

Modeling language constructs with fuzzy sets: some approaches, examples and interpretations

Pavlo Kapustin
University of Bergen
pavlo.kapustin@uib.no

Michael Kapustin
Moscow Institute of Physics and Technology
michael.kapustin@gmail.com

Abstract

We present and discuss a couple of approaches, including different types of projections, and some examples, discussing the use of fuzzy sets for modeling meaning of certain types of language constructs. We are mostly focusing on words other than adjectives and linguistic hedges as these categories are the most studied from before. We discuss logical and linguistic interpretations of membership functions. We argue that using fuzzy sets for modeling meaning of words and other natural language constructs, along with situations described with natural language is interesting both from purely linguistic perspective, and also as a meaning representation for problems of computational linguistics and natural language processing.

1 Introduction

The use of fuzzy sets for representing meaning of some types of natural language constructs was first proposed and described in earlier works of Lotfi Zadeh (Zadeh, 1971, 1972). Representation based on fuzzy sets is very expressive as it allows to quantitatively model the nature of the relationship between different concepts, and represent vagueness and imprecision that are so common to natural language.

Nowadays, fuzzy sets seem to be relatively little known among linguists, and little used in natural language processing (Carvalho et al., 2012; Novák, 2017). Most of the examples described in the literature include certain types of adjectives and linguistic hedges.

We would like to contribute to this field by describing a couple of approaches, including different types of projections, that can be used for modeling meaning of some types of language constructs using fuzzy sets. We describe and discuss examples that include some adjectives, adverbs and prepositions. We discuss logical and linguistic interpretations of membership functions (Hersh and Caramazza, 1976), and argue for importance of distinguishing between them when modeling language constructs with fuzzy sets.

2 Related work

Here we briefly mention some of the work related to the use of fuzzy sets as a meaning representation.

In his early works, Lotfi Zadeh suggests modeling meaning of certain types of adjectives (e.g. “small”, “medium”, “large”) as fuzzy sets, and some linguistic hedges (e.g. “very”, “slightly” — as operators, acting on these fuzzy sets (Zadeh, 1971, 1972). Hersh and Caramazza (1976) introduce logical and linguistic interpretations of membership functions.

Novák (2017) describes Fuzzy Natural Logic, a mathematical theory that attempts to model semantics of natural language, including Theory of Evaluative Linguistic Expressions (Novák, 2008). Some ways of modeling meaning of words like nouns and verbs have also been suggested (Novák, 1992, 2017; M. Kapustin and P. Kapustin, 2015). Novák et al. (2016) includes an example of evaluative linguistic expressions that contain perceptions like “near” and “far”.

In M. Kapustin and P. Kapustin (2015) we describe a framework for computational interpreting of natural language fragments, and suggest modeling meaning of words as operators. P. Kapustin (2015) describes an application that implements and tests some features of this framework in a simplified setting.

There is some work aiming to make fuzzy sets easier to learn from data. For example, Runkler (2016) describes an approach for generation of linguistically meaningful membership functions from word vectors. We describe compatibility intervals, a meaning representation that is closely related to fuzzy sets (P. Kapustin and M. Kapustin, 2019b).

We discuss how people relate some language constructs to compatibility intervals in an experimental study (P. Kapustin and M. Kapustin, 2019a).

3 Method

3.1 Projections

In this paper, we describe modeling meaning of language constructs by approximating it with a set of *projections* of this construct on different properties (here term “property” is used in a relatively general sense).¹ Each such projection is defined by a fuzzy set and a corresponding membership function that describes compatibility of the construct with different values that the respective property may take. The intuition behind this approach is simple: language constructs contain information about different properties, and information about each property can be modeled as an independent projection.²

Consider fig. 1. Presented membership functions attempt to quantitatively relate constructs “expected”, “common”, “possible”, “extraordinary” to surprisingness of a certain result. Of course, meaning of mentioned words is complex and cannot be fully described in terms of surprisingness, but they do tell us something about it, among other things. So, these membership functions may be seen as *projections* of the meanings of these constructs onto property “surprisingness”.

3.2 Membership function arguments and values

Regarding values of membership function arguments (in this case, values of “surprisingness”), here we are using a relative scale ranging from zero to one. Choice of scale, including its type (linear, logarithmic, etc.), and mapping between real values and relative values is a topic of separate research and is beyond the scope of this paper.

We look at interpreting membership functions values similarly to Zadeh (1975, 1978): values of membership function can be seen as degrees of compatibility between the value of the function argument and the construct the membership function is describing. Consider fig. 1: $\mu_{\text{expected}}(1) = \mu_{\text{common}}(1) = 0$, because constructs “expected” and “common” are not compatible with high values of surprisingness, and $\mu_{\text{extraordinary}}(1) = 1$, because “extraordinary” is highly compatible with high values of surprisingness (μ is denoting degree of membership).

3.3 Membership functions: different interpretations

Similarly to Hersh and Caramazza (1976), we distinguish between two different interpretations of membership functions: *logical* (modeling what is “logically”, or “technically” correct), and *linguistic* (modeling how the word is used).

Consider fig. 2: “young1” corresponds to logical interpretation, reflecting the fact that infants and newborns are, indeed, as young as one can be. On the other hand, “young2” corresponds to linguistic interpretation, reflecting the fact that when people use the word “young”, they usually refer to ages other

¹We propose a related but a bit more specific definition of properties in M. Kapustin and P. Kapustin (2015).

²This approach may be seen as a slight generalization of the ideas described in Lotfi Zadeh’s early works (Zadeh, 1971, 1972), where construct meaning is modeled as one fuzzy set (one projection), and as a special case of the approach we suggest in M. Kapustin and P. Kapustin (2015), where each concept is modeled as an operator.

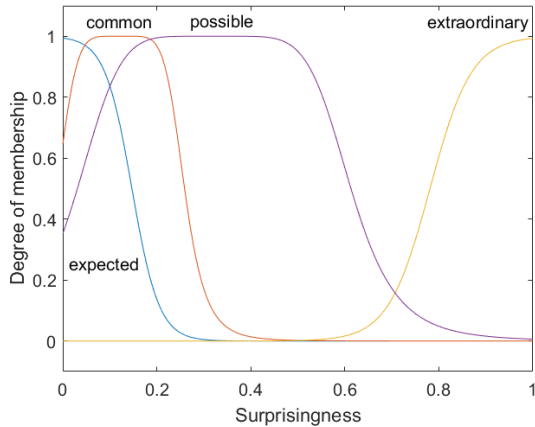


Figure 1: “Expected”, “common”, “possible”, “extraordinary” related to “surprisingness”.

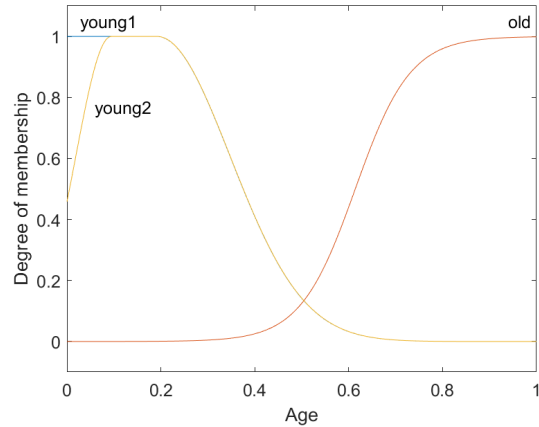


Figure 2: Logical (young1) and linguistic (young2) interpretations.

than newborns and infants. However, for the word “old” its usage does not differ from what is “logically” correct: we may say “old” about someone who is 80 or 100 years old.

Let’s consider fig. 1 again: $\mu_{\text{expected}}(0) > \mu_{\text{common}}(0) > \mu_{\text{possible}}(0)$. This corresponds to linguistic interpretation and models that, even though highly anticipated results are probably both common and possible, “expected” might be a better word than “common” (and especially than “possible”) to describe such results (of course, this only takes “surprisingness” into account).

We believe that many, but probably, not all of the differences between logical and linguistic interpretations are related to scalar implicatures and related phenomena, and believe that this needs to be investigated further.

Differing logical and linguistic interpretations have some interesting implications. Consider fig. 3. Here we apply negation, implemented as Zadeh’s complementation (Zadeh, 1972), to constructs “young1”, “young2” and “old”. While such negation seems to work well with the logical interpretation, it gives somewhat unexpected results with the linguistic interpretation: according to $\text{not}(\mu_{\text{young2}})$, it appears that infants are less “not young” than newborns, which is not correct.

We think that logical and linguistic interpretations complement each other, each of them modeling different aspects of the meaning of the language constructs, and for some words may need to be modeled as separate membership functions. Examples in this paper follow linguistic interpretation (unless mentioned otherwise).

3.4 Choice of constructs, projections and membership functions

The choice of constructs, projections and membership functions in this paper is subjective and serves as an illustration. For the experimental study, please see P. Kapustin and M. Kapustin (2019a).

4 One-dimensional projections

One-dimensional projection is a projection onto one property that allows to model how a language construct relates to this property.

4.1 One-dimensional projections: time references

Here we describe how one-dimensional projections can be used for modeling meaning of words like “after”, “afterwards”, “later”, “until” and “since”. In these examples we choose to focus on the meaning

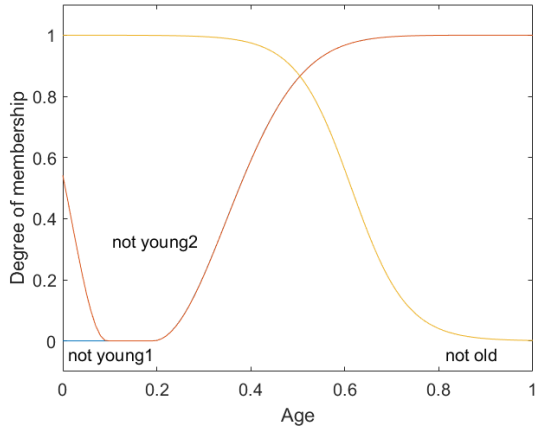


Figure 3: Negation of young2 gives somewhat unexpected results.

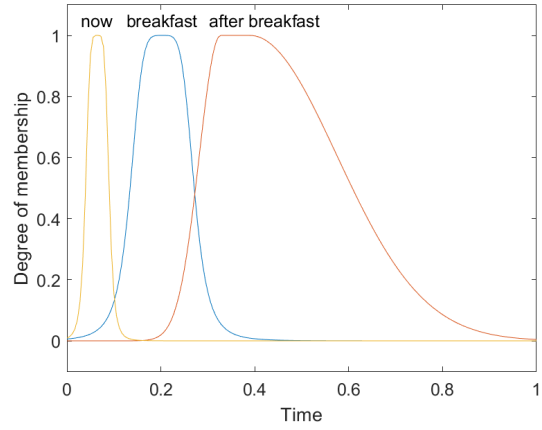


Figure 4: Time reference given by “after breakfast” in “you can play after breakfast”.

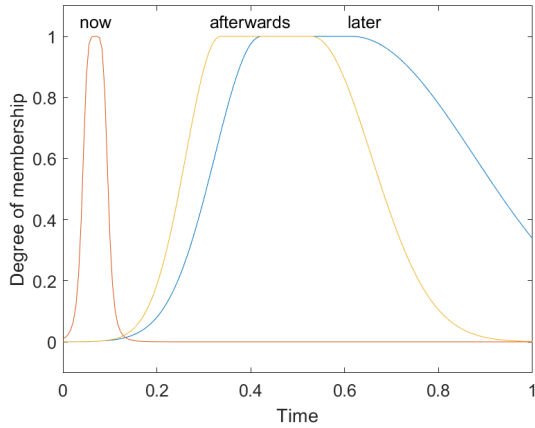


Figure 5: Time references given by “afterwards” and “later” in “let’s discuss this afterwards / later”.

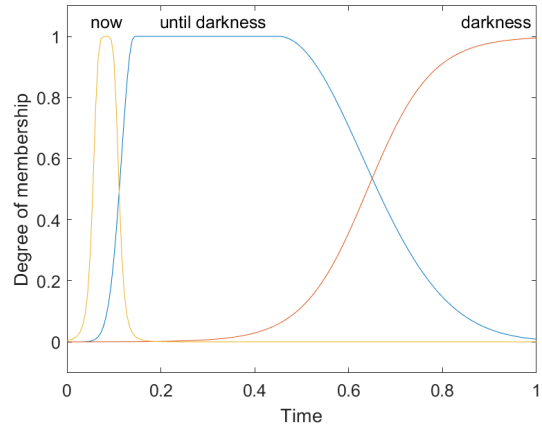


Figure 6: Time references given by “until darkness” in “you can play until darkness”.

aspect of the words that has to do with providing a time reference relative to the time of utterance (given by “now”).

Consider fig. 4. Here we choose to model “after” as suggested by Vocabulary.com (2018a): “happening at a time subsequent to a reference time”, that’s why the membership function for “after breakfast” is decreasing relatively rapidly (this would be different if we chose to model “after” as in “the world has changed after the Second World War”). “Before” may be modeled in a similar way, but we do not include a figure here for brevity.

Consider fig. 5. Here we choose to model “afterwards” as a function that decreases relatively rapidly after a certain point, agreeing with dictionaries mentioning that a certain reference time is usually assumed (Vocabulary.com, 2018b; Cambridge.org, 2018a). On the other hand, “later” is modeled as “at some time in the future” (Vocabulary.com, 2018c; Cambridge.org, 2018b), that’s why the function is decreasing slower, $\mu_{\text{later}} > \mu_{\text{afterwards}}$ in more distant future, and $\mu_{\text{later}}(1) > 0$. This would be different if we chose to model “later” as a synonym for “afterwards” (this meaning of “later” is also suggested by the same dictionaries).

Consider figs. 6 and 7. The fact that the time references given by “darkness” and “summer” are relatively vague is modeled by slow decrease of $\mu_{\text{untilDarkness}}$ and slow increase of $\mu_{\text{sinceSummer}}$.

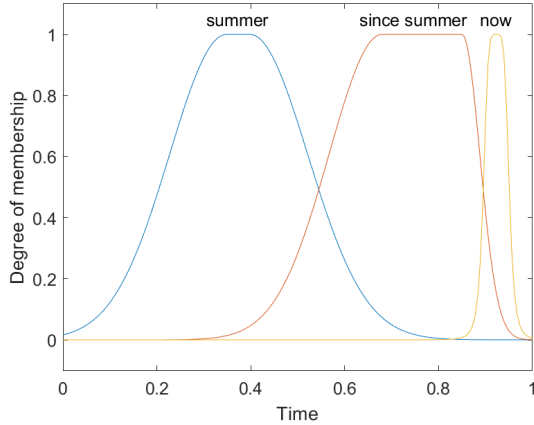


Figure 7: Time reference given by “since summer” in “you have had the book since summer”.

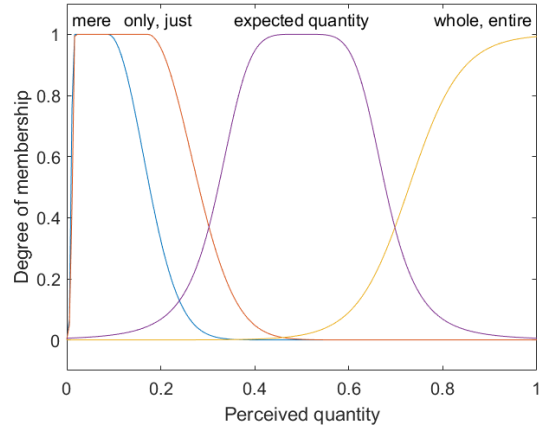


Figure 8: “Mere”, “only”, “just”, “whole”, “entire” related to perceived quantity in “only two days”, “whole room”, “mere one percent”.

4.2 One-dimensional projections: perception of quantities

Consider fig. 8. Here we suggest how one-dimensional projections can be used to model meaning of words like “only”, “just”, “whole”, “entire”, “mere”. In these examples we choose to focus on what these words tell us about certain quantity compared to our expectations (e.g. Zeevat, 2009; Berkeley.edu, 2019). We use name “perceived quantity” for the property.

Here we let $\mu_{\text{whole}}(1) = \mu_{\text{entire}}(1) = 1$ to model the fact that words “whole” and “entire” may be used with something perceived as very large (e.g. “entire universe”). On the other hand, we let $\mu_{\text{mere}}(0) = \mu_{\text{only}}(0) = \mu_{\text{just}}(0) = 0$, because we cannot think of examples when these words are used with zero quantities (e.g. “a mere zero”, “only nothing” and “just no one” sound strange). Also, here we choose to model “mere” as a more specific word than “only” and “just”, as suggested by OxfordDictionaries.com (2019b): “used to emphasize how small or insignificant someone or something is”. Here we do it by letting μ_{mere} cover less area than $\mu_{\text{only, just}}$ on fig. 8.

4.3 One-dimensional projections on related properties: repeating events

Consider fig. 9. Here we are attempting to model what the words “seldom”, “occasionally”, “regularly”, “often” tell us about event frequency (as in “I often play chess”). The words “occasionally” and “regularly” seem to be less specific than the words “seldom” and “often”, and we model this by letting their membership functions cover larger area under the curve.

Consider fig. 10, where we are attempting to model what the words “seldom”, “occasionally”, “regularly”, “usually”, “often” tell us about expectedness of an event (as in “I often play chess when we meet with my friends”)³. We model “regularly” as a more specific word on fig. 10 than on fig. 9, because we believe that “I regularly play chess when I meet with my friends” means a rather high expectedness of the game of chess if the meeting happens (but lower than for “usually” or “often”).⁴ Note that we include “usually” on fig. 10, but not on fig. 9, because it is possible to say “I usually play chess when I meet with my friends”, while “I usually play chess” sounds strange.

In this example words “seldom”, “occasionally”, “regularly”, “often” have two independent projections on related properties: “frequency” and “expectedness”. In general, we think that having multiple independent projections on related properties is interesting, in particular because it may help the systems

³Here by “expectedness” we mean “the quality or state of being expected” (CollinsDictionary.com, 2019).

⁴We discuss how people relate these and other constructs to compatibility intervals (P. Kapustin and M. Kapustin, 2019b), a representation closely related to fuzzy sets, in an experimental study (P. Kapustin and M. Kapustin, 2019a).

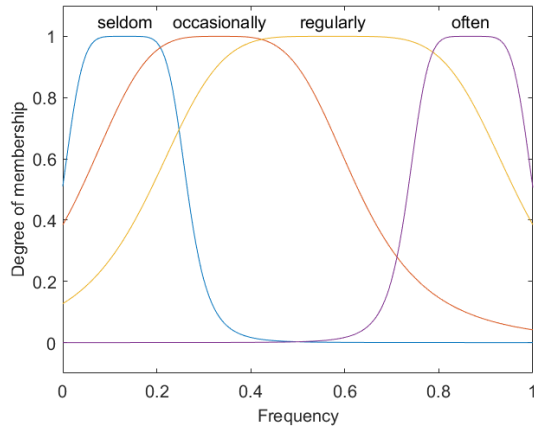


Figure 9: “Seldom”, “occasionally”, “regularly”, “often” related to event frequency.

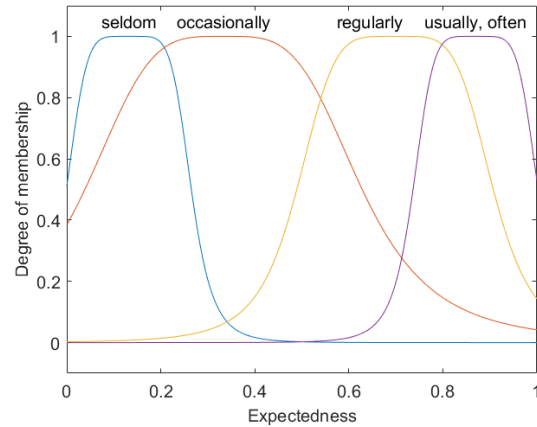


Figure 10: “Seldom”, “occasionally”, “regularly”, “usually”, “often” related to event expectedness.

learn more about the relation between these properties, and needs more research.

5 Membership functions that depend on other functions

5.1 Membership functions that depend on other functions: sufficiency and excess

Consider figs. 11 and 12. Here we are attempting to model what constructs “enough”, “not enough”, and “too much” tell us about the amount of certain property with respect to how much property is desirable/acceptable, modeled with a separate desirability/acceptability function.⁵ We believe that for the construct “not enough” to be meaningful, there should be a place where desirability/acceptability function is increasing (e.g. “not enough air pollution” usually does not make sense). Likewise, the construct “too much” (or too expensive, etc.) only makes sense if there is a place where desirability/acceptability function is decreasing (e.g. “I have too much money” would often require an explanation to answer why having less money would be more desirable).

Here we follow linguistic interpretation for μ_{enough} , modeling the fact that we would normally use words other than “enough”, when the amount of property is much higher than the amount qualifying as “enough”, and that is why μ_{enough} is gradually decreasing after a certain point. We let $\mu_{\text{enough}}(1) > 0$ as “enough” may still be used in such situations (e.g. “he earns enough” may be used about a millionaire when one prefers to be less specific).

It is interesting to note that figs. 11 and 12 present examples when both membership functions and the scale of members function arguments depend on another function (in this case, desirability / acceptability). We believe that such dependencies need further research for such models to become practically useful for problems of computational linguistics and natural language processing.

6 Multi-dimensional projections

Sometimes modeling meaning of certain constructs requires membership functions that take several arguments, when it is the relation of the arguments is what defines the concept. Here we are discussing several examples of this kind.

Consider figs. 13 and 14. Like many other constructs, “already” and “still” have several related meanings with subtle differences. Here we are focusing on modeling surprise at the fact that something happens or will happen earlier or later than expected (e.g. Zeevat, 2009, 2013; Cambridge.org, 2019a,

⁵Similar notion of admissibility is used in Meier (2003).

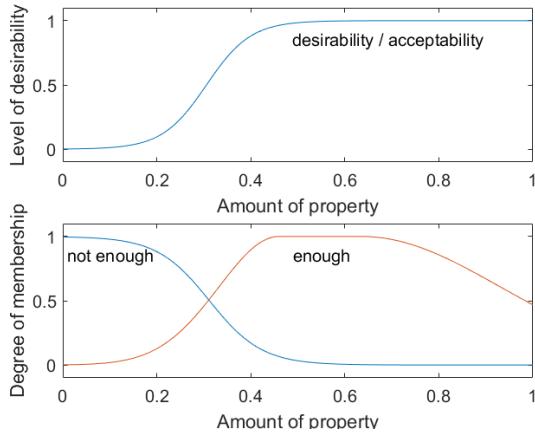


Figure 11: “Enough” and “not enough” related to the amount of property in “enough / not enough for everyone”.

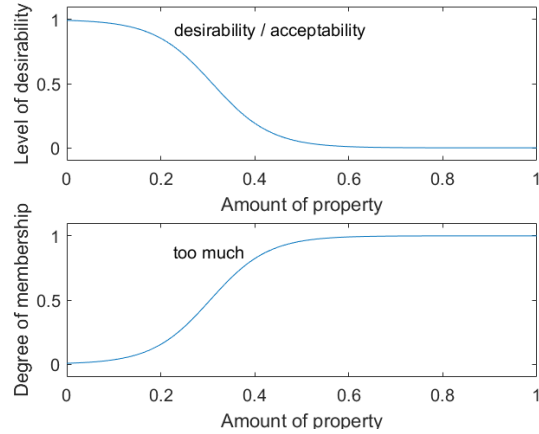


Figure 12: “Too” related to the amount of property in “I think that owning a car is too expensive these days”.

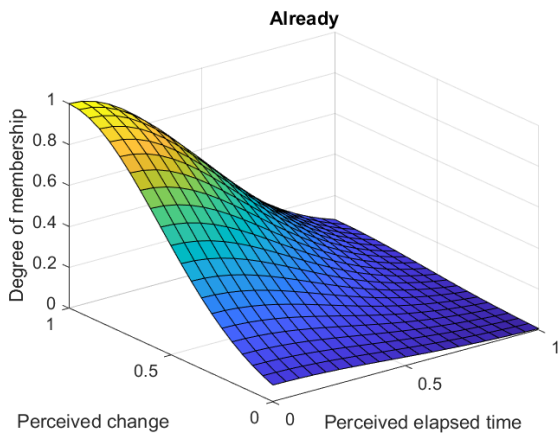


Figure 13: “Already” related to perceived change and perceived elapsed time in “it is already finished”.

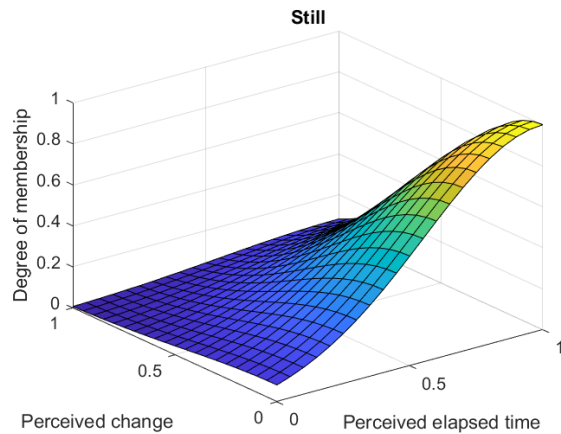


Figure 14: “Still” related to perceived change and perceived elapsed time in “they are still working”.

2019d). We represent these constructs by relating properties “perceived change” and “perceived elapsed time”. “Already” means that perceived elapsed time is relatively low, and perceived change is relatively high, while “still” means the opposite.

Consider fig. 15. Here we model construct “efficient” by relating properties “progress” and “elapsed time”: “efficient” means that elapsed time is relatively low, and progress is relatively high.

Consider fig. 16. Many dictionaries define “lately” as “recently” or “not long ago” (OxfordDictionaries.com, 2019a; Cambridge.org, 2019c; Merriam-Webster.com, 2019). However, Cambridge.org (2019b) explains that “lately” is used for states or repeating events, mostly with present perfect, and is not used for single events. Here we choose to model “lately” in this meaning, as a word that describes recent state of things: when the time is close to zero (further in the past), pretty much all states are compatible with the construct “it rains a lot lately”. In other words, we have no information about the state of things, and this is modeled by membership degree of “lately” being approximately equal to one, as long as time is close to zero. When the time is closer to one (recent past), only the states with high average rainfall are compatible with the construct. It seems that “lately” is sometimes used as a word that contrasts recent situation with earlier situation, however we believe that this can be very context dependent, and choose not to model it here: according to fig. 16, we don’t know how things were in the past.

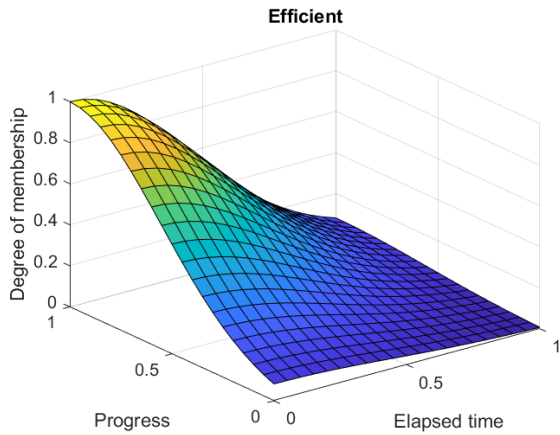


Figure 15: “Efficient” related to progress and time in “this dryer is very efficient”.

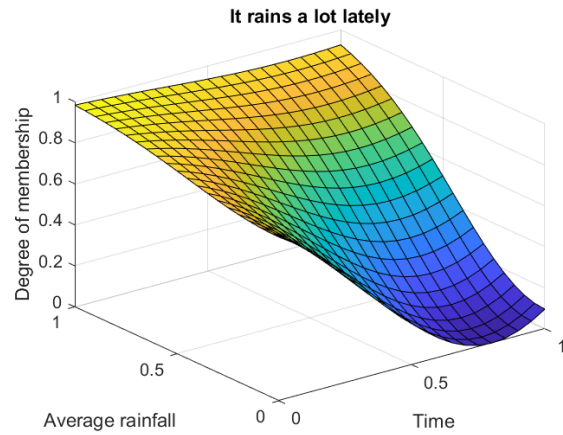


Figure 16: “Lately” related to time and average rainfall in “it rains a lot lately”.

7 Discussion

Although the choice of membership functions used in the examples is subjective, we hope that they are a useful illustration to the approaches and ideas described in the paper, as well as to the importance of distinguishing between logical and linguistic interpretations of membership functions. For the experimental study, please see P. Kapustin and M. Kapustin (2019a).

One can see applications of such models both in natural language understanding and natural language generation. When a system meets a language construct, it can understand it in terms of “underlying” properties, e.g. “often” — in terms of “frequency”, and “already” — in terms of the relation between “perceived change” and “perceived elapsed time”. Similarly, having information about possible values of property or properties, a system can attempt to describe the situation with appropriate words, e.g. information about “progress” and “time” can be described using words like “efficient”.⁶

We think that wider adoption of fuzzy sets in computational linguistics and natural language processing may benefit from the research that will help to make such models easier to learn from data. For example, Runkler (2016) describes an approach for generation of linguistically meaningful membership functions from word vectors. We suggest a meaning representation that is closely related to membership functions, but may be somewhat easier to learn from data (P. Kapustin and M. Kapustin, 2019b).

In many cases, when trying to understand how the membership functions should behave, and even qualitatively compare membership functions for related words, it was not that easy to find linguistic evidence in the literature. In some cases we had a feeling that dictionary definitions left some important parts of the construct meaning unexplained (but it was clear from the examples or explanations found elsewhere). We noticed these things because of our attempts to model meanings of the constructs in a more formal way (in this case using membership functions).

We argue that fuzzy sets and membership functions are useful tools that are interesting both from purely linguistic perspective, and also as a meaning representation for problems of computational linguistics and natural language processing, and hope that more researchers become interested in this area.

Acknowledgements

We thank Vadim Kimmelman and Csaba Veres for helpful discussions and comments. We thank anonymous reviewers for helpful feedback.

⁶We suggest one approach for describing situations in words in M. Kapustin and P. Kapustin (2015).

References

- Berkeley.edu (2019). Frame:RankedExpectation. https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Ranked_expectation.
- Cambridge.org (2018a). Afterwards. <https://dictionary.cambridge.org/dictionary/english/afterwards>.
- Cambridge.org (2018b). Later. <https://dictionary.cambridge.org/dictionary/english/afterwards>.
- Cambridge.org (2019a). Already. <https://dictionary.cambridge.org/grammar/british-grammar/already>.
- Cambridge.org (2019b). Late or lately. <https://dictionary.cambridge.org/grammar/british-grammar/late-or-lately>.
- Cambridge.org (2019c). Lately. <https://dictionary.cambridge.org/dictionary/english/lately>.
- Cambridge.org (2019d). Still. <https://dictionary.cambridge.org/grammar/british-grammar/still>.
- Carvalho, J. P., F. Batista, and L. Coheur (2012). A critical survey on the use of fuzzy sets in speech and natural language processing. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, 1–8. IEEE.
- CollinsDictionary.com (2019). Expectedness. <https://www.collinsdictionary.com/dictionary/english/expectedness>.
- Hersh, H. M., and A. Caramazza (1976). A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General* 105 (3): 254.
- Kapustin, M., and P. Kapustin (2015). Modeling meaning: computational interpreting and understanding of natural language fragments. *arXiv preprint arXiv:1505.08149*.
- Kapustin, P. (2015). Computational comprehension of spatial directions expressed in natural language. Master's thesis, The University of Bergen.
- Kapustin, P., and M. Kapustin (2019a). Language constructs as compatibility intervals: an experimental study. In preparation.
- Kapustin, P., and M. Kapustin (2019b). Modeling meaning of language constructs using compatibility intervals. In submission.
- Meier, C. (2003). The meaning of too, enough, and so... that. *Natural Language Semantics* 11 (1): 69–107.
- Merriam-Webster.com (2019). Lately. <https://www.merriam-webster.com/dictionary/lately>.
- Novák, V. (1992). The alternative mathematical model of linguistic semantics and pragmatics. In *The Alternative Mathematical Model of Linguistic Semantics and Pragmatics*, 87–183. Springer.
- Novák, V. (2008). A comprehensive theory of trichotomous evaluative linguistic expressions. *Fuzzy Sets and Systems* 159 (22): 2939–2969.
- Novák, V. (2017). Fuzzy logic in natural language processing. In *Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on*, 1–6. IEEE.
- Novák, V., I. Perfilieva, A. Dvůrák, et al. (2016). *Insight into Fuzzy Modeling*. John Wiley & Sons.

- OxfordDictionaries.com (2019a). Lately. <https://en.oxforddictionaries.com/definition/lately>.
- OxfordDictionaries.com (2019b). Mere. <https://en.oxforddictionaries.com/definition/mere>.
- Runkler, T. A. (2016). Generation of linguistic membership functions from word vectors. In *Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on*, 993–999. IEEE.
- Vocabulary.com (2018a). After. <https://www.vocabulary.com/dictionary/after>.
- Vocabulary.com (2018b). Afterwards. <https://www.vocabulary.com/dictionary/afterwards>.
- Vocabulary.com (2018c). Later. <https://www.vocabulary.com/dictionary/later>.
- Zadeh, L. A. (1971). Quantitative fuzzy semantics. *Information Sciences* 3 (2): 159–176.
- Zadeh, L. A. (1972). A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges. *Journal of Cybernetics*.
- Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—I. *Information sciences* 8 (3): 199–249.
- Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems* 1 (1): 3–28.
- Zeevat, H. (2009). “Only” as a mirative particle.
- Zeevat, H. (2013). Expressing surprise by particles. In *Beyond Expressives: Explorations in Use-Conditional Meaning*, 297–320. Brill.