Dutch colour terms

Frank Van Eynde, Vincent Vandeghinste,
Leen Sevens & Ineke Schuurman

Centrum voor Computerlinguïstiek

From a physical point of view all colours are equal, but our senses and our minds treat some as more equal than others. Red and blue, for instance, are generally felt to be basic, while crimson and indigo are not. The question why has occupied philosophers, artists, physicists and psychologists over the centuries. Among the most famous are Aristotle, Leonardo da Vinci, Isaac Newton, Wolfgang Goethe, Arthur Schopenhauer and Ludwig Wittgenstein. Linguists have also contributed to the debate, since the distinction between basic colours and other colours is also reflected in our languages.

A seminal work from the linguistic side is Berlin & Kay (1969). Unconvinced by the then prevailing view in linguistics and anthropology that the colour spectre is divided differently in different languages and that there is, hence, no *raison d'être* for semantic universals in the field of colour terms, Brent Berlin and Paul Kay did a comparative investigation of the colour terms of 98 languages from different language families. It provided them with evidence for two universals. The first concerns the existence of "a total universal inventory of exactly eleven basic colour categories ... *white, black, red, green, yellow, blue, brown, purple, pink, orange* and *grey*." (o.c., 2). The second concerns a partial ordering in this inventory. It is depicted in Figure 1 (o.c., 4). The left to right order reflects a series of implications: If a language has a term for the colour at the right hand side of an arrow then it also has terms for the colours at the left hand side. For instance, a language which has a term for blue has also terms for white, black, red, yellow and green, but not necessarily for grey or brown.

The distinction between basic colours and other colours has also drawn the attention of Dany Jaspers, as evidenced by Jaspers (2012), and it will probably keep drawing his attention for some time to come, as evidenced by Jaspers (ms), a piece of work in progress.

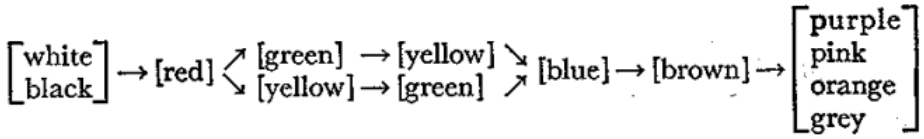

Figure 1: Berlin & Kay's hierarchy of basic colour categories

1 Colour percepts and colour terms

In contrast to Berlin & Kay, who started from a crosslingual inventory of colour *terms*, Jaspers starts from human colour *percepts*. More specifically, his point of departure is the anatomic fact that the human retina contains three types of cones which are activated by the wavelengths of red, green and blue light respectively. Using bits to represent activation, the colours are represented by bitstrings of length three, where 100 stands for the percept *RED*, 010 for the percept *GREEN* and 001 for the percept *BLUE*. These are the level 1 colours. The level 2 colours involve activation of pairs of cone types. *YELLOW*, for instance, involves activation of the cones for red and green (110), *MAGENTA* of the cones for red and blue (101) and *CYAN* of the cones for green and blue (011). Activation of all three cone types corresponds to the level 3 colour *WHITE* (111). Conversely, if none of the cones is activated we get the level 0 colour *BLACK* (000). A summary is provided by the Hasse-diagram in Figure 2. The boxed colour categories stand out as nonbasic. To name them "one has to resort to nonbasic (non-naturally learned) specialised colour terms such as *cyan* ... or alternatively to binary lexicalisations such as Dutch *appelblauwzeegroen* (appleblue-seagreen) or complex expressions such as *bluish green* or *greenish blue*." (Jaspers, 2012, 234)

In spite of the differences in approach, there are some striking similarities between Berlin & Kay (1969) and Jaspers (2012): Both take a universalist stance and elaborate a system in which colour terms have a stable extra-linguistic denotation, grounded in a collection of 329 colour chips by Berlin & Kay and in the effect of wavelenghts on the retina by Jaspers. On a more detailed level, it is noteworthy that the inventory of basic colour percepts in the Hasse-diagram is identical to the inventory of basic colour terms in languages with six colour terms in the Berlin & Kay hierarchy, see Figure 3. The main difference concerns the treatment of the other colour terms, such as *brown* and *orange*. Berlin & Kay include these among the basic colour categories, while Jaspers treats them as

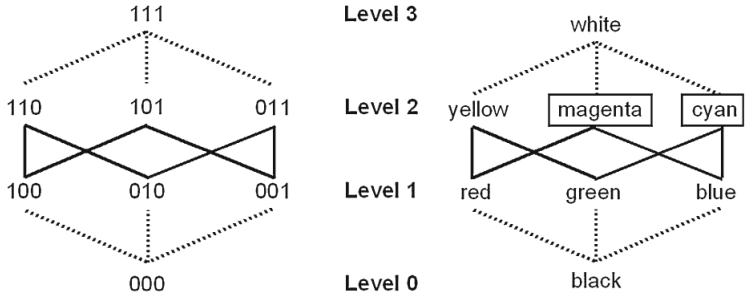

Figure 2: Jaspers' Hasse-diagram for colour percepts

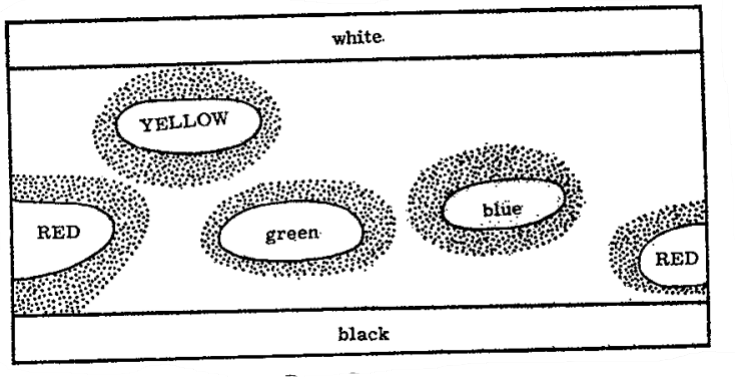

Figure 3: Colour terms of a stage V language in Berlin and Kay (p. 20)

'in-between' colours: "for in-between colours (ORANGE for instance), the range of possible bitstring values has to be increased from just 0 and 1 to a system with in principle any number of possible values in between 0 and 1." (Jaspers, 2012, 244).[1]

This obviously begs the question of what it means for a colour—or a colour term—to be basic. The issue is addressed explicitly in (Berlin & Kay, 1969, 6), which provides four criteria for a colour *term* to be basic:

1. it is *monolexemic*; that is, its meaning is not predictable from the meaning of its parts.

2. its signification is not included in that of any other colour term.

3. its application must not be restricted to a narrow class of objects.

4. it must be psychologically salient for informants. Indices of psychological salience include, among others, (1) a tendency to occur at the beginning of elicited lists of colour terms, (2) stability of reference across informants and across occasions of use, and (3) occurrence in the ideolects of all informants.

The first criterion excludes morphologically complex terms, such as *greenish, salmon-colored*, and paraphrases, such as *the color of the rust on my aunt's old Chevrolet*. The second excludes terms like *crimson* and *scarlet*, since their denotation is included in that of *red*. The third excludes terms like *blond*, whose application is restricted to objects like hair, beer and tobacco. The fourth is used to confirm the results of the other criteria.

For Jaspers it is mainly the colour *percepts* that matter, but the way in which they are expressed in languages also plays a role, as is clear from his remarks about *cyan* and *magenta*: "Logically perfectly conceivable, and perceptually feasible, but, so it appears, not resulting in a percept that constitutes a cognitively natural and focal unary colour, but rather in a binary colour percept, nonnaturally lexicalised in language by means of a specialised term rather than an everyday natural colour term." (Jaspers, ms, 32) What it means for a colour *term* to be natural is not spelled out explicitly. We assume it comes close to Berlin & Kay's criterion of psychological salience.

The purpose of this paper now is to expand a bit on what it means for a colour *term* to be basic. More specifically, we will propose a morpho-syntactic test for

[1] Cone activation is, hence, not a binary matter, but a matter of degree, see (Jaspers, ms, 31).

differentiating basic from non-basic colour terms, not as a replacement but rather as an addition to the criteria of Berlin & Kay. In a nutshell, we will argue that there are two types of colour terms: those which are used both as a noun and an adjective and those which are only used as a noun. Anticipating the results, we will demonstrate that (1) the basic colour terms all belong to the first group, and (2) that the second group is entirely made up of non-basic colour terms. We will use the Dutch colour terms for exemplification, but hasten to add that the criterion is applicable to all languages which differentiate between nouns and adjectives.

To back up our claims we use fragments from the SoNaR corpus Oostdijk et al. (2013). This corpus contains 500 million words and is automatically tagged and lemmatized. To retrieve the relevant data we use OpenSoNaR Reynaert et al. (2014), an online tool for corpus search with queries that are formulated in the Corpus Query Language (CQL) Jakubíček et al. (2010). It allows to combine search in different fields (such as word, lemma or part-of-speech tag) with regular expressions. All examples are taken from the SoNaR corpus and contain an identifier.

2 Two types of colour terms

Syntactically speaking, the Dutch colour terms are singular neuter mass nouns. That they are nouns is clear from the compatibility with a determiner and/or adjectival modifier, as in (1).

(1) Het is een intens *rood* [dpc-cam-001283-nl-sen.p7.s3]

That they are singular and neuter is clear from the fact that they combine with *het*, rather than *de*.

(2) De partij staat voor ruim twee ton in het *rood* [wsuea-233.p6.s3]

That they are mass nouns is clear from the lack of a plural form, modulo coercion, as in (3), where *gelen* stands for kinds of yellow.

(3) De verf bezwijkt onder zijn eigen gewicht en glijdt in een web van *gelen* over de puistige verfkorst. [wrpph-11787.p2.s3]

Many of the Dutch colour terms are also used as adjectives, as in (4).

(4) in het kristal is een heel klein *rood* lampje gemaakt [wsuea-224.p37.s4]

Since the adjective is homophonous to the noun and since their meanings are obviously related, it is not always easy to distinguish them. For that reason we will present some diagnostic criteria, focussing first on their use in prenominal positions (2.1) and then on their use in nominal positions (2.2). An additional criterion, relating to the compatibility with comparative and superlative affixes, is discussed in 2.3.

2.1 In prenominal position

Characteristic of prenominal adjectives is the alternation between declined and non-declined forms. The general rule is that they take the declension affix *-e*, as in *een kleine tafel* 'a small.DCL table', unless the NP is singular, neuter and indefinite, as in *een klein stuk* 'a small piece'. Nouns which occur in prenominal position do not show this alternation: They do not take the declension affix, also in NPs where the affix is obligatory for adjectives, as in *een kleine aluminium tafel* 'a small.DCL aluminium table' and *de goedkopere standaard uitrusting* 'the cheaper.DCL standard equipment'.

Applying this distinction to the colour terms, it is clear that *rood* 'red' has an adjectival use, since it shows the characteristic alternation between declined and non-declined forms, taking the affix in (5), but not in the singular neuter indefinite NP in (4).

(5) om uit de *rode* cijfers te komen [wsuea-40.p4.s4]

In fact all of the basic colour terms of Berlin & Kay are not only used as nouns, but also as adjectives. To demonstrate this and –at the same time– to get some quantitative data about the use of the colour terms we sampled the SoNaR corpus, employing queries, such as those in (6).

(6) a. [word="rood" & pos="ADJ\(prenom.*"]
b. [word="rode" & pos="ADJ\(prenom.*"]

A summary of the results is given in Table 1. The upper half consists of the basic colours of Jaspers' diagram, which in turn are identical to the first six of Berlin & Kay. Interestingly, they are also the colour terms with the highest frequency in the corpus. The figures in the fourth column specify the percentage of non-declined forms.

For *oranje* 'orange' the table only mentions the sum of the uses, since it has the same form in singular neuter indefinite NPs, as in (7), and other NPs, such as

	nondeclined	#	%	declined	#	Sum
black	*zwart*	6377	13.02	*zwarte*	42591	48968
white	*wit*	4917	12.16	*witte*	35509	40426
red	*rood*	4215	7.34	*rode*	53230	57445
green	*groen*	4309	14.81	*groene*	24775	29084
yellow	*geel*	1852	6.90	*gele*	24971	26823
blue	*blauw*	2153	12.16	*blauwe*	15545	17698
brown	*bruin*	1420	18.41	*bruine*	6294	7714
grey	*grijs*	1313	13.83	*grijze*	8179	9492
purple	*paars*	449	8.00	*paarse*	5162	5611
pink	*roos*	112	1.80	*roze*	6112	6224
orange	*oranje*			*oranje*		2675

Table 1: Colour adjectives in prenominal position

the plural one in (8).

(7) … met een rood / *oranje* achterstuk … [wrpei-20972.p4.s206]

(8) … dat verpakt is in *oranje* doeken … [wsuea-232.p26.s11]

This lack of variation is due to shwa deletion. It applies to all adjectives whose stem ends in a shwa, such as *timide, frigide, morbide, …*, and it also affects other morpho-phonological processes, such as the formation of plural nouns, as in *bedienden, boden, …*, and comparative adjectives, as in *timider*.

A special case is *roze* 'pink'. Most speakers use the form *roze* both in singular neuter indefinite NPs, as in (9), and in other NPs, as in (10).[2]

(9) was het een *roze* of een blauw kussentje … [wrppg-196281.p13.s5]

(10) De *roze* laag behoort zichtbaar aanwezig te zijn in de stad [wrpph-46.p11.s1]

For these speakers it is similar to *oranje*. There are also speakers, though, who use the form *roos* in singular neuter indefinite NPs, as in (11).[3]

[2] Alongside *roze* one also finds the form *rose*, as in *de rose trui van de Giro d'Italia* [ws-u-e-a-0000000048.p27.s3]. It is dismissed as archaic in normative statements, see http://onzetaal.nl/taaladvies/roze-rose/.

[3] The 112 hits for prenominal *roos* include 36 instances of the singular non-neuter noun *roos*

(11) Die gaat vervolgens in een *roos* zakje … [wrppg-671991.p13.s3]

For these speakers it overtly shows the declension alternation.

By contrast, a non-basic colour term, such as *magenta*, does not show the declension alternation. In (12), for instance, we do not find the declined form which is standardly required in non-neuter nominals and which is indeed present on the colour term in the other conjunct, i.e. *blauwe* 'blue'.

(12) Verder zijn onder meer een blauwe en een *magenta* cowboy te zien. [wrpph-36501.p1.s2]

In this respect, *magenta* is just like the prenominal noun *aluminium* in *een kleine aluminium tafel* 'a small.DCL aluminium table'. The same holds for *lila, sepia, kaki, indigo*: They also lack the declension affix in plural and non-neuter nominals, as in (13)–(16).

(13) Marion (74), een oma gehuld in beige en *lila* tinten [wrppg-185373.p13.s1]

(14) … vol *sepia* foto's en prenten van Shanghai in de roaring twenties [wrppg-700881.p1.s1]

(15) een *kaki* tas voor bij m'n vest [wrpea-205450.p11.s1]

(16) de ronde maan aan de *indigo* hemel [wrppb-182.p679.s1]

Notice that the absence of the declension affix is not due to the fact that these colour terms end in a clear vowel, since adjectives which end in a clear vowel do take the affix, as in *naë verwanten, gedweeë schapen, cruë opmerkingen* and *moeë studenten*.

2.2 In nominal position

Characteristic of adjectives in nominal position is the presence of the *-e* affix, as in *een rijke* 'a rich.NOM' and *die dove* 'that deaf.NOM'. The affix is homophonous to the declension affix, but should be distinguished from it, not only because it has another function, but also because it has another range of application. Adjectives whose stem ends in *-en*, for instance, do not take the declension affix (*zijn eigen(*e) schuld* 'his own fault' and *een open(*e) deur* 'an open door'), but they can take the nominalization affix, as in *het eigene van deze behandeling*

(34 for the flower and 2 for the bull's eye) and 1 of the singular neuter mass noun *roos*, standing for the colour 'pink'. These are annotation errors.

'the particularity of this treatment' and *het opene van zijn karakter* 'the openness of his character'. Moreover, nominalized adjectives have a plural counterpart, while declined adjectives do not. It is marked by the affix *-n*, as in *de rijken* 'the rich.NOM.PL' and *die doven* 'those deaf.NOM.PL'. A characteristic property of this plural is that it can only refer to humans.

By contrast, nouns which appear in nominal positions do not take the *-e* affix, and if they have a plural counterpart it is canonically formed by the addition of *-s* or *-en*. Moreover, plural nouns do not necessarily refer to humans.

Colour terms in nominal positions are, hence, adjectives if they take the *-e* affix and if their plural counterpart uniquely refers to humans, as in (17)–(18).

(17) Als die Vlaamse *groene* er dan toch moet komen, dan ... [wrppi-124.p8.s5]

(18) De *groenen* zijn definitief gelanceerd. [wiki-5.p29.s21]

This contrasts with the noun *groen*, as used in (19).

(19) Verscholen achter het *groen* in de Betuwe ligt het dorpje Zoelmond. [wsuea-226.p8.s1] To retrieve the nominalized colour adjectives in the corpus we employed queries, such as those in (20).

(20) a. [word="groene" & pos="ADJ\(nom.*"]
b. [word="groenen" & pos="ADJ\(nom.*"]

The results for Berlin & Kay's basic colour terms are summarized in Table 2.

The absence of any instances for *gelen, paarsen* and *oranjen* is due to the fact that they are erroneously tagged as common nouns. The occurrences of *oranjen* and *gelen* in (21), for instance, do concern nominalised adjectives.

(21) De belangrijkste partij vandaag is niet die van de groenen, die van de blauwen, van de *oranjen*, de *gelen* of de licht- of donkerzwarten, maar wel die van de grijzen en kalenden. [wrppg-517257.p2.s2]

Notice that the plural form of the adjective *oranjen* differs from the plural form of the homophonous noun *oranjes*, as used in (22).

(22) VVD en D66 hebben geen behoefte aan een maatschappelijk debat over het persifleren van de *Oranjes*. [wrppg-32094.p2.s1]

The colour terms which were identified as invariably nominal in section 2.1, such as *magenta, lila, sepia, kaki, indigo*, behave as nouns in nominal positions: They do not take the nominalization affix, and since they are mass nouns, they have

		#	plural, human	#	Sum
black	*zwarte*	206	*zwarten*	2361	2567
white	*witte*	145	*witten*	5	150
red	*rode*	183	*roden*	15	198
green	*groene*	71	*groenen*	2986	3057
yellow	*gele*	73	*gelen*	0	73
blue	*blauwe*	331	*blauwen*	29	360
brown	*bruine*	34	*bruinen*	200	234
grey	*grijze*	127	*grijzen*	36	163
purple	*paarse*	9	*paarsen*	0	9
pink	*roze*	517	*rozen*	1	518
orange	*oranje*	218	*oranjen*	0	218

Table 2: Colour terms in nominal position

no plural counterpart, modulo coercion, in which case their plural form is marked by *-s*.

2.3 Comparative and superlative forms

In contrast to nouns, adjectives may take comparative and superlative affixes. The presence of those affixes in (23)–(25), hence, confirms that *groen, grijs* and *roze* have an adjectival counterpart.

(23) natuurlijk is de éne gemeente *groener* dan de andere [dpc-vla-1171-nl-sen.p98.s2]

(24) 22 jaar later is zijn haar een stuk *grijzer*, maar ... [dpc-ind-1648-nl-sen.p8.s2]

(25) ... in Cada Um door een *rozere* bril ... [wrppg-186151.p1.s4]

Similarly, the presence of the superlative affixes in (26)–(27) confirms that *groen* and *zwart* have an adjectival counterpart.

(26) En wist je dat Brussel de *groenste* hoofdstad van Europa is? [dpc-vla-1171-nl-sen.p61.s5]

(27) 1 juli blijft de *zwartste* dag uit de Britse geschiedenis [wrpph-28.p4.s6]

	comparative	#	superlative	#
black	*zwarter(e)*	204	*zwartst(e)*	397
white	*witter(e)*	511	*witst(e)*	47
red	*roder(e)*	322	*roodst(e)*	22
green	*groener(e)*	1040	*groenst(e)*	207
yellow	*geler(e)*	58	*geelst(e)*	2
blue	*blauwer(e)*	124	*blauwst(e)*	9
brown	*bruiner(e)*	173	*bruinst(e)*	12
grey	*grijzer(e)*	186	*grijst(e)*	5
purple	*paarser(e)*	14	*paarst(e)*	4
pink	*rozer(e)*	24		
orange	*oranjer(e)*	2		

Table 3: Comparative and superlative forms

To retrieve the comparative and superlative forms in the corpus we employed queries, such as those in (28).

(28) a. [word="zwartere?"]
b. [word="zwartste?"]

The results are summarized in Table 3.

There are no occurrences of *rozest(e)* or *oranjest(e)* in SoNaR, but there are combinations with *meest*, as in (29)–(30).

(29) de *meest roze* stad van de States, op de *meest roze* dag van het jaar [wrppg-244891.p5.s6]

(30) naar binnen, naar de *meest oranje* plek [wrppg-166951.p12.s2]

As expected, the colour terms which are only used as nouns, are not compatible with the affixes of the comparative and the superlative.

3 Conclusion

As demonstrated in section 2, the basic colour terms of Berlin & Kay are used as both nouns and adjectives. Conversely, the colour terms which are only used as nouns are typical examples of non-basic colour terms. An advantage of this test

is that it can also be applied to colour terms for which it is less clear whether they qualify as basic. As an example let us take *violet*. It is clearly monolexemic, but it is less clear whether its use is restricted to a narrow class of objects, or whether it is psychologically salient. The part of speech test, however, gives a clear result. Given that *violet* shows the declension alternation in prenominal position and that it is compatible with a comparative affix, it is clear that it is not only used as a noun, but also as an adjective.

(31) Ze zien alleen de kleuren in het ultra *violette* deel van het licht [wrpei-20972.p4.s220]

(32) Men moet er echter op bedacht zijn dat door beperkingen in de beeldschermtechniek de monitor het blauw te cyaan weergeeft : de correcte kleur is *violetter* [wrpej-14532.p3.s5]

It is, hence, a candidate for basic colour term status. Whether it *is* a basic colour term, depends on whether it also passes the other tests. Ideally these should be made as clear and easy to apply as the part of speech test and the monolexemicity test. We assume that this is feasible for the test of restrictedness to a narrow class of objects, for instance, by checking the range of nouns that are modified by colour adjectives in a corpus such as SoNaR. This enterprise, though, is left for future work. In the meantime, we are eagerly looking forward to Jaspers (ms).

References

Berlin, Brent & Paul Kay. 1969. *Basic color terms. their universality and evolution*. Berkeley & Los Angeles: University of California Press.

Jakubíček, Miloš, Adam Kilgarriff, Diana McCarthy & Pavel Rychlý. 2010. Fast syntactic searching in very large corpora for many languages. In *Paclic*. 741–747.

Jaspers, Dany. 2012. Logic and colour. *Logica Universalis* 6. 227–248.

Jaspers, Dany. ms. Logic of colours.

Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste & Ineke Schuurman. 2013. The construction of a 500-million-word reference corpus of contemporary written Dutch. In Peter Spyns & Jan Odijk (eds.), *Essential Speech and Language Technology for Dutch. Results by the STEVIN programme*. 219–247. Springer.

Reynaert, Martin, Matje van de Camp & Menno van Zaanen. 2014. OpenSoNaR:

user-driven development of the SoNaR corpus interfaces. In *Proceedings of COLING 2014: System demonstrations*. 124–128. Dublin, Ireland: Dublin City University and Association for Computational Linguistics.