Note that the Sorbian group is distinguished by its lack of an ORANGE category. Lechitic (represented by Polish) is distinguished by its basic term for BLUE *niebiesk*, as opposed to *modr* for the rest of West Slavonic. Terms for BROWN act to distinguish all three groups: *hned* / *hned* for Czech-Slovak, *brun* for Sorbian and *brazowy* for Lechitic (Polish). Terms for GREY distinguish Czech-Slovak from the rest of West Slavonic: *šed* versus *šer-* / *šer-* / *szar-*. Finally there is no evidence of a LIGHT BLUE category.

4.1.3 South Slavonic

Table 8: Basic colour term in South Slavonic.

Colour category	EASTERN SOUTH SLAVONIC		WESTERN SOUTH SLAVONIC	
	Bulgarian	Macedonian	Serbo-Croat	Slovene
<i>Elemental</i>				
WHITE	bjal	bel	beo	bel
BLACK	ceren	crn	crn	crn
RED	cerven	crven	crven	rdec
GREEN	zelen	zelen	zelen	zelen
YELLOW	ža lt	žovt	žut	rumen
BLUE	sin	sin / plav	plav	modr
<i>Compound</i>				
BROWN	kafjav	kafeav	braon	rjav
PINK	rozov	-	-	roza
PURPLE	lilav	violetev	ljubicast	vijolien
ORANGE	oranžev	portokalov	narancast	oranžn
GREY	siv	siv	siv	siv
LIGHT BLUE	-	-	-	-

The ‘missing’ category is PINK, affecting one language from each South Slavonic group: Macedonian for Eastern South Slavonic, and Serbo-Croat for Western South Slavonic. BLUE is interesting, since each language behaves differently: Bulgarian has *sin*, Macedonian has *sin* for some speakers and *plav* for others, Serbo-Croat has *plav* and Slovene has *modr*. BROWN is also an interesting area. First the non-basic term *oker* is highly salient in all languages except for Bulgarian, with a ranking of 11th or 12th. And second, three terms are used amongst the four languages: *kafjav* / *kafeav* for Eastern South Slavonic, *braon* for Serbo-Croat and *rjav* for Slovene. Each of these is basic in the relevant language. No language has a basic term for ‘light blue’, in contrast to the East Slavonic languages.

4.2. Diachronic account

We have operated on a limited area where there are typological constraints, and on a family where solid comparative work has already been done. This has allowed us to check the predictions of the model, both typologically and in relation to conventional wisdom.

First, the structure of the induced inheritance tree that emerged came close to the conventional wisdom about the shape of the language family's historical tree (see Figure 4). For example the induction process has drawn together East and West Slavonic together, under one node. Remarkably this is entirely consistent with the historical view: for example Schenker (1993: 114-115) on the basis of morphological evidence groups East and West Slavonic together as North Slavonic, the label we adopt for this node. This is a deliberate choice of label. Schenker gives a morphological example as evidence for a North versus South Slavonic split. The historical class of *-o-* stem verbs has an instrumental in *-om?* in South Slavonic and *-?m?* in North Slavonic. For example for 'year' we have *god-om?* in Old Church Slavonic, an extinct East South-Slavonic language, and *god-?m?* in Old Russian, the ancestor language of Russian, Ukrainian and Belarusian.

Second, if the hierarchy represents the historical development, the central question is, how does the distribution of the colour terms match up to the evolutionary claims of Berlin and Kay (1)? Terms for BLACK, WHITE, RED, GREEN and YELLOW appeared at the root node *Common_Slavonic*. All of these are Proto-Slavonic basic terms, as claimed in Schenker (1993: 111-12). They correspond to Berlin and Kay Stages 1 to 4. In other words, our default induction method has produced a hypothesis based on the synchronic data of the list task which is consistent with Berlin and Kay's evolutionary claims, and reflects the historical situation, as outlined by Schenker amongst others. Do we find a similar story as we move along Berlin and Kay's evolutionary stages? The category that evolves at Stage 5 is BLUE, and we find different BLUE terms at the East, West and South Slavonic nodes: at East_Slavonic we have *sin´-*, at West-Slavonic *modr-* and at South Slavonic *plav-*. Thus the hypothesis fits the evolutionary claim: BLUE was the next basic term to evolve, and as such was an innovation for each of the three main branches. Terms for GREY also appear 'high up' in the hierarchy, and act to split North and South Slavonic. The induced hypothesis based on an induced hierarchy of contemporary data is that GREY evolves at an earlier stage than BLUE. However, we discussed this as possibility for Slavonic in section 2.2. From the hierarchy generated in (9) the hypothesis is that it is positioned between Stage 4 GREEN and Stage 5 BLUE. According to our model, the next stage is the emergence of BROWN, again a result consistent with the typology. This is dramatically shown in West Slavonic in Figure 5.

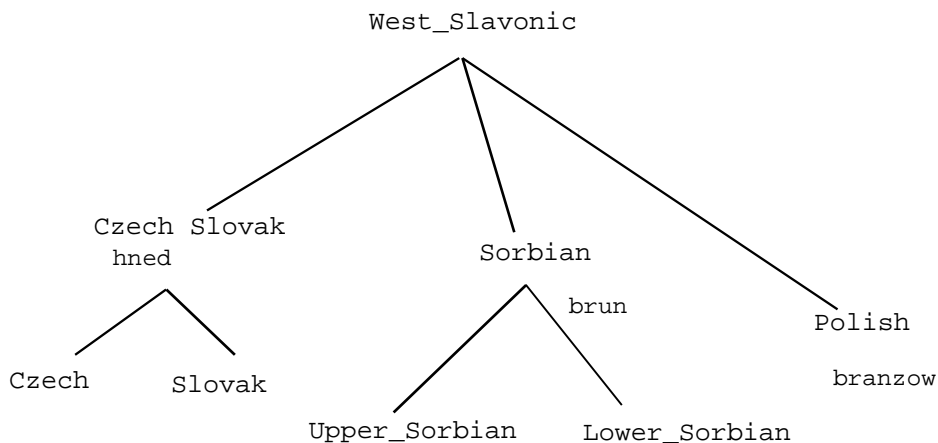


Figure 5. BROWN as emergent in West Slavonic

Here we see different terms for each of the main branches of West Slavonic. However some terms were found to be in the ‘wrong’ place, notably PURPLE which appeared at the root node. There is no evidence of a PURPLE terms in Proto-Slavonic. A separate hierarchy was constructed representing the stages of lexicalization colour categories, according to Berlin and Kay. The Evolutionary Colour Stages hierarchy was placed orthogonally to the main Slavonic languages hierarchy. In this way we were able to use the Berlin and Kay theory to constrain the availability of terms at the various historical stages of Slavonic. An important aspect of the Evolutionary Colour Stages hierarchy is that it can only prevent the evaluation of terms, but cannot generate them. In other words, the constraining hierarchy functions only to refine the hypothesis that has already been induced.

As a default inheritance tree the model acted to ‘fill in the gaps’ about the history of Slavonic. Overrides were used to pinpoint areas of obsolescence, innovation and semantic shift. For example, the Proto-Slavonic term *siv* ‘grey’ is overridden in Contemporary South and West Slavonic since it is now obsolete in languages belonging to these branches. Lexical replacement is captured as the overriding of a basic term in favour of new term, for example the replacement of *cervlen(yj)* by *krasn(y)* in C18 Russian. The shift in *gned* from a restricted term in Proto-Slavonic to an unrestricted basic term in Czech is captured by the overriding of this non-basic term at the Czech node.

We have worked on languages where there is plentiful data to establish relationships within the family. However, we allowed our model access only to a minute portion of the data, namely the colour term inventory. Our results show a remarkably close fit to the received wisdom based on the full data. This demonstrates the success of the project and the plausibility of our hypothesis.

References

- Berlin, B. & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley and Los Angeles: University of California Press.
- Comrie, B. & Corbett, G. G. (eds.). (1993). *The Slavonic Languages*. London: Routledge.

- Corbett, Greville G; and Davies, Ian R. L. (1997). Establishing basic color terms: measures and techniques. In Hardin, C. L. & Maffi, L. (eds.), *Colour categories in thought and language*. Cambridge: Cambridge University Press. 197-223.
- Gazdar, G. (1990). An introduction to DATR. In Evans, R. & Gazdar, G. (eds.), *The DATR papers I*. Cognitive Science Research Paper CSRP 139. Brighton: University of Sussex. 1-14.
- Hippiisley, A. (2001). Basic BLUE in East Slavonic. *Linguistics* **39** (1). 151-179.
- Hippiisley, A. & Gazdar, G. (2000). Inheritance hierarchies and historical reconstruction: towards a history of Slavonic colour terms'. *Chicago Linguistics Society 35: Main Session*. Chicago: Chicago Linguistics Society.125-140.
- Hippiisley, A. & and Davies, Ian R. L. (2001). Basic colour terms of Lower and Upper Sorbian. University of Surrey manuscript.
- Kay, P. & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language* **54**. 610-646.
- Priestly, T. (1987). On Basic Color Terms in Early Slavic and Ukrainian. *The Annals of the Ukrainian Academy of Arts and Sciences in the United States* **15** (39-40). 243-251.
- Schenker, A. M. (1993). Proto-Slavonic. Comrie, B. & Corbett, G. G. (eds.), *The Slavonic Languages*. London: Routledge. 60-121. London: Routledge.