Tilburg University

The Use of Colour in Reference Production

Sekicki, Mirjana; Viethen, Jette; Goudbeek, M.B.; Krahmer, E.J.

Published in:
Proceedings of the 36th Annual Conference of the Cognitive Science Society (2014)

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

The Use of Colour in Reference Production: A Comparison between Dutch and Greek

Mirjana Sekicki (mirjana_sekicki@yahoo.com)
Faculty of Philosophy; University of Novi Sad, Serbia

Jette Viethen (jette.viethen@mq.edu.au)
Centre for Language Technology; Macquarie University, Sydney, Australia

Martijn Goudbeek (m.b.goudbeek@uvt.nl)
Tilburg center for Cognition and Communication (TiCC); Tilburg University, The Netherlands

Emiel Krahmer (e.j.krahmer@uvt.nl)
Tilburg center for Cognition and Communication (TiCC); Tilburg University, The Netherlands

Abstract

The important role of referring expressions in human communication has inspired much research in the fields of computational linguistics and psycholinguistics. Building on the research done by Viethen, Goudbeek and Krahmer (CogSci, 2012) the present study takes a cross-linguistic perspective on examining the use of the colour attribute in distinguishing a target referent. It aims at answering the following research question: Does the availability of adequate basic colour terms in a language affect the use of colour in reference production? We conducted a language production experiment with native speakers of Dutch and Greek. Our results confirm that the use of the colour attribute in reference production depends on the colour term resources of a particular language. In addition, we have recorded a large cross-linguistic difference in the proportion of the colour use, which we relate to the particular colour nuances used.

Keywords: language production; reference production; colour; cross-linguistic study.

Introduction

Speakers often need to distinguish one object (the *target referent*) from other objects in the same scene that are not the intended referent (the *distractors*). Both spoken and written discourse often include noun phrases of the structure similar to the following: "the tall red bike at the corner". Such verbal descriptions produced in order to point to an object and put it in focus of the particular discourse are called *referring expressions*. When producing referring expressions a speaker needs to determine which attributes of a target (such as its colour, type, size or location) to include. This process, which is known as semantic content selection, has been a topic of extensive research in cognitive science.

Colour has been found to be readily used and even favoured in reference production tasks (Pechmann, 1989; Sedivy, 2003). It is frequently included redundantly in a target description (Koolen, Goudbeek & Krahmer, 2013; Viethen & Dale, 2011). This is claimed to be due to the ease of colour perception, since it is perceived instantly and independent of context (Pechmann, 1989). In contrast, an object's size is a relative attribute and its value can be determined only in comparison to the objects of the same type (Brown-Schmidt & Tanenhaus, 2006). Thus, the use of size in referring expression production has been shown to depend on how distinguishing size is of the target object (Sedivy, 2003; Viethen & Dale, 2011). The findings of Pechmann (1989) show that people tend to start articulating a referring expression before having thoroughly examined the whole scene. Hence, the adjectives that denote easily perceivable features tend to be articulated first. In his study, Pechmann (1989) found that such a feature was almost exclusively colour. People often produce overspecified referring expressions, since they tend to start uttering their description with the most easily perceivable characteristic and only later include the most relevant ones. Moreover, the results of Belke and Meyer (2002) are in agreement with Pechmann (1989) showing that speakers usually mention the absolute characteristics first, without reflecting upon how distinguishing they are of the target.

Viethen, Goudbeek and Krahmer (2012) were the first to further investigate the mentioned tendency for using colour. They examined whether the preference for using the colour attribute is reduced when the colour of the target is relatively similar to that of the distractors. In addition, they considered the notion of colour term basicness.

Berlin and Kay (1969) argued for eleven basic perceptual colour categories universal to human vision that act as referents for eleven or fewer *Basic Colour Terms* (BCTs) in any language. In order to be considered basic a colour term has to conform to various parameters.[1] The authors found

[1]Berlin and Kay (1969) set the following criteria for considering a colour term basic: (a) the colour term should be monolexemic – its meaning is not predictable from the meaning of its parts; (b) its meaning should not be included in the meaning of any other term; (c) its use should not be restricted to a certain class of objects; (d) it should be psychologically salient for speakers, which implies the tendency to occur at the beginning of elicited lists of colour terms, the stability of reference across informants and occasions of use, and the occurrence in the idiolects of all informants.

that the number of BCTs can vary greatly among languages, i.e. a language can have between two and eleven BCTs.

Interestingly, the results of Viethen et al. (2012) show that colour use does not differ significantly between the conditions with colours that are perceptually highly different (e.g., red and blue) and with colours of low perceptual difference but high codability, i.e., for which BCTs are available (e.g., red and pink); but they did record a reduction in colour use for colour pairs of low codability, i.e., nameable only with a morphologically complex modification of a BCT (e.g., light-blue and dark-blue). Thus, they conclude that the difference in colour use between conditions is entirely due to the difficulty in naming similar colour nuances. However, their findings do not clearly point as to what causes this difference, namely, whether it is the lack of separate BCTs or morphological complexity. Using morphologically complex terms potentially asks for higher cognitive effort. However, in order to require less cognitive effort a mono-morphemic term should also be salient.

In the present study, we take a cross-linguistic perspective and examine what influence the colour vocabulary resources of different languages have on the assumed preference for using the colour attribute in reference production. We follow Viethen et al. (2012) but include two languages that differ in the number of BCTs and thus allow for stronger evidence for the impact of colour codability on the use of colour. The language of the experiment conducted by Viethen et al. (2012) was Dutch. In order to produce comparable results, the present study employs Dutch, but with the addition of Greek, a language that according to Androulaki et al. (2006) includes a 12th BCT, *ghalazio* (γαλάζιο: light-blue), and thus has the maximum number of BCTs found in a language. These two languages enable us to create stimuli with colour nuances for which separate BCTs are available in one language but not in the other and thus record if colour codability yields any differences in colour attribute use both within a language and cross-linguistically. In addition, we might be able to shed more light on the apparent inconsistency of the claim that Greek has 12 BCTs with the theory of Berlin and Kay (1969).

In sum, we compare the use of the colour attribute in three different settings: (a) when the colours of the target and the distractors are very different, (b) when these colours are similar with available BCTs for naming them, and (c) when the colours are similar with no available BCTs. In addition to within-language comparisons, this set-up allows us to examine the potential cross-linguistic differences.

We aim at answering the following research question: *Does the availability of adequate BCTs in a language affect the use of colour in reference production?* Following Viethen et al. (2012) we expect no difference in the use of the colour attribute in conditions where the colours of the target and of the distractors are very different in comparison with the conditions where the colours are similar. However, we expect this to be so only in the conditions where BCTs exist for both the colour of the target and the distractors. Thus, we postulate the following hypotheses:

H1: *Low perceptual difference between the colours of the target and those of the distractors alone does not moderate the proportion of colour use in reference production.*

H2: *Low codability of the colours required to distinguish the target from the distractors leads to a reduction in the proportion of colour use in reference production.*

One of the practical implications of the present study is the further development of Referring Expression Generation (REG) algorithms. None of the existing algorithms considers what modifies the human readiness to use colour. However, aiming at producing human-like output, REG algorithms should take into account the findings that show whether the perception of some type or combination of colour nuances leads to (dis)preferring the use of colour in reference production and whether these preferences differ for languages with different BCT inventories.

Experiment

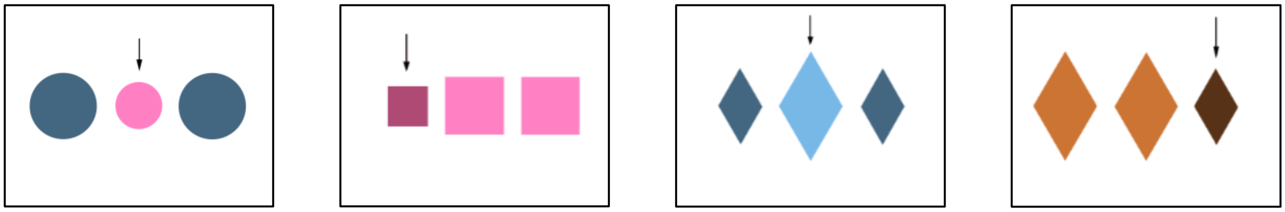
We conducted a language production experiment presenting the participants with a number of images consisting of three geometrical objects, displayed on a computer screen. The participants were instructed to verbally refer to one particular object in the scene so that an imaginary partner could successfully distinguish that object. The same experiment was conducted both with Greek and Dutch participants, in their respective languages.

Method

Participants 35 native Greek speakers, all but two students of Tilburg University, participated in the experiment voluntarily. They had spent a maximum of one year in the Netherlands and none of them spoke Dutch. There were 19 male and 16 female participants. Their age ranged from 19 to 30 years (M = 24.63). In addition, 30 native Dutch speakers, students of Tilburg University, participated either voluntarily or in return for course credit. There were 13 male and 17 female participants. Their age ranged from 18 to 29 years (M = 21.83).

Materials and Design To ensure comparability, the experimental design and the stimuli employed largely resemble those of Viethen et al. (2012). Each participant was presented with 40 critical images and 80 filler items. The critical trials consisted of three simple two-dimensional geometrical figures of the same type. The target figure differed in colour and size from the distractors and the two distractor figures were identical. Thus, either the use of colour or size sufficed to fully distinguish the target. The area of the target figure was always two times bigger (or smaller) than the area of the distractor figure.

The main experimental manipulation concerned the colour of the figures. In a pilot study, native speakers of both Greek



(b) *hidiff*: one small pink and two big dark-blue circles.

(c) *lodiff-hicode*: one small purple and two big pink squares.

(d) *lodiff-dep_code*: one big light-blue and two small dark-blue diamonds.

(e) *lodiff-locode*: one small dark-brown and two big light-brown diamonds.

Figure 1: Example stimuli from the four experimental conditions

and Dutch participated in a colour naming pre-test which ensured they named the chosen colours as anticipated.

Four different colour conditions were created and named in resemblance to the conditions used by Viethen et al. (2012). The *hidiff* condition included images with colours that differed considerably, namely, pink and dark-blue, and purple and light-blue. Half of the images included objects in purple and light-blue and another half in pink and dark-blue. Colours of a pair differ greatly not only in terms of hue and saturation, but also in terms of their brightness, as pink was never used in combination with light-blue and purple never with dark-blue. Figure 1(a) illustrates the *hidiff* condition.

The *lodiff-hicode* condition used colours that differ only in brightness, namely, pink and purple as shown in Figure 1(b). Thus, this condition used colours of lower difference but high codability since BCTs exist for them in both languages.

Figure 1(c) illustrates the *lodiff-dep_code* condition. Colours differ only in terms of brightness, as in *lodiff-hicode*; however, different coding options are available for them in the two languages. There are no separate BCTs for light-blue and dark-blue in Dutch, but the BCT *blauw* (blue) has to be modified with *licht* (light) or *donker* (dark) in order for the distinction to be made. Greek, on the other hand, has two BCTs, *ghalazio* (γαλάζιο) and *ble* (μπλε), respectively.[2] For the Greek participants, this condition is equivalent to *lodiff-hicode*, for the Dutch, to *lodiff-locode*.

Finally, Figure 1(d) illustrates the *lodiff-locode* condition that included light-brown and dark-brown, colours for which both languages use the morphologically complex structure, *lichtbruin* (light-brown) and *donkerbruin* (dark- brown) in Dutch; and anihto kafe (ανοιχτό καφέ: light-brown) and skuro kafe (σκούρο καφέ: dark-brown) in Greek.

In order to set the colour values, we used the Hue Saturation Brightness (HSB) colour model. Since there were three pairs of colours (pink/purple, light-blue/dark-blue, light-brown/dark-brown), we had to make sure that the difference between the colours in each pair was identical. Tightening up the colour choice strategy of Viethen et al. (2012), we decided that the only difference between the two colours in a pair was to be a 40% difference in brightness, keeping the values of hue and saturation constant. Table 1 shows the HSB values for the colour pairs used. Note that *hidiff is* the only condition where the colours of the target and distractor objects differ in all three values, not only brightness. This ensured a high and more easily perceivable difference between the colours than in the other conditions.

The filler items were designed to make sure that the overall number of occurrence of each colour was kept approximately the same. The number of target objects in each colour was kept balanced, as well as the position of the target object in the scene, namely, left, middle or right.

Filler Items Two thirds of the total number of trials were filler items, half of which were geometrical fillers and the other half Greeble fillers.

Geometrical fillers were similar to the critical items in that they used the same types of geometrical figures and the same colours. However, they differed in various ways in order not to prime the participants to develop strategies in creating their descriptions. First, the target object could be distinguished from the distractors in terms of pattern (vertical stripes or dots) and/or type. Colour and size were never fully distinguishing. Second, the distractors were never identical, but differed in colour, type and/or pattern. Third, the size of the figures in a scene was the same.

The Greeble items[3] consisted of two 3D purple objects each. These objects are quite complex and difficult to differentiate and differ greatly from the critical items and geometrical fillers. As has been the case in previous studies (e.g., Koolen et al., 2013), they have proved to be an excellent distractor leading the majority of participants in our experiment to consider them the objects of our attention.

Procedure The stimuli were shown on a computer screen in a silent and dimly lit room. Half of the participants were presented with the randomized order of stimuli and the other half with the reversed version of this order. In both cases, a geometrical filler was always followed by a Greeble filler, which was followed by a critical item.

[2]It should be noted that speakers of Greek have the alternative of saying *anihto ble* (ανοιχτό μπλε: light-blue), instead of *ghalazio* (γαλάζιο) for this colour nuance.

[3]The Greebles are courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University; www.tarrlab.org.

Table 1: The table of the colour values

C	pink	purple	l-blue	d-blue	l-brown	d-brown
H	330°	330°	210°	210°	30°	30°
S	50%	50%	40%	40%	70%	70%
B	100%	60%	90%	50%	70%	30%

The experiment included written instructions that requested the participants to imagine a situation in which they were to assist a colleague who was presented with the same set of images. This imaginary colleague was supposed to click on the target object in each of the images, but they did not know which object to click on. An arrow was used to point out the intended target object to the participants, whose task it was to verbally distinguish the target from the distractors. Each image included the beginning of an appropriate sentence, namely, *Click on the…* in the respective language. The sentence was included in order to remind the participants of their task and to lead them to be more concrete in their answers, not needing to introduce the context for each image. The participants were instructed not to use spatial information of the target object, i.e., not to characterize it as the left, right or the middle one in the scene. This restriction assured that the participants would use only the object characteristics controlled by the researcher. Moreover, the time for producing the answers was limited. Each image was displayed for only 4.5 seconds. After that a fixation cross was displayed for 1.5 seconds. This was done in order to prevent the participants from producing extensive descriptions of the scenes, as well as from meditating upon their answers.

After the experiment a post-test was conducted where the participants were presented with the colour pairs used in the experiment, and were requested to name the colours. The post-test was introduced in order to confirm that the right nuances of colours were used in order to elicit the expected colour terms in both languages.

Table 2 gives the proportion of the colour attribute use in the experimental conditions in the two languages.

Results

Coding of the Independent Variables The main dependent measure we analysed is the proportion of colour use in the different conditions. We consider a description to contain colour if the term used is true of the target object, regardless of its distinguishing value. Moreover, we analysed the use of the size attribute.

The recorded responses were first transcribed and then annotated by a Dutch native-speaker and a near-native speaker of Greek, for Dutch and Greek data, respectively.

Data Analysis We used a within-subjects ANOVA to compare the use of attributes in the four experimental conditions. Subsequently, pairwise comparisons further investigated the differences. In addition, the differences between conditions in the two languages were assessed with a mixed 2 (languages) x 4 (conditions) ANOVA, again, followed by pairwise comparisons.

The use of the colour attribute in the Greek sample showed a significant overall effect [F (3,102) = 4.09, p = .009]. After six pairwise comparisons with Bonferonni's correction for α-level mistakes (p < .008), the only significant difference in colour use [t (34) = 3.22, p < .008, r = .48] was recorded between the *hidiff* condition (M = .81, SD = 0.33) and *lodiff-dep_code* (M = .71, SD = 0.40).

For the use of size in the Greek sample, the within-subjects ANOVA revealed a significant overall effect [F (3,102) = 4.84, p < .003]. Pairwise comparisons showed a significant increase in the use of size [t (34) = -3.17, p < .008, r = .48] between the *hidiff* condition (M = .79, SD = 0.25) and *lodiff-dep_code* (M = .89, SD = 0.18). No significant difference was found between *hidiff* (M = .79, SD = 0.25) and *lodiff-hicode* (M = .89, SD = 0.18), [p = .012], however, with the strict α-level correction we employed this can be seen as a trend towards significance.

The ANOVA revealed a significant difference in colour use among the conditions in the Dutch sample [F (3,87) = 5.38, p = .002]. Consequently, pairwise comparisons showed that the difference between *hidiff* (M = .51, SD = 0.44) and *lodiff-locode* (M = .40, SD = 0.48), exhibits a trend towards significance [t (29) = 2.80, p = .009]. Also, a trend towards significance was found between *lodiff-hicode* (M = .45, SD = 0.47) and *lodiff-locode* (M = .40, SD = 0.48) [t (29) = 2.80, p = .009], and between *lodiff-dep_code* (M = .47, SD = 0.47) and *lodiff-locode* (M = .40, SD = 0.48) [t (29) = 2.73, p = .011]. Concerning the use of size in the Dutch sample, we found a significant overall effect [F (3,87) = 4.59, p < .005] and a significant increase in the use of this attribute [t (29) = -3.25, p < .008, r = .52] between *hidiff* (M = .85, SD = 0.21) and *lodiff-hicode* (M = .95, SD = 0.12).

Concerning the comparison between the two languages, the mixed ANOVA revealed a significant main effect of colour use among the conditions [F (3,189) = 6.79, p < .001] and a significantly higher use in the Greek sample [F (1,63) = 8.87, p= .004, r = .35]. There was no significant interaction between language and condition [F (3,189) = 2.55, p = .06]. Figure 2 shows the difference in the overall colour use between the two language samples.

Table 2: The proportion of colour use in the conditions. The shaded fields present the conditions where no difference in colour use was expected.

	Greek		Dutch	
Conditions	Mean	SD	Mean	SD
hidiff	.81	0.33	.51	0.44
lodiff-hicode	.76	0.36	.45	0.47
lodiff-dep_code	.71	0.40	.47	0.47
lodiff-locode	.75	0.40	.40	0.48

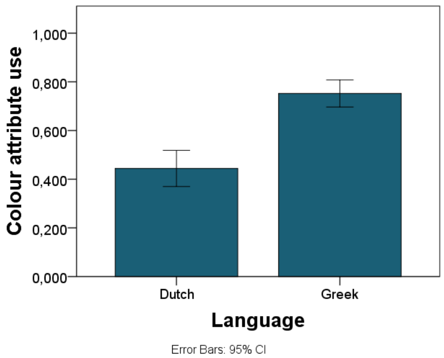


Figure 2: The use of the colour attribute in the two language samples, Dutch and Greek

There was a significant effect of size use among the conditions [F (3,189) = 8.92, p< .001], but no difference between languages [F (1, 63) = 1.595, p = .21], and no significant interaction [F (3,189) = 0.34, p = .79].

Finally, we have calculated the proportion of use of the anticipated colour term, out of the cases in which that term could have been used and a colour term was used. As shown in Table 3, the anticipated terms were less readily used in the experiment than in the colour naming post-test in both languages, even though participants were in general able to correctly identify the colours. Interestingly, this reduction in colour use compared to the colour naming test is lower, or non-existent for pink and purple.

Discussion

The results support the first hypothesis since there was no significant difference in colour use between the *hidiff* and the *lodiff-hicode* condition in either of the languages. Even though the colours used in *lodiff-hicode* were more similar than those in *hidiff*, this did not lead to a significant difference in colour use between the two conditions, since the colours in both conditions could be named using BCTs. This is in agreement with the results of Viethen et al. (2012).

For the Greek speakers, *lodiff-dep_code* was similar to *lodiff-hicode* since the differences between colours in the pairs were the same and separate BCTs exist to name them. Consequently, we expected no reduction in colour use between these conditions in the Greek sample; however, the results proved the contrary. The colour naming post-test showed that colours were successfully recognized; but there was a reduced use of these terms in actual reference production. This suggests a potential difference between using an adequate term in a colour naming task and in reference production. In addition, perhaps *ghalazio* (γαλάζιο) is not as established as a BCT as Androulaki et al. (2006) suggested.

The second hypothesis expected a reduction in colour use compared to *hidiff* in *lodiff-dep_code* and *lodiff-locode* in the Dutch, and in *lodiff-locode* in the Greek sample. Since only a tendency towards significance was recorded for *lodiff-locode* in the Dutch sample, this hypothesis is not supported by the present results. For the Greek sample, this interesting finding may have to do with the availability of multiple colour terms in this condition since it provides an option for using a BCT (*ghalazio* (γαλάζιο)) or a modified BCT (*anihto ble* (ανοιχτό μπλε). In addition, the colours used were not prototypical, but were highly constrained by the rules set in order to strictly control the stimuli. Thus, unclear colour nuances paired with multiple naming options may have led to uncertainty and confusion as to how to name them and consequently to the reduction in colour use.

Considering the between languages comparison, there was a large main effect of language on colour use. Greek participants used colour more readily in all experimental conditions. However, there was no significant difference in size use between the languages. Thus, the Greek sample included more overspecified references. There is a much greater readiness for using size in the present study than in Viethen et al. (2012). This may also be due to the fact that our colour nuances were less prototypical for their expected terms, which potentially led to reduced colour use in the Dutch and a high size use in both samples. In addition, one possible explanation for the higher colour use in Greek is tuning of the colours to the Greek speakers, i.e., the nuance for light-blue was designed to elicit the expected Greek colour term that does not exist in Dutch. Moreover, Greek speakers make frequent use of morphologically simple, yet non-basic terms. Thus, an effort was made not to elicit unwanted terms from the plethora of readily used ones in Greek, e.g., we had to create a nuance of *pink* that would not be called *fuchsia*. Tuning the colours to one particular cultural and language group might have led to these colours being found non-typical and difficult to name to speakers of the other background, leading to the observed low rate of colour use in the Dutch sample.

Another possible explanation for the unexpectedly low colour use in Dutch is that the experimental design ensured that there were an equal number of occurrences of every colour. Since most of the colours (four out of six) required a morphologically complex term in Dutch and were therefore less easily nameable, this might have led the participants to abandon colour use altogether.

Table 3: The proportions of use of the *anticipated* colour terms: (I) in the experiment, in instances where a colour term was used, (II) in the colour naming post-test

	I	Dutch	Greek	II	Dutch	Greek
pink		100	98		93	100
purple		97	90		97	91
light-blue		50	66		97	80
dark-blue		61	88		97	97
light-brown		71	81		80	86
dark-brown		77	73		97	97

Thus, the present study's use of non-prototypical colour nuances has potentially led to the scarce colour use by the Dutch and the extensive size use by the Greek participants. However, prototypical basic colours are not readily found in nature. Colour naming options are much wider in reality. The results of the present study suggest that this fact is not to be neglected. Consequently, future studies should perhaps examine to what extent the use of dull colour nuances moderates the use of the colour attribute.

Finally, the colour nuances used might have led to colour becoming a less salient attribute. However, since the experimental setting gave the participants the freedom to choose any linguistic means to describe the target object, there is a possibility that colour use was influenced more by the frequency of colour terms in the respective languages than by any perceptual effects.

Consequences for Computational Modelling The Incremental Algorithm (Dale & Reiter, 1995) is considered most successful in producing human-like referring expressions. It is based on a preference list over attributes where all the values of an attribute are treated the same, e.g., all colour nuances are equally likely to be used in a referring expression. Our results suggest that different preference ratings should exist for the different colour values. For instance, colour may be the preferred attribute in the context of prototypical colours; however, size may become preferred over colour in the context of non-prototypical values. In addition, we have shown that tuning colours to members of one language and cultural group might cause difficulties in naming colours for members of a different group. This implies that rankings in the preference list should be different for different languages.

Conclusions

This study was set to investigate whether it is morphological complexity or colour term basicness that led to the findings of Viethen et al. (2012). The present results show, first and foremost, that there are more factors that influence people's tendency to use colour. Assuming that *ghalazio* is actually not an established BCT in Greek could account for the reduction in colour use in the Greek *lodiff-dep_code* condition, however, it does not explain the lack of reduction in *lodiff-locode*. Since we recorded no significant reduction in *lodiff-locode* in either language, morphological complexity and lack of adequate BCTs alone proved not to be sufficient to cause the reduction in colour attribute use.

The most important effect we found is the difference in colour use between the two languages. The colour nuances used in the experiment were finely tuned to Greek speakers with special attention to using the right nuances to elicit the expected colour terms. Subsequently, Greek speakers used colour extensively. Dutch speakers used colour to a much lower degree, even less than in Viethen et al. (2012). Hence, using dull, non-prototypical colour nuances might lead to the reduction in colour use with speakers of a language where those nuances are not linguistically specified.

Future work in the field of colour attribute use should consider the choice of colour nuances. The improvement of REG algorithms should include the context and source of the colour of the target, whether the target is an object found in nature and to what extent its colour can be regarded as prototypical of a colour term from a given language.

Acknowledgments

The reported research forms part of the VICI project "Bridging the Gap between Psycholinguistics and Computational Linguistics: the Case of Referring Expressions", funded by the Netherlands Organization for Scientific Research (NWO grant 277-70-007). We thank Kristel Bartels for help with transcribing the Dutch data.

References

Androulaki, A., Gômez-Pestaña, N., Mitsakis, Ch., Jover, J., Coventry, K. & Davies, I. (2006). Basic Colour Terms in Modern Greek. *Journal of Greek Linguistics*, *7*, 3-47.

Belke, E., & Meyer, A. S. (2002). Tracking the Time Course of Multidimensional Stimulus Discrimination: Analyses of Viewing Patterns and Processing Times During "Same"–"Different" Decisions. *European Journal of Cognitive Psychology*, *14*(2), 237-266.

Berlin, B., & Kay, P. (1969). *Basic Colour Terms: Their Universality and Evolution*. Berkeley and Los Angeles, Calif.: University of California Press.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the Eyes When Talking about Size: An Investigation of Message Formulation and Utterance Planning. *Journal of Memory and Language*, *54*, 592-609.

Dale, R., & Reiter, E. (1995). Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, *18*, 233-263.

Koolen, R., Goudbeek, M. & Krahmer, E. (2013). The Effect of Scene Variation on the Redundant Use of Colour in Definite Reference. *Cognitive Science, 37(2)*, 395-411.

Pechmann, T. (1989). Incremental Speech Production and Referential Overspecification. *Linguistics, 27(1)*, 89–110.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32(1)*, 3–23.

Viethen, J., & Dale, R. (2011). GRE3D7: A Corpus of Distinguishing Descriptions for Objects in Visual Scenes. In *Proceedings of the Workshop on Using Corpora in Natural Language Generation (NLG): Language Generation and Evaluation (UCNLG+Eval)*, 12-22, Edinburgh, UK.

Viethen, J., Goudbeek, M. & Krahmer, E. (2012). The Impact of Colour Difference and Colour Codability on Reference Production. *Proceedings of the 34th annual meeting of the Cognitive Science Society (CogSci 2012),* 1084-1089, Sapporo, Japan.