Fuzzy Logic for Vagueness Management in Referring Expression Generation[1]

N. Marín, G. Rivas-Gervilla, and D. Sánchez

Dept. Computer Science and A.I., and CITIC-UGR

University of Granada, 18071 Granada, Spain

{nicm,griger,daniel}@decsai.ugr.es

Abstract

In this work we overview some of the contributions regarding the use of Fuzzy Logic in referring expression generation. We also discuss recent advances that can help to overcome the arguments in the literature against the use of Fuzzy Logic in Natural Language Generation.

1 Introduction

Different types of vagueness are present in natural language employed for human communication (van Deemter, 2010). In this paper we are concerned with the kind of vagueness related to concepts, words, and linguistic expressions that allow for borderline cases in which fulfilment is not clear (van Deemter, 2010). This kind of vagueness cannot be properly represented and managed using classical logics, hence alternative tools are necessary (van Deemter, 2010). A widely employed tool for dealing with this kind of vagueness is Fuzzy Logic (Kacprzyk and Zadrozny, 2010).

Fuzzy Logic has been used in Natural Language Generation (NLG) for different purposes (Marín and Sánchez, 2016; Ramos-Soto et al., 2016): modeling the semantics of concepts and expressions, uncertainty representation and quality measurement, among others. In this paper we focus on the use of Fuzzy Logic for the aforementioned purposes in the setting of referring expression generation (REG), one of the crucial tasks of NLG (Gatt and Krahmer, 2018). We also briefly discuss recent contributions that enlarge the available Fuzzy Logic toolbox with new capabilities, solving some issues

that have been argued in order to disregard the use of Fuzzy Logic in NLG.

2 Referring expression generation

Given a set of objects $\mathcal{O}$ with properties in $\mathcal{P}$, and given $o \in \mathcal{O}$, the objective of the REG task is to provide a linguistic expression able to identify o within $\mathcal{O}$. It is usual to distinguish two steps in REG: *extraction* and *expression* (Marín and Sánchez, 2016). In the first one, the semantics of the referring expression is represented by means of some knowledge representation formalism, typically a formal logic. In the second one, an appropriate sentence in natural language is provided.

In this paper we are concerned only with the extraction phase in which, in its most basic form, a referring expression is a conjunction of properties in $\mathcal{P}$, usually represented by the set of properties that appear in that conjunction. More general structures can be employed involving negation and disjunction, as well as generalized quantifiers. Other generalizations to the problem are the reference to sets of objects, the use of relational properties represented by mathematical relations between objects, the use of collective properties defined for sets of objects, and the use of *gradual* concepts (Krahmer and Van Deemter, 2012). The latter are the object of our interest here.

3 Vagueness in words and linguistic expressions

Vagueness due to borderline/intermediate cases appears typically because of the use of gradual concepts in language[2]. One example is the concept *large* regarding the size of an object since, within

---

[2]The terms *gradual* or *fuzzy* are the usual ones in the Soft Computing area for this kind of properties. The term *gradable* is also common in the literature (van Deemter, 2016).

the set of all possible sizes, some values match the concept *large*, some others do not match the concept at all, whilst the rest are intermediate cases that match the concept to a certain extent. Hence, fulfilment of a concept becomes a matter of degree.

Gradual concepts are products of the human mind and abound in human language and communication. As a consequence, they are of primary interest in Artificial Intelligence, particularly when it comes to developing systems for linguistic interaction with humans. One crucial problem is how to represent the semantics of such concepts. It is well known that crisp sets are not well suited for that purpose, since elements either fully belong to the set or to its complement. The only way to represent the semantics of *large* using a crisp set is by giving a size threshold above which *large* holds. However, this solution gives counterintuitive results, since a small variation in size near the threshold is enough to turn a large size into a non-large one when, in fact, both sizes may be even indistinguishable. Indeed, when asked about whether something is *large*, humans do not always provide a yes/no answer, but expressions like "more or less", "so-so", etc. that cannot be represented by a crisp set. Another example is the definition of the concept *heap*, leading to the classical Sorites Paradox (van Deemter, 2010).

Fuzzy Logic is recognized as a suitable tool for representing the semantics of gradual concepts (Rosch, 2013). A fuzzy set F assigns a fulfilment degree in $[0, 1]$ to each value of the domain X where the concept is defined, by means of a *membership function* $\mu_F : X \to [0, 1]$. This way, the semantics of *large* can be represented by means of a continuous function on the set of sizes, in which "small" differences in size produce "small" differences in membership. That is, the transition from being *large* (1) to not being *large* (0) becomes gradual, allowing to represent intermediate cases intuitively by assigning them degrees in $(0, 1)$.

Atomic gradual concepts like *large* can be combined to form *derived* gradual concepts, by means of logical connectives. One of the particularities of Fuzzy Logic is that different operators can be used in order to compute the membership function of derived properties: a wide range of t-norms for intersection, t-conorms for union, and fuzzy negations for complement are available. Other kind of derived gradual concepts can be obtained from the application of *linguistic hedges* that modify the semantics of concepts. The semantics of such derived

concepts (like *very large*) are obtained by means of a composition of a function associated to the hedge (*very*, a typical function being $very(x) = x^2$) and the membership function representing the semantics of the concept being modified (*large*).

Gradual concepts can be also employed to form more complex expressions called *protoforms*, that are one of the main objects of study of the *Computing with Words* (CW) area (Zadeh, 1999; Kacprzyk and Zadrozny, 2010). Fulfilment of protoforms is also gradual, degrees in $[0, 1]$ being "computed" from the semantics of the concepts involved.

Protoforms can be expressed linguistically, a paradigmatic example being quantified statements like "*most* of the *large* animals are *slow*", which is a particular instantiation of the protoform "Q of D are A". In this protoform, Q is a gradual quantifier (*most* in our previous example) with semantics represented by a fuzzy set $\mu_Q : [0, 1] \to [0, 1]$ assigning fulfilment degrees of the quantifier to percentages in $[0, 1]$. For instance, a particular semantics of *most* is given by the following continuous and piecewise-linear function:

$$Q(x) = \begin{cases} 0 & x \leq 0.5 \\ 4x - 2 & 0.5 \leq x \leq 0.75 \\ 1 & 0.75 \leq x \end{cases} \quad (1)$$

On its turn, both D and A are fuzzy subsets of the same set X (animals), induced by gradual concepts (*large* and *slow*, respectively). Techniques for computing the fulfilment degree of such sentences, including more complex sentences involving generalized quantifiers, are available (Delgado et al., 2014; Díaz-Hermida et al., 2018).

Graduality can appear in combination with other sources of uncertainty in protoforms, like probability (Zadeh, 1999). Besides, gradual concepts can be used for different purposes in protoforms. A particular case is the *possibilistic use*, in which gradual concepts are employed as restrictions representing the available knowledge. An example of instantiation of such protoforms is "John is *old*", where we lack some knowledge about the actual age of John, but we know it to be restricted to the set of ages that match the gradual concept *old*, membership degrees representing our preference for some ages against others if we had to guess. Note the difference with "I like *old* cars", in which the same gradual concept appears under a *veristic use* and, contrary to the previous case, there is no

uncertainty (the expression claims that I like every old car, membership degrees corresponding to the degree to which I like every car because of its age).

Let us remark that fulfilment of certain proto-forms can be computed from other protoforms using rule-based inference, among other kind of reasoning, like in *granular linguistic models of phenomena* (Triviño and Sugeno, 2013).

The use of fuzzy sets has been discussed specifically for REG in (Gatt et al., 2016), where some of the properties in $\mathcal{P}$ are assumed to be gradual. As a consequence, features like referential success of referring expressions become gradual, as we shall discuss in the next section. Though in (Gatt et al., 2016) only conjunctions of atomic properties are considered, it is immediate to extend the discussion to both derived properties and protoforms like those discussed before, for instance under a possibilistic use of gradual concepts (Gatt and Portet, 2016; Marín et al., 2019). The case of quantified statements is also particularly interesting in REG, as it has been shown by using crisp generalized quantification in (Ren et al., 2010).

4 Vagueness and measures

Quality assessment is a fundamental question in any intelligent system, Data2text systems (including REG approaches) not being an exception. Quality assessment models are necessary for two fundamental reasons: (i) they must guide searching processes and (ii) they must allow the results obtained to be evaluated and compared (Bugarín et al., 2015b,a).

The look for quality models is not trivial because they are usually context-dependent and combine aspects that, in many cases, turn out to be subjective and interdependent/conflicting: it is necessary to obey the user's preferences in the context the system is executed (Marín and Sánchez, 2016). In general, it is a multidimensional problem that requires multi-objective optimization algorithms (Castillo-Ortega et al., 2012).

An important part of the mentioned quality models focuses on the definition of measures that allow to assess different aspects of quality with values in $[0, 1]$, even when gradual concepts are not involved. For example we can consider:

- The accuracy, that is, the degree of fulfillment of the expression by the target object set. We have discussed about accuracy of gradual concepts and protoforms in the previous section.

- The brevity, that is, minimizing the length of the expression. Brevity can be measured in $[0, 1]$ when length is divided by the maximum possible length.

- The salience, related to how easy is to perceive the properties employed in the expression. In general, more salient expressions are preferred by users.

As can be seen in these three examples, measures can be gradual, particularly –but not necessarily– in the presence of gradual properties. Fuzzy Logic provides both mechanisms to solve the problem of *symbol grounding* and an extensive prior knowledge in the form of a wide variety of well known fuzzy measures, which can be used as a basis for measuring such quality related aspects.

In the case of referring expressions, the case of referential success deserves special mention. In a *crisp* environment, referential success can be considered as an inherent quality of a referring expression since it makes no sense to use a referring expression lacking referential success. However, in the presence of gradual properties, determining the referential success is also a matter of degree (Gatt et al., 2016): referring expressions with a high degree of referential success are then searched, preferably the best for each target set.

As we have previously mentioned, Fuzzy Logic permits defining gradual measures of referential success. For example, in (Gatt et al., 2016), a definition of referential success based on accuracy is proposed, measuring to what extent the accuracy occurs in the target set whilst it does not occur in the other distractors. Referential success measurement is also a clear example of how the broad background in measures of Fuzzy Logic can be useful for solving problems in the REG field. For example, studies can be found in the literature (Marín et al., 2016b, 2017a, 2018b) which show that referential success can be assessed using the well-known measures of specificity of Fuzzy Logic (Yager, 1982). In (Gatt et al., 2018) interested readers can find an experimental analysis with users regarding a variety of specificity based measures of referential success.

Additional measures have to be considered in the possibilistic setting mentioned in the previous section. Consider, for example, datasets in which we find objects satisfying a property among a given set of properties, but not knowing exactly which

one. The presence of this type of uncertainty in the information is usually translated into uncertainty in the set of objects that satisfy a given expression. In (Marín et al., 2019) a novel fuzzy set based approach to this problem can be found, where innovative notions like *possible referent* and *necessary referent* of an expression are defined and used as basis for a gradual measure of referential success.

In addition to quality measures of final referring expressions, other fuzzy measures have been proposed for guiding the REG process, such as measures of discriminatory power, for example by means of index-type specificity measures (Marín et al., 2017b), among others.

5 Discussion

Different arguments have been put forward against the use of Fuzzy Logic in modeling and reasoning with concepts in REG and NLG in general.

Fuzzy Logic was deemed unsuitable for representing and dealing with gradual concepts (Osherson and Smith, 1981). Recent work has shown that such claims are erroneous, mainly due to misunderstandings and misconceptions such as using a set theory as a theory of concepts, (Belohlávek et al., 2009; Belohlávek and Klir, 2011; Rosch, 2013). Such use is wrong since, though derived concepts can be obtained by using Fuzzy Logic operations, not every complex concept allows to obtain its semantics by operations on sets, requiring specific modeling of their membership functions instead.

Defining appropriate membership functions representing the semantics of gradual concepts is a complex problem because of subjectivity and context-dependence (Cadenas et al., 2014), among other reasons (van Deemter, 2010). This has been employed as an argument against using Fuzzy Logic as well. However, the same can be said of using crisp sets; for instance, the definition of concepts like *large* using crisp sets requires to define subjective and context-dependent thresholds, the semantics of such models being much more sensitive to small changes on thresholds than when using Fuzzy Logic. In general, the symbol grounding problem is shared by every knowledge representation formalism (Harnad, 1990).

Fortunately, more and more techniques for appropriately solving this problem in different settings are available in the literature, see (Chamorro-Martínez et al., 2017; Ramos-Soto et al., 2019) for recent proposals, the second one in relation to REG.

Also related to REG is the proposal in (Marín et al., 2018a), in which the context-dependent semantics of terms like *large*, *medium*, and *small* – interpreted as *the largest*, etc. (van Deemter, 2006) – are automatically calculated according to the collection of size values of the objects in the context, without requiring the intervention of humans beyond fixing once and for all the axioms that the models must satisfy. A similar idea has been employed for modeling the semantics of crisp contextual properties in (Fernández, 2009).

Regarding operations and reasoning, the availability of different ways for performing usual set operations with fuzzy sets and the fact that they are truth-functional (which imply that no Fuzzy Set Theory is a Boolean algebra) has been pointed out as a disadvantage (van Deemter, 2010). The recent development of level-based representations (RLs) as an alternative to fuzzy sets can solve this problem (Dubois and Prade, 2008; Sánchez et al., 2008, 2012; Martin, 2015). RLs have been used in REG (Marín et al., 2016a).

The approach in (Sánchez et al., 2012) represents gradual concepts by means of functions $\rho : (0, 1] \rightarrow 2^X$ instead of fuzzy sets, going beyond Fuzzy Logic in several respects:

- Every classical set operation is extended to the gradual case uniquely in a non-truth functional way by performing the operation in each level of $(0, 1]$ independently, keeping all Boolean properties.

- Fuzzy sets are employed as input (by using α-cuts) and output (measuring membership) only. This way, the understandability and modeling resources of fuzzy sets and the algebraic properties of RLs operations are combined into *RL-systems* that solve some of the drawbacks associated to fuzzy reasoning.

This discussion leads us to believe that Fuzzy Logic, Computing with Words, and RL-systems can help in dealing with graduality/uncertainty in REG and other NLG tasks.

References

Radim Belohlávek and George J. Klir, editors. 2011. *Concepts and fuzzy logic*. The MIT Press, Cambridge, Massachussets.

Radim Belohlávek, George J. Klir, Harold W. Lewis III, and Eileen C. Way. 2009. Concepts and fuzzy sets:

Misunderstandings, misconceptions, and oversights. *Int. J. Approx. Reason.*, 51(1):23–34.

Alberto Bugarín, Nicolás Marín, Daniel Sánchez, and Gracián Triviño. 2015a. Aspects of quality evaluation in linguistic descriptions of data. In *2015 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2015, Istanbul, Turkey, August 2-5, 2015*, pages 1–8. IEEE.

Alberto Bugarín, Nicolás Marín, Daniel Sánchez, and Gracián Triviño. 2015b. Fuzzy knowledge representation for linguistic description of time series. In *2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15), Gijón, Spain., June 30, 2015*. Atlantis Press.

José Tomás Cadenas, Nicolás Marín, and María Amparo Vila Miranda. 2014. Context-aware fuzzy databases. *Appl. Soft Comput.*, 25:215–233.

Rita Castillo-Ortega, Nicolás Marín, Daniel Sánchez, and Andrea Tettamanzi. 2012. Quality assessment in linguistic summaries of data. In *Advances on Computational Intelligence - 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2012, Catania, Italy, July 9-13, 2012. Proceedings, Part II*, volume 298 of *Communications in Computer and Information Science*, pages 285–294. Springer.

Jesús Chamorro-Martínez, José Manuel Soto-Hidalgo, Pedro Manuel Martínez-Jiménez, and Daniel Sánchez. 2017. Fuzzy color spaces: A conceptual approach to color vision. *IEEE Trans. Fuzzy Syst.*, 25(5):1264–1280.

Kees van Deemter. 2006. Generating referring expressions that involve gradable properties. *Computational Linguistics*, 32(2):195–222.

Kees van Deemter. 2016. *Computational models of referring: a study in cognitive science*. The MIT Press, Cambridge, Massachussets.

Miguel Delgado, M. Dolores Ruiz, Daniel Sánchez, and María-Amparo Vila. 2014. Fuzzy quantification: a state of the art. *Fuzzy Sets Syst.*, 242:1–30.

Félix Díaz-Hermida, Martin Pereira-Fariña, Juan Carlos Vidal, and Alejandro Ramos-Soto. 2018. Characterizing quantifier fuzzification mechanisms: A behavioral guide for applications. *Fuzzy Sets Syst.*, 345:1–23.

D. Dubois and H. Prade. 2008. Gradual elements in a fuzzy set. *Soft Computing*, 12:165–175.

Raquel Fernández. 2009. Salience and feature variability in definite descriptions with positive-form vague adjectives. In *Proceedings Workshop on the Production of Referring Expressions: Bridging the gap between computational and empirical approaches to reference (PRE-CogSci 2009)*.

Albert Gatt and Emiel Krahmer. 2018. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *J. Artif. Intell. Res.*, 61:65–170.

Albert Gatt, Nicolás Marín, François Portet, and Daniel Sánchez. 2016. The role of graduality for referring expression generation in visual scenes. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems - 16th International Conference, IPMU 2016, Eindhoven, The Netherlands, June 20-24, 2016, Proceedings, Part I*, volume 610 of *Communications in Computer and Information Science*, pages 191–203. Springer.

Albert Gatt, Nicolás Marín, Gustavo Rivas-Gervilla, and Daniel Sánchez. 2018. Specificity measures and reference. In *Proceedings of the 11th International Conference on Natural Language Generation, Tilburg University, The Netherlands, November 5-8, 2018*, pages 492–502. Association for Computational Linguistics.

Albert Gatt and François Portet. 2016. Multilingual generation of uncertain temporal expressions from data: A study of a possibilistic formalism and its consistency with human subjective evaluations. *Fuzzy Sets Syst.*, 285:73–93.

Stevan Harnad. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1):335 – 346.

Janusz Kacprzyk and Slawomir Zadrozny. 2010. Computing with words is an implementable paradigm: Fuzzy queries, linguistic data summaries, and natural-language generation. *IEEE Trans. Fuzzy Systems*, 18(3):461–472.

Emiel Krahmer and Kees Van Deemter. 2012. Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1):173–218.

Nicolás Marín, Gustavo Rivas-Gervilla, and Daniel Sánchez. 2016a. A measure of referential success based on alpha-cuts. In *Scalable Uncertainty Management - 10th International Conference, SUM 2016, Nice, France, September 21-23, 2016, Proceedings*, volume 9858 of *Lecture Notes in Computer Science*, pages 345–351. Springer.

Nicolás Marín, Gustavo Rivas-Gervilla, and Daniel Sánchez. 2016b. Using specificity to measure referential success in referring expressions with fuzzy properties. In *2016 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2016, Vancouver, BC, Canada, July 24-29, 2016*, pages 563–570. IEEE.

Nicolás Marín, Gustavo Rivas-Gervilla, and Daniel Sánchez. 2017a. Referential success of set referring expressions with fuzzy properties. In *Proceedings of the 10th International Conference on Natural Language Generation, INLG 2017, Santiago de Compostela, Spain, September 4-7, 2017*, pages 247–251. Association for Computational Linguistics.

Nicolás Marín, Gustavo Rivas-Gervilla, and Daniel Sánchez. 2018a. An approximation to context-aware size modeling for referring expression generation. In *2018 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8.

Nicolás Marín, Gustavo Rivas-Gervilla, and Daniel Sánchez. 2019. Referring under uncertainty. In *2019 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2019, New Orleans, LA, USA, June 23-26, 2019*, pages 1–6. IEEE.

Nicolás Marín, Gustavo Rivas-Gervilla, Daniel Sánchez, and Ronald R. Yager. 2017b. On families of bounded specificity measures. In *2017 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2017, Naples, Italy, July 9-12, 2017*, pages 1–6. IEEE.

Nicolás Marín, Gustavo Rivas-Gervilla, Daniel Sánchez, and Ronald R. Yager. 2018b. Specificity measures and referential success. *IEEE Trans. Fuzzy Systems*, 26(2):859–868.

Nicolás Marín and Daniel Sánchez. 2016. On generating linguistic descriptions of time series. *Fuzzy Sets and Systems*, 285:6–30.

Trevor P. Martin. 2015. The X-μ representation of fuzzy sets. *Soft Computing*, 19(6):1497–1509.

Daniel N. Osherson and Edward E. Smith. 1981. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9(1):35–58.

Alejandro Ramos-Soto, José M. Alonso, Ehud Reiter, Kees van Deemter, and Albert Gatt. 2019. Fuzzy-based language grounding of geographical references: From writers to readers. *Int. J. Comput. Intell. Syst.*, 12(2):970–983.

Alejandro Ramos-Soto, Alberto Bugarín, and Senén Barro. 2016. On the role of linguistic descriptions of data in the building of natural language generation systems. *Fuzzy Sets and Systems*, 285:31–51.

Yuan Ren, Kees Van Deemter, and Jeff Z Pan. 2010. Generating referring expressions with OWL2. In *23rd International Workshop on Description Logics DL2010*, page 420.

Eleanor Rosch. 2013. Neither concepts nor Lotfi Zadeh are fuzzy sets. In *On Fuzziness - A Homage to Lotfi A. Zadeh - Volume 2*, volume 299 of *Studies in Fuzziness and Soft Computing*, pages 591–596. Springer.

Daniel Sánchez, Miguel Delgado, and María-Amparo Vila. 2008. A restriction level approach to the representation of imprecise properties. In *Proceedings Int. Conference on Information Processing and Management of Uncertainty IPMU'08*, pages 153–159.

Daniel Sánchez, Miguel Delgado, María-Amparo Vila, and Jesús Chamorro-Martínez. 2012. On a non-nested level-based representation of fuzziness. *Fuzzy Sets and Systems*, 192(1):159–175.

Gracián Triviño and Michio Sugeno. 2013. Towards linguistic descriptions of phenomena. *Int. J. Approx. Reasoning*, 54(1):22–34.

Kees van Deemter. 2010. *Not Exactly: in Praise of Vagueness*. Oxford University Press.

Ronald R. Yager. 1982. Measuring tranquility and anxiety in decision making: an application of fuzzy sets. *International Journal of General Systems*, 8(3):139–146.

Lotfi A. Zadeh. 1999. From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 46(1):105–119.