The role of graduality for referring expression generation in visual scenes

Albert Gatt[1], Nicolás Marín[2],
François Portet[3], and Daniel Sánchez[2]

[1] Institute of Linguistics, University of Malta
[2] Department of Computer Science and A.I.,
University of Granada, 18071, Granada, Spain
[3] Laboratoire d'Informatique de Grenoble,
Grenoble Institute of Technology
albert.gatt@um.edu.mt,nicm@decsai.ugr.es,
francois.portet@imag.fr,daniel@decsai.ugr.es

Abstract. Referring Expression Generation (REG) algorithms, a core component of systems that generate text from non-linguistic data, seek to identify domain objects using natural language descriptions. While REG has often been applied to visual domains, very few approaches deal with the problem of fuzziness and gradation. This paper discusses these problems and how they can be accommodated to achieve a more realistic view of the task of referring to objects in visual scenes.

Keywords: Referring expression, fuzziness, linguistic description, visual scenes

1 Introduction

The aim of systems for automatically generating linguistic descriptions of visual scenes is to generate text describing the scene contained in an image, simulating the results provided by humans in the same task. In recent years, scene description has attracted a lot of attention as part of a growing trend to explore the interface between vision and language in Natural Language Generation (NLG) [1], as is the case of other kind of data [2–4].

Research in this area tends to focus on learning correspondences between parts of an image and textual descriptions to generate descriptive captions [5–7]. A somewhat different focus is offered by work on Referring Expression Generation (REG), one of the most important tasks in NLG. A referring expression is a noun phrase whose communicative purpose is to identify an object to the hearer, for which it must be a distinguishing description, i.e., it must be "an accurate description of the entity being referred to, but not of any other object in the current context set", the context set being "the set of entities that the hearer is currently assumed to be attending to" [8]. The most standard way of approaching REG computationally is by seeing the problem as a content selection problem [9]. If a particular object is to be identified linguistically among a set of

other objects, then the challenge is to find a set of properties of the object that identifies it uniquely and to do so in a way that matches what human speakers would do in a similar situation.

Formally, this means that there are two important criteria for a REG algorithm to be successful. Let $re = \{p_1, \ldots, p_n\}$ stand for a referring expression (a set of properties) returned by such an algorithm for some target referent o. The description re is considered *accurate* if, for all $p \in re$, $o \in [\![\, p \,]\!]$, that is, every property in the description is true of o. Furthermore, re is a *successful* description if it uniquely identifies o, that is:

$$\bigcap_{p \in re} [\![\, p \,]\!] = \{o\} \tag{1}$$

The choice problem represented by REG has been explored in a number of algorithms reviewed in detail in [9], though it has only recently been tackled in the context of complex visual scenes [10, 11]. Referring expressions in images use visual properties and concepts for characterizing univocally certain regions in the image, corresponding to objects or areas of interest.

In many cases these concepts and properties are fuzzy or gradable in nature, a factor that has been ignored in many REG algorithms, which typically treat a target's properties as symbols with a crisp extension. While exceptions to this trend exist [12], they have focussed on prototypically gradable properties, such as an object's size. Yet, even properties such as an object's colour admit of boundary cases and gradations. Another simplification in existing REG models has to do with the notion of an object itself: frequently, REG algorithms assume that the relevant components of a visual scene are entities and their properties. However, even the concept of an 'entity' can be argued to be fuzzy. An example where this complication becomes an important issue is a weather forecasting application, which needs to refer to a geographical region in order to describe the weather conditions holding within it. As real-world applications show [13], regions corresponding to a particular set of predicates describing the weather conditions will often not be easily identifiable using place names or simple expressions based on cardinal directions.

Thus, in addition to the problem of modelling the semantics of fuzzy concepts with respect to their correspondence with properties of objects [14, 15], referring expressions using fuzzy concepts may also match objects in the scene to a certain extent, and this extent has to be calculated, introducing an additional complexity in the REG problem. In particular, distinguishability of an object with respect to others in terms of a referring expression involving fuzzy concepts turns out to be a matter of degree. In this paper we address this problem.

The paper is organized as follows: in Section 2 we briefly discuss fuzzy visual concepts. In Section 3 we deal with graduality in referring expressions. Implications of graduality in REG are discussed in Section 4. Section 5 concludes the paper.

2   Fuzzy concepts in visual scenes

Concepts related to low-level features like colour, texture, and shape are among the most employed in linguistic description [16], and are usually affected by fuzziness. Colour terms used as adjectives like *red*, *vivid yellow*, etc. are paradigmatic examples of fuzzy concepts, and can be represented by means of fuzzy subsets of crisp colours in some colour space. That is, each crisp colour (usually represented by a triplet of numbers, whose domain and meaning vary from one colour space to another) has a membership degree to each colour term. Several proposals are available in the literature for determining the membership function representing each colour term [17]. Once the terms are defined, they have to be matched to the objects under study. This is not a trivial issue because, unless all the pixels comprising the object have the same colour, approaches for associating colour terms to the whole object are necessary, like for instance the notion of *colour dominance* [18]. Fuzzy approaches for defining shape (round, regular, convex, etc.) and texture properties (coarseness, contrast, orientation, etc.) are also available, see [19, 20] and references therein.

Not only properties, but regions themselves can be fuzzy as well, as they are usually defined in terms of their correspondence to properties [16]. In this sense, a fuzzy region is defined as a fuzzy subset of pixels in the image. As an example of this, if we take the homogeneity in colour as the criterion for determining regions, we obtain fuzzy regions like the ones in the image in Figure 1(a), which shows a collection of cells as seen under the microscope. The regions corresponding to cells are characterized by being *red* but, whilst some of them have clear boundaries, some others do not, since the colour of some pixels has partial membership to *red* for some cells. Dozens of methods for fuzzy segmentation are available [21].

The lack of clear boundaries of fuzzy regions is another source of fuzziness for other properties which further complicates the task of calculating a region's other properties, including its size, shape, and its spatial relationships with other regions and with the image framework (location), among others. These concepts, which are themselves fuzzy even for crisp regions, have been an object of study with fuzzy techniques [22, 16, 23]. It is easy to see examples of concepts that are fulfilled only to some degree by objects in the image depicted in Figure 1(a), such as *round shape*, *above* (spatial relationship), *in the middle* (location), etc.

In order to illustrate our discussion, let us consider, for the objects in Figure 1(a), the fuzzy properties *small*, *dark*, and *upperleft*. Table 1 shows example fulfilment degrees of these properties for all the named objects in Figure 1(a). For the property *small*, a fuzzy subset on the domain of the proportion of pixels in the image has been employed. The property *dark* is calculated as the dominance of the property *low luminance* of the pixels in an object. The property *upperleft* is defined as a fuzzy region with square core and support, core with side 1/4 of the image, and support with side 3/8, with membership decreasing linearly. We omit more specific details about how these degrees have been calculated for reasons of space.
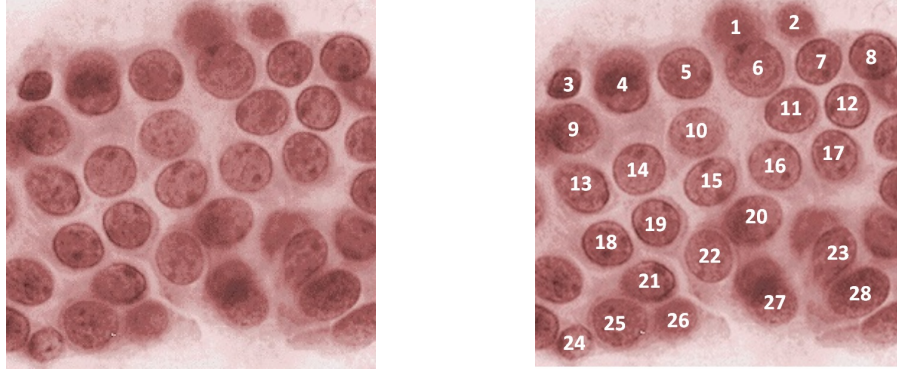
(a) An image containing objects with fuzzy properties

(b) Labelled cells in the image

Fig. 1. An image with fuzzy properties and its labelled counterpart

3   Gradation in referring expressions

When referring expressions are comprised of fuzzy properties, both their accuracy and their ability to discriminate an object become a matter of degree, contrary to the view usually assumed by REG algorithms (see Section 1). In this section we propose different ways to assess these, as well as some other properties which are also important when generating referring expressions and assessing their quality.

3.1   Accuracy

As we have noted, the accuracy of a referring expression with respect to a given object measures the extent to which what the expression says is true for the object, that is, the degree to which *all* the properties that appear in the expression hold for the object. When the referring expression is composed of fuzzy properties, to calculate the accuracy of the expression for a given object we first have to compute the fulfilment degree of each property; subsequently we have to appropriately aggregate these degrees into a single value.

Let us once again consider a referring expression re involving properties $\{p_1, \ldots, p_n\}$ and an object o. Let also $p_i(o)$ be the fulfilment degree of property p_i for object o. The accuracy of re for object o is calculated as follows:

$$a_{re}(o) = \bigotimes_{i=1}^{n} p_i(o) \tag{2}$$

where $\otimes$ is a t-norm.

As an example, consider again Figure 1(a) (Figure 1(b) assigns numbers to the cells) and the three referring expressions: *The small object in the upper-left* (re_1), *The dark object in the upper-left* (re_2), and *The small and dark object* (re_3).

Cell	Dark	Small	Upperleft	dark & small	dark & upperleft	small & upperleft
1	0,00	0,29	0	0,00	0,00	0,00
2	0,00	0,71	0	0,00	0,00	0,00
3	1,00	1,00	0,96	1,00	0,96	0,96
4	0,83	0,00	0,85	0,00	0,83	0,00
5	0,00	0,00	0,05	0,00	0,00	0,00
6	0,00	0,00	0	0,00	0,00	0,00
7	0,00	0,14	0	0,00	0,00	0,00
8	0,33	0,14	0	0,14	0,00	0,00
9	0,33	0,00	0,33	0,00	0,33	0,00
10	0,00	0,00	0,01	0,00	0,00	0,00
11	0,00	0,14	0	0,00	0,00	0,00
12	0,00	0,43	0	0,00	0,00	0,00
13	0,00	0,00	0	0,00	0,00	0,00
14	0,00	0,00	0	0,00	0,00	0,00
15	0,00	0,00	0	0,00	0,00	0,00
16	0,00	0,00	0	0,00	0,00	0,00
17	0,00	0,14	0	0,00	0,00	0,00
18	0,17	0,14	0	0,14	0,00	0,00
19	0,17	0,14	0	0,14	0,00	0,00
20	0,00	0,00	0	0,00	0,00	0,00
21	0,17	0,43	0	0,17	0,00	0,00
22	0,00	0,00	0	0,00	0,00	0,00
23	0,17	0,29	0	0,17	0,00	0,00
24	0,00	0,86	0	0,00	0,00	0,00
25	0,00	0,00	0	0,00	0,00	0,00
26	0,00	0,57	0	0,00	0,00	0,00
27	0,50	0,43	0	0,43	0,00	0,00
28	0,67	0,00	0	0,00	0,00	0,00

Table 1. Fulfilment degree of several fuzzy properties and accuracy of three referring expressions for the image in Figure 1(a).

According to data in table 1, and using the minimum as t-norm, we have that:

- $a_{c_3}(re_1) = \min(small(c_3), upperleft(c_3)) = 0.96$
- $a_{c_3}(re_2) = \min(dark(c_3), upperleft(c_3)) = 0.96$
- $a_{c_3}(re_3) = \min(small(c_3), dark(c_3)) = 1$

3.2   Referential success

As noted in Section 1, referential success is the extent to which the referring expression unambiguously identifies the intended referent to the reader or hearer. Consider again a referring expression re and a set of objects $O = \{o_1, \ldots, o_m\}$. The referential success of re with respect to a referent object o_i is the degree to which re is true for o_i and false for the rest of objects, that is:

$$rs_{re}(o_i) = a_{re}(o_i) \otimes \left(\bigotimes_{o_j \in O \wedge j \neq i} \neg(a_{re}(o_j)) \right) \in [0, 1] \qquad (3)$$

where $\otimes$ is a t-norm and $\neg$ is a fuzzy negation. Note that this notion of referential success is quite different from the more traditional definition given

in REG (see 1 above). As an example, consider again Figure 1(a) and the three referring expressions in the previous section. In this case, if we once more use the minimum as t-norm and the standard fuzzy negation operator, we have that:

– $rs_{re_1}(c_3) = \min(a_{re_1}(c_3), \min_{c_j \neq c_3}(1 - a_{re_1}(c_j))) = \min(0.96, 1) = 0.96$
– $rs_{re_2}(c_3) = \min(a_{re_2}(c_3), \min_{c_j \neq c_3}(1 - a_{re_2}(c_j))) = \min(0.96, 0.17) = 0.17$
– $rs_{re_3}(c_3) = \min(a_{re_3}(c_3), \min_{c_j \neq c_3}(1 - a_{re_3}(c_j))) = \min(1, 0.57) = 0.57$

As can be observed, though the three considered referring expressions are almost equally true for cell 3, they have rather different degrees of referential success. The two last expressions fail in the objective of unambiguously identifying the referent object to the reader. This is due to the fact that the accuracy of re_2 regarding cell 4 is also high ($a_{re_2}(c_4) = 0.83$) and the accuracy of re_3 regarding cell 27 is not low ($a_{re_2}(c_27) = 0.43$).

The presented measure is a natural extension of the conventional concept of referential success when the accuracy of the referring expression for a given object is fuzzy. However, this measure produces similar values when evaluating referring expressions independently of the number of objects they *fuzzily* refer to.

For example, consider the following sentences:

– re_4 =The dark object.
– re_5 =The object in the upper-left.

In both cases, the referential success with respect to cell 3 is rather low ($rs_{re_4}(c_3) = 0.17$, $rs_{re_5}(c_3) = 0.15$). However, if we do not have any other alternative, we will probably choose re_5 because, as can be seen in table 1, there are fewer objects for which re_5 is true to some extent than there are for re_4. That is, the fuzzy sets of objects of which re_5 is to some extent true is *more specific*, i.e. there is less uncertainty about which referent is intended.

To asses this uncertainty, Fuzzy Set Theory offers a wide variety of measures of specificity of fuzzy sets. If O_{re} is the fuzzy set of objects for which re is true to some extent, with $O_{re}(o) = a_{re}(o)$, then the specificity of re, $Sp(re)$, should satisfy, among others, the following:

– $Sp(re) = 1$ iff O_{re} is a crisp singleton.
– If $O_{re} = \emptyset$, then $Sp(re) = 0$.
– If $card(O_{re}) \leq card(O_{re'})$ and O_{re} and $O_{re'}$ are normal fuzzy sets, then $Sp(re) \geq Sp(re')$, where $card(\cdot)$ is a suitable cardinality measure.

3.3   A multiobjective problem

In order to asses the goodness or adequacy of a referring expression, other quality dimensions can be considered, apart from the ones previously discussed, similarly to what happens in complete linguistic descriptions [3]. The REG literature contains a number of such proposals. Thus, *brevity* has been proposed as a criterion of adequacy [24] based on the conversational maxim, due to Grice [25],

that a cooperative communicator should say no more than is required for the purposes of the exchange. Similarly, *humanlikeness* has been explored based on a variety of psycholinguistic findings, such as for example the extent to which the properties selected in a referring expression are salient or 'preferred' by speakers [26, 27]. Finally, it has been argued that the use of certain properties in a referring expression depends on the relevance to the user given their current communicative task [28] as well as their knowledge state [29].

Though in the example used in this section it has been possible to determine the *best* referring expression for cell 3, in general, the problem of generating a referring expression for a given object does not always lead to a single *best* solution. It is quite normal to find that different measures of quality are correlated to different degrees; for instance, they might exert opposing influences, be contradictory, and/or have a negative correlation. For example, as we add properties to a referring expression, we increase the specificity but we also decrease brevity and probably also accuracy and ease of comprehension. Interestingly, empirical work comparing a large number of REG algorithms against shared datasets suggests that in fact, different evaluation metrics corresponding to different quality dimensions often do not correlate with each other [27].

This means that there is in general no such thing as the optimum or best referring expression for a given object, and the referring expression generation is a multiobjective optimization problem, as explained in [30, 31] for complete linguistic descriptions.

4   REG in the presence of graduality: a discussion

As we have noted, gradability and fuzziness have not received much attention in the REG literature. An exception is the work of van Deemter [12], who proposes a semantics for gradable properties, such as those related to the size of an object, and an algorithm to incorporate these properties in a REG procedure. van Deemter proposes to represent gradable properties numerically, preceding the standard REG content determination procedure with a step in which such properties are converted into inequalities. For example, if o is defined as having a value n for its height, this would be 'unpacked' into a set of inequalities of the form $n > m$, for all m which are values of height that are less than n. A standard REG algorithm can now use these inequalities as it would any other property, assuming that content selection is followed by an appropriate realisation step to convert them into an appropriate natural language string.

This approach is arguably less general than the one proposed in this paper. In particular, if we abandon the view that the relationship between objects and their properties is necessarily boolean (in the sense that an object either has a property or it doesn't) and adopt the more graded notion of *degree of fulfilment*, it is possible to treat all properties as potentially graded, or rather, all objects in a domain as having a particular property to different degrees. The contrast between the definition of accuracy and referential success given in (1) and (2-3) makes this explicit.

4.1 Discriminatory power

From the perspective of knowledge representation, this also eschews a core simplifying assumption made in many approaches to REG, namely, that properties are mutually exclusive. For example, consider an object's colour. In a domain in which some objects are *red* and some are *pink*, it is typically assumed that $[\![\, red\,]\!] \cap [\![\, pink\,]\!] = \emptyset$, so that a REG algorithm that determines, for a target referent *o*, that $o \in [\![\, red\,]\!]$, can safely ignore all objects in the extension of *pink*, since the latter is a completely different property.

This assumption has had an important implication for the development of heuristics to control the search performed by a REG algorithm during content selection, especially where the notion of *discriminatory power* is concerned. Discriminatory power refers to the informativeness of a property, that is, the extent to which it is true of the target referent, but false of other objects in the domain [24]. Thus, algorithms which define the quality of a referring expression in terms of brevity or informativity make use of this notion in trying to generate brief descriptions, by prioritising properties with higher discriminatory power, under the assumption that this will help the reader or hearer to identify the object more efficiently [24, 32, 33]. Returning to our example of colour, if the target referent happens to be the only *red* object in the domain, these algorithms will tend to select this property. The problem is that the object in question might also have a non-zero degree of fulfilment with respect to the property *pink*, based on the closeness of this colour to *red*, so that the discriminatory value of the property selected is not as clear-cut as was initially assumed.

4.2 Salience and contextual knowledge

Explicitly accounting for fuzziness also has implications for the way in which another important quality heuristic is modelled in a REG algorithm, namely, salience or preference. Many algorithms [26, 34] give priority to properties, not based on discriminatory power, but on how salient, relevant or preferred they are. This notion of preference is usually defined by appealing to psycholinguistic findings, for example, the finding that speakers tend to avoid prototypically gradable properties such as size, unless they are absolutely required, in contrast for example to colour which, as we have seen, is used very frequently [35, 36]. However, many such findings are based on experimental domains in which colours are maximally salient and distinct. A fuzzy approach can of course handle such cases, given that the fulfilment degree of an object with respect to a colour which is maximally salient in a domain would be higher than it would be in case the property is less salient, that is, closer to other colours. However, this approach would also enable more nuanced models, in which preference or salience can be computed in less straightforward cases, as in the example in Figure 1(a) above. Indeed, the human tendency to use colouring referring expressions decreases significantly in the presence of colours that are similar, or close to each other [37]. Similarly, the tendency to use size increases, when the size difference between a target referent and its distractors is large [38].

Note that a graded notion of salience is also more faithful to the findings in the vision literature, where many computational models of visual salience identify salient regions based on a combination of the properties we have discussed in this paper, treating these properties as gradable [39, 40]. The framework proposed here could potentially be adapted to the findings from the vision literature, by incorporating a psychologically plausible definition of salience into the computation of degree of fulfilment that determines the likelihood for a property to be selected by a REG procedure.

A further trend in computational research on vision is the incorporation of top-down, or contextual, knowledge [41, 15]. Certain models [42, 43] have successfully modelled human shifts of attention in visual scenes by incorporating both feature-based salience and knowledge of the type of scene being viewed. Thus, the salience of an object depends not only on the distinctiveness of its properties, but on human expectations concerning its location and function in a particular scene. Returning to our running example, when viewing an array of cells, such as those in Figure 1(a), a biologist's attention might be drawn less by their colour (perhaps because this is relatively unimportant, or expected, where cells and tissues are concerned), but by their other properties. Future research on REG will need to take into account these top-down, expectation-based mechanisms, which are also known to influence the way humans refer to objects [44, 45]. It is an open question whether the computation of the degree of fulfilment of a fuzzy property can be modulated by such top-down considerations.

4.3   Spatial relations

Finally, another area in which fuzzy sets can play an important role in REG is in reference to objects using (spatial) relations to other elements of a scene. Since the early work of Dale and Haddock [46], the problem has been mostly defined in logical terms: the elements of a scene that can serve as potential anchors for a target referent in a spatial relation such as *in front of* are themselves encoded explicitly in the REG input [47], though there has been some work that takes a more realistic perspective in selecting salient landmarks [34, 10]. Once again, the simplifying assumptions underlying REG algorithms can be relaxed using a fuzzy approach to spatial relations, especially where these are 'uncertain' and not necessarily mutually exclusive. For instance, in Figure 1(b) if the object to be identified is cell 1, it can be seen that it is above cell 6 but also overlaps with it to some degree. It is not clear whether a good referring expression in this case should combine the two relations (*the small cell at the top which is above and slightly overlaps another*), select only one, perhaps the most salient one, or simply rely on a non-relational frame of reference (*the one at the top towards the middle*). Many techniques have been described to handle fuzzy spatial relationships, and we believe that REG would benefit from an exploration of this literature. Some related work has shown that fuzzy sets can be useful in generating appropriate natural language descriptions of uncertain temporal relations between events [48]; applying similar strategies to spatial relations is a promising way forward.

5   Conclusions

In this paper, we have tackled a classic problem in the generation of natural language from non-linguistic data, namely, referring expression generation. Focussing on visual scenes, we have sought to extend the remit of REG algorithms, by (a) abandoning some simplifying assumptions related to knowledge representation, especially the assumption that properties are crisp sets and that they are either true or false of an object; and (b) extending the notions of accuracy and referential success to deal with the more complex picture that gradability affords. Our discussion has also pointed to several directions for future work, notably the possibility of integrating the types of models proposed in this paper with findings from the vision literature concerning salience. We believe this is a promising direction in which to take research on REG; indeed, it potentially offers a way to bridge the gap between research on fuzzy sets, research on Natural Language Generation, and work in computer vision.

Acknowledgments

References

1. Reiter, E., Dale, R.: Building Natural Language Generation Systems. Cambridge University Press, Cambridge, UK (2000)
2. Kacprzyk, J., Zadrozny, S.: Computing with words is an implementable paradigm: Fuzzy queries, linguistic data summaries, and natural-language generation. IEEE Trans. Fuzzy Systems **18**(3) (2010) 461–472
3. Marín, N., Sánchez, D.: On generating linguistic descriptions of time series. Fuzzy Sets and Systems **285** (2016) 6–30
4. Ramos-Soto, A., Bugarín, A., Barro, S.: On the role of linguistic descriptions of data in the building of natural language generation systems. Fuzzy Sets and Systems **285** (2016) 31–51
5. Mitchell, M., Dodge, J., Goyal, A., Yamaguchi, K., Stratos, K., Han, X., Mensch, A., Berg, A., Han, X., Berg, T., Daume III, H.: Midge: Generating Image Descriptions From Computer Vision Detections. In: EACL 2012), Avignon, France, Association for Computational Linguistics (2012) 747–756
6. Kulkarni, G., Premraj, V., Ordonez, V., Dhar, S., Li, S., Choi, Y., Berg, A.C., Berg, T.L.: Baby talk: Understanding and generating simple image descriptions. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(12) (2013) 2891–2903
7. Yatskar, M., Galley, M., Vanderwende, L., Zettlemoyer, L.: See No Evil , Say No Evil : Description Generation from Densely Labeled Images. In: Proceedings of the Third Joint Conference on Lexical and Computation Semantics (*SEM). (2014)
8. Reiter, E., Dale, R.: A fast algorithm for the generation of referring expressions. In: COLING'92. (1992) 232–238

9. Krahmer, E., van Deemter, K.: Computational generation of referring expressions: A survey. Computational Linguistics **38**(1) (2012) 173–218
10. Elsner, M., Rohde, H., Clarke, A.D.F.: Information Structure Prediction for Visual-world Referring Expressions. In: EACL 2014), Gothenburg, Sweden, Association for Computational Linguistics (2014) 520–529
11. Kazemzadeh, S., Ordonez, V., Matten, M., Berg, T.L.: ReferItGame: Referring to Objects in Photographs of Natural Scenes. In: EMNLP 2014), Doha, Qatar, Association for Computational Linguistics (2014) 787–798
12. van Deemter, K.: Generating referring expressions that involve gradable properties. Computational Linguistics **32**(2) (2006) 195–222
13. Turner, R., Sripada, S., Reiter, E., Davy, I.P.: Selecting the Content of Textual Descriptions of Geographically Located Events in Spatio-Temporal Weather Data. In: Applications and Innovations in Intelligent Systems XV. (2008) 75–88
14. Reiter, E., Sripada, S., Hunter, J., Yu, J., Davy, I.: Choosing words in computer-generated weather forecasts. Artif. Intell. **167**(1-2) (2005) 137–169
15. Cadenas, J.T., Marín, N., Miranda, M.A.V.: Context-aware fuzzy databases. Appl. Soft Comput. **25** (2014) 215–233
16. Castillo-Ortega, R., Chamorro-Martínez, J., Marín, N., Sánchez, D., Soto-Hidalgo, J.M.: Describing images via linguistic features and hierarchical segmentation. In: FUZZ-IEEE 2010. (2010) 1–8
17. Soto-Hidalgo, J.M., Chamorro-Martínez, J., Sánchez, D.: A new approach for defining a fuzzy color space. In: FUZZ-IEEE 2010. (2010) 1–6
18. Chamorro-Martínez, J., Medina, J.M., Barranco, C.D., Galán-Perales, E., Soto-Hidalgo, J.M.: Retrieving images in fuzzy object-relational databases using dominant color descriptors. Fuzzy Sets and Systems **158**(3) (2007) 312–324
19. Chamorro-Martínez, J., Martínez-Jiménez, P.M., Soto-Hidalgo, J.M., León-Salas, A.: A fuzzy approach for modelling visual texture properties. Inf. Sci. **313** (2015) 1–21
20. Chamorro-Martínez, J., Martínez-Jiménez, P.M., Soto-Hidalgo, J.M., Prados-Suárez, B.: Fuzzy sets on 2D spaces for fineness representation. Int. J. Approx. Reasoning **62** (2015) 46–60
21. Prados-Suárez, B., Chamorro-Martínez, J., Sánchez, D., Abad, J.: Region-based fit of color homogeneity measures for fuzzy image segmentation. Fuzzy Sets and Systems **158**(3) (2007) 215–229
22. Hudelot, C., Atif, J., Bloch, I.: Fuzzy spatial relation ontology for image interpretation. Fuzzy Sets and Systems **159**(15) (2008) 1929–1951
23. Buck, A.R., Keller, J.M., Skubic, M.: A memetic algorithm for matching spatial configurations with the histograms of forces. IEEE Trans. Evolutionary Computation **17**(4) (2013) 588–604
24. Dale, R.: Cooking up referring expressions. In: ACL 1989), Vancouver, BC, Association for Computational Linguistics (1989) 68–75
25. Grice, H.P.: Logic and conversation. In: Syntax and Semantics 3: Speech Acts. Elsevier, Amsterdam (1975) 41–58
26. Dale, R., Reiter, E.: Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. Cognitive Science **19**(2) (apr 1995) 233–263
27. Gatt, A., Belz, A.: Introducing shared task evaluation to NLG: The TUNA shared task evaluation challenges. In Krahmer, E., Theune, M., eds.: Empirical methods in natural language generation. Springer, Berlin and Heidelberg (2010)
28. Jordan, P.W., Walker, M.a.: Learning content selection rules for generating object descriptions in dialogue. Journal of Artificial Intelligence Research **24** (2005) 157–194

29. Janarthanam, S., Lemon, O.: Adaptive Generation in Dialogue Systems Using Dynamic User Modeling. Computational Linguistics **40**(4) (dec 2014) 883–920
30. Castillo-Ortega, R., Marín, N., Sánchez, D., Tettamanzi, A.: Quality assessment in linguistic summaries of data. In: IPMU 2012. (2012) 285–294
31. Bugarín, A., Marín, N., Sánchez, D., Triviño, G.: Aspects of quality evaluation in linguistic descriptions of data. In: FUZZ-IEEE 2015. (2015) 1–8
32. Gardent, C.: Generating Minimal Definite Descriptions. In: ACL 2002. (July 2002) 96–103
33. Frank, M.C., Goodman, N.D.: Predicting pragmatic reasoning in language games. Science (New York, N.Y.) **336**(6084) (may 2012) 998
34. Kelleher, J.D., Kruijff, G.J.G.J.: Incremental generation of spatial referring expressions in situated dialog. In: COLING-ACL 2006, Sydney, Australia, Association for Computational Linguistics (2006) 1041–1048
35. Pechmann, T.: Incremental speech production and referential overspecification. Linguistics (1989)
36. Belke, E., Meyer, A.S.: 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) (apr 2002) 237–266
37. Viethen, J., Goudbeek, M., Krahmer, E.: The impact of colour difference and colour codability on reference production. In: CogSci 2012, Austin, TX, Cognitive Science Society (2012) 1084–1089
38. van Gompel, R.P., Gatt, A., Krahmer, E., van Deemter, K.: Overspecification in reference: Modelling size contrast effects. In: AMLAP 2014. (2014)
39. Itti, L., Koch, C.: Computational modelling of visual attention. Nature reviews neuroscience **2**(3) (mar 2001) 194–203
40. Erdem, E., Erdem, A.: Visual saliency estimation by nonlinearly integrating features using region covariances. Journal of vision **13**(4) (2013) 11, 1–20
41. Oliva, A., Torralba, A.: The role of context in object recognition. Trends in cognitive sciences **11**(12) (dec 2007) 520–527
42. Torralba, A., Oliva, A., Castelhano, M.S., Henderson, J.M.: Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. Psychological Review **113**(4) (2006) 766–786
43. Kanan, C., Tong, M.H., Zhang, L., Cottrell, F.W.: SUN: Top-down saliency using natural statistics. Visual Cognition **17**(6-7) (2009) 979–1003
44. Sedivy, J.C.: Pragmatic versus form-based accounts of referential contrast: evidence for effects of informativity expectations. Journal of psycholinguistic research **32**(1) (jan 2003) 3–23
45. Westerbeek, H., Koolen, R., Maes, A.: Stored object knowledge and the production of referring expressions: the case of color typicality. Frontiers in Psychology **6**(July) (2015) 1–12
46. Dale, R., Haddock, N.: Generating referring expressions involving relations. In: EACL 1991, Berlin, Germany (1991) 161–166
47. Areces, C., Koller, A., Striegnitz, K.: Referring expressions as formulas of description logic. In: INLG 2008. (2008) 42–49
48. Gatt, A., Portet, F.: Multilingual generation of uncertain temporal expressions from data: A study of a possibilistic formalism and its consistency with human subjective evaluations. Fuzzy Sets and Systems **285** (2016) 73–93