Compositionality for perceptual classification

Staffan Larsson
Centre for Linguistic Theory and Studies in Probability (CLASP)
Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg
sl@ling.gu.se

Abstract

We compare three approaches to compositional semantics for words that are modeled (in part) using perceptual classifiers: "meanings as sets", "meanings as transparent functions", and "meanings as opaque functions". The approaches are evaluated according to whether they fulfill a set of desiderata, including dealing with non-intersective compositionality, working with state of the art classifiers, and accounting for learning of perceptual meanings. We find that none of the current approaches fulfill all our desiderata.

1 Introduction

The idea of using classifiers in modeling perceptual meanings has recently received much attention in a variety of research fields, mostly if not exclusively focusing on visual domains. A couple of different overall approaches can be discerned, and in this paper we provide an overview and comparison of these approaches. Among perceptual domains[1] we count both visual and other domains. Among the visual domains, meanings of spatial terms and colour terms have so far received the most attention from the research community. In work on vagueness, the domain of height is often used.

By perceptual words we mean words with perceptual meanings. By perceptual meaning we mean the aspect of meaning of an expression e which enables an agent to decide, based on perceptual information derived from some situation s, whether e holds of (is true of, describes etc.) s. Of course, each individual perceptual domain has its own unique features and needs to be studied separately. However, by regarding perceptual meanings as a more or less uniform phenomenon, we are assuming that useful generalizations can be made across specific perceptual domains.

The general problem of compositionality for perceptual classification can be described as follows: Given an NL expression (phrase or sentence) e that includes $n \geq 2$ perceptual words (or subexpressions) $e_1, \ldots, e_n$, how can an agent decide whether e correctly describes a visual scene s? This minimally requires computing the meaning c of e and classifying a situation s as being described (or not) by c. The different approaches to compositionality for classifiers give different solutions to how this is to be done[2].

2 Desiderata on solution

In this section, we list a couple of possible desiderata on solutions to the problem of compositionality for perceptual meanings. For each feature, each approach will be marked with "+", "-" or "?". A "+" means that at least a proof of concept solution addressing the desideratum in question exists, and "-" means that no proof of concept exists. Unclear cases are marked "?".

A solution to the problem of compositionality for perceptual classification should...

[1] We use this term to mean something more specific than "sensory modality" (which include visual, auditory, tactile, olfactory, and taste modalities).

[2] Note that we are limiting our analysis to the modeling of perceptual words using classifiers, excluding e.g. vector space analyses of word meanings. While there may be some affinities, exploring these is left for future work

...handle intersective compositionality [COM]. Most work on compositionality for classifiers focuses on intersective compositionality, where a classifier c is composed from two classifiers c_1 and c_2 by saying that c classifies s positively to the extent that both c_1 and c_2 classify s positively. For example, we may assume that "light green" can be interpreted as "light and green", or that "upper right" means "upper and right".

...handle non-intersective compositionality [NON]. There seem to be cases where intersective compositionality does not work, and this is an area where the different approaches differentiate in interesting ways, as will be shown below. For example, "sort of green" probably does not mean "sort of and green" in the way that "light green" can be (possibly) modeled as meaning "light and green"[3].

...account for learning of perceptual meanings [LEA]. This amounts to updating classifiers based on sensory observations of visual scenes and associated linguistic descriptions. This can be done in different ways, e.g. from corpora or from interaction with humans.

...account for vagueness [VAG]. Given that there will always be some variation in human judgement (McMahan and Stone, 2015), one may also want a solution to account for (replicate or at least explain) the observed variability. This is especially relevant for vague judgements (which may possibly include all perceptual judgements)[4]

...work with state of the art classifiers [SOA]. There are many approaches to visual classification, and recently deep learning approaches have made great strides. It is of course an advantage if an account of compositionality for visual classifiers can benefit from these advances, and therefore it is desirable that the account is neutral to the type of classifier, as far as possible.

...connect perceptual meanings to other semantic phenomena [SEM]. For example, inference, quantification and modality. Being able to reason based on observations, bringing together perceptual and linguistic information, would presumably be useful in e.g. robotics. Reasoning about modality, e.g. distinguishing necessary, possible, impossible and actual states of affairs, can also be expected to be useful in many AI applications.

3 Existing approaches to compositionality for visual classifiers

In this section, we review and compare a few different approaches to the use of classifiers for judging whether a sentence holds of an observed situation.

3.1 Meanings as sets

In work on formal semantics in the context of perception in robots, several researchers (Matuszek et al., 2012; Krishnamurthy and Kollar, 2013) have used classifiers in conjunction with NL semantics based on first order logic (FOL) in the Possible Worlds Semantics (PWS) tradition (Montague, 1974).

The idea here is to apply all classifiers to all objects in the scene, producing a first order model where meanings of predicates are sets of referents (or n-tuples of referents in the case of n-place relations). This makes it possible to apply standard FOL compositional semantics to sentences, and to evaluate these using standard model-theoretic semantics.

[3] In general, it appears that (at least) adverbs that modify (perceptual) adjectives cannot be modeled using intersective compositionality.

[4] Vague words are often also context sensitive, so that e.g. a tall basketball player is generally taken to be taller than a tall person. However, the two phenomena are in principle independent. While accounting for context-sensitivity is a possible desideratum on accounts of perceptual semantics, we have not included it here.

The main advantage of this approach is that it is pretty straightforward and uses standard machinery of formal semantics, which means that inference, quantification etc. are taken care of **[SEM+]**. Note, however, that on this approach predicates are not evaluated against objective referents in a world (encompassing everything there is), but against mental and subjective representations of observed and classified referents in a specific situation (typically encompassing only a limited part of the world). This seems to depart from foundational assumptions of PWS and may possibly put into question existing accounts of e.g. quantification and modality. Another advantage is that there are no limits on the kinds of classifier that can be used **[SOA+]**.

Existing accounts of compositionality rely exclusively on set intersection. For example, Krishnamurthy and Kollar (2013) analyze "blue mug on table" as the intersection of the sets [blue], [mug] and [on-table][5]**[COM+]**. A problem for this approach, however, is non-intersective compositionality. Intersective compositionality will not work when a phrase containing more than one perceptual predicate cannot be analysed as a conjunction, and (as far as we know) there is currently no set-theoretic analysis of non-intersective compositionality **[NON-]**.

There is some work on learning and vagueness in the PWS tradition(Barker, 2002) which appears to be compatible with the meanings-as-sets approach and that may be applicable to perceptual semantics, although this remains to be shown **[LEA?, VAG?]**.

3.2 Meanings as transparent functions

In general, a parameterised function takes an input (domain) and a set of parameters, and yields an output (range). We distinguish *transparent* and *opaque* functions as follows. A transparent function is a parameterised function where the input and parameters have clear interpretations understandable to humans, and where the effects on the output of manipulating the parameters are predictable. For example, a simple threshold classifier (parameter: threshold) or a noisy threshold classifier (parameters: threshold, standard deviation). An *opaque function* is a parameterised function where the parameters do *not* have clear interpretations understandable to humans. Examples include most neural networks, including deep neural nets, and probability distributions collected from observations[6]. Note that the distinction is slightly vague and that there are borderline cases. For example, a n-input neuron with a threshold implements to a transparent linear classifier function in n-dimensional space.

Gapp (1994) models spatial concepts using cubic spline functions[7], essentially parameterized (transparent) functions mapping points in 3D space onto "degrees of applicability" in the interval [0..1][8]. These functions are transparent in the sense that there is an observable and understandable connection between the function parameters and the resulting distribution of degrees of applicability across the spatial domain. McMahan and Stone (2015) model meanings of colour words as probability distributions over colour spaces. These distributions are represented by probabilistic thresholds derived from a corpus of colour descriptions. Transparent functions are also used for modeling perceptual concepts in Larsson (2011) (using "(to the) right" as a proof of concept) and Larsson (2013) (adding compositionality, exemplified by "far right" and "upper right") and incorporating vagueness (Fernández and Larsson, 2014) ("tall"). Here, the meaning of "(to the) right" is modeled as a two-input perceptron, which is transparent in the sense that it the connection between the parameters and the resulting classification results is understandable (with some effort). Specifically, the weights on the two inputs modify the slant of the line dividing "(to the) right" from "not (to the) right", while the threshold modifies the distance of this line from the centre.

On the approach taken in Larsson (2013), high-level logical aspects of meaning are represented together with low-level aspects (classifiers etc.) in a single representational system (Type Theory with

[5]We are here simplifying somewhat by not seeing "table" as a separate predicate, as done in the original.

[6]Note that if a probability distribution derived from observations is analysed using some kind of curve-fitting, it may become transparent.

[7]Cubic splines are constructed of piecewise third-order polynomials which pass through a set of control points (`http://mathworld.wolfram.com/CubicSpline.html`)

[8]Note that degrees of applicability do not form a probability distribution.

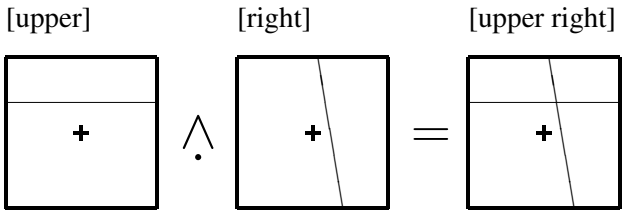

Figure 1: Intersective compositionality for "upper right" on the transparent functions approach

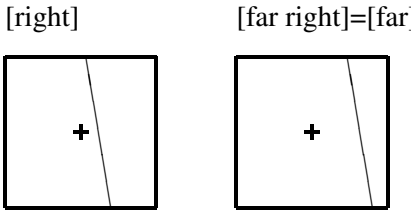

Figure 2: Non-intersective compositionality for "far right" on the transparent functions approach

Records or TTR (Cooper, 2012)). A wide variety of semantic phenomena have been described using TTR including modeling of intensionality and mental attitudes (Cooper, 2005), dynamic generalised quantifiers (Cooper, 2004), co-predication and dot types in lexical innovation, frame semantics for temporal reasoning, reasoning in hypothetical contexts (Cooper, 2011) and enthymematic reasoning (Breitholtz and Cooper, 2011) **[SEM+]**.

On the transparent function approach, intersective compositionality can be handled in various ways. Gapp (1994) analyses "between X and Y" as meaning (roughly) "in front of X and in front of Y, with X and Y facing each other", with the applicability score for a point being "between" an two facing objects X and Y computed as the arithmetic mean of the scores for p for "in-front-of X" and "in-front-of Y". As a proof of concept of intersective compositionality for transparent classifiers, Larsson (2013) shows how to compute the intersective meaning of "upper right" from the meanings of "upper" and "right" (see Figure 1). This is done by simply conjoining the classifiers for "upper" and "right" so that a point (in 2D space) is to the upper right iff it is "upper" and "(to the) right" **[COM+]**.

Larsson (2013) also provides a proof of concept of non-intersective compositionality for degree modifiers, accounting for "far" in "far right". The proposal is that "far" takes parameters of the "right" classifier and yields modified classifier for "far rightness" with an increased threshold (see 2). Together, these two examples indicate how both intersective and non-intersective can be handled in this version of the transparent functions approach **[NON+]**.

In Fernández and Larsson (2014), the above analysis is supplemented with an account of vagueness, using probabilistic TTR (Cooper et al., 2015). In line with Lassiter (2011), both Fernández and Larsson (2014) and McMahan and Stone (2015) uses a noisy threshold, and in Fernández and Larsson (2014) the meaning of "tall" is modeled as a transparent function parameterised by mean and standard deviation **[VAG+]**.

In Larsson (2013), learning is accounted for by modifying parameters of classifiers based on observed judgements (e.g. by a teacher in the case of supervised learning). One could also imagine how extending the corpus underlying the probabilistic thresholds in McMahan and Stone (2015) could result in updated thresholds, thus providing the basis for a learning mechanism. **[LEA+]**

On the downside, on this approach probability distributions are computed from transparent functions, which excludes deep learning, thus forcing the use of lower quality classifiers compared to the state of the art **[SOA-]**.

3.3 Meanings as opaque functions

As mentioned above, opaque functions are functions whose parameters (if any) are understandable to human interpreters. We believe that a couple of related approaches to meanings as classifiers can be seen as examples of this overall approach.

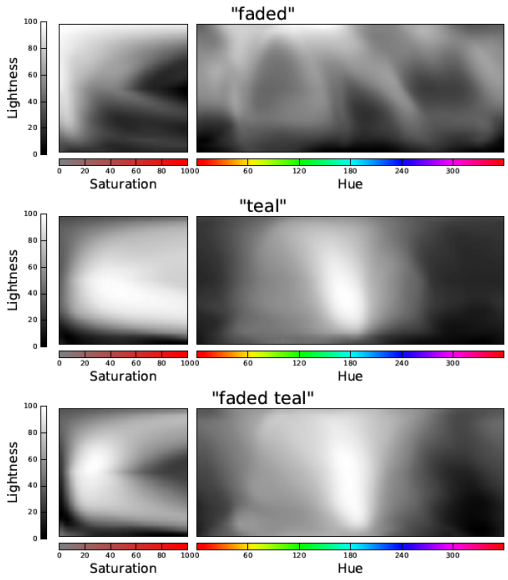

Figure 3: Intersective compositionality for ”faded teal” on the opaque functions approach. Picture from Monroe et al. (2016).

Monroe et al. (2016) build on McMahan and Stone (2015) but, instead of explicitly computing thresholds, encode probability distributions implicitly in a recurrent neural network sequence decoder. Feeding a word sequence and a colour sample to this network yields the probability of the sequence as describing the colour sample, equal to the product of probabilities of each successive word in the sentence conditioned on the colour sample input and the preceding words. On this approach, classification is done by getting the probability for the point in conceptual space corresponding to the situation. This is similar to the approach of Logan and Sadler (1996), although the latter model ”degrees of goodness” rather than probabilities. Monroe et al. (2016) includes an account of intersective compositionality for dealing with descriptions not found in the training set. If one ignores the word sequence probabilities, Monroe et al. (2016) can be interpreted as regarding intersective composition as equivalent to multiplying together probability distribution matrices (see Figure 3)[9]. Schlangen et al. (2016) model classifiers as neural networks taking feature representations of objects and returning ”appropriateness score”; these classifiers presumably correspond (for visual space) to something like the ”degrees of goodness” of Logan and Sadler. Intersective composition is done not on the classifier level (the matrices) but on the combined classifier outputs for an object described by words $w_1, \ldots, w_n$, by simple multiplicative composition (for nominal constructions). Åstbom (2017) shows that geometric mean provides a better analysis than arithmetic mean when composing a matrix for ”upper and to the right” from matrices for ”upper” and ”to the right” **[COM+]**.

Regarding non-intersective modifiers, the question is to what extent they can be represented as distributions of appropriateness scores or probabilities in perceptual spaces, and thus handled by the regular mechanisms of intersective compositionality. We have not been able to find any account of non-intersective compositionality on the opaque functions approach **[NON-]**.

Perhaps the main advantage of this approach is that since it does not matter how probability distributions are encoded, deep learning can be used to yield high quality classifiers **[SOA+]**. Also, the distributions can be derived from observed judgements, possibly amended with a smoothing function **[LEA+]**. Furthermore, provided that vagueness can be modeled as probability distributions, this approach is well suited to model vagueness **[VAG+]**.

[9]However, the output probability for a word in a multi-word sequence will also be conditioned on the previous words, which enables this approach at least in principle to learn a separate meaning for ”teal” when it follows ”faded”, provided ”faded teal” is the training corpus. Furthermore, since the output of the hidden layer of each node in an RNN is fed forward into the next node, they can also model more complex phenomena such as long distance dependencies.

Meanings as...	INT	NON	LEA	VAG	SOA	SEM
... sets	+	-	?	?	+	+
... transparent functions	+	+	+	+	-	+
... opaque functions	+	-	+	+	+	?

Table 1: How the different approaches satisfy the desiderata (INT=intersective compositionality, NON=non-intersective compositionality, LEA=learning perceptual meanings, VAG=accounting for vagueness, SOA=work with state of the art classifiers, SEM=connect to other semantic phenomena

Since this approach has not so far been connected with work on formal semantics, it is unclear to what extent it connects to other semantic phenomena. There is work on inference in deep neural networks (Kumar et al., 2015) but as of yet, many semantic phenomena (such as quantification, modality, intensional contexts, and co-predication) remain to be captured in terms of this approach **[SEM?]**. However, there seems to be no principled reason why one could not connect opaque function classifiers to TTR in the same way as has been done for transparent function classifiers (see above), yielding probabilistic type judgments.

4 Discussion

Table 1 shows an overview of our conclusions. The lack of an account of non-intersective compositionality is, on our view, a serious shortcoming of the "meanings as sets" approach. The main drawback of the "transparent functions" approach is that it does not work with state of the art classifiers, such as deep neural nets. The reason for this is of course that such classifiers are not transparent functions. Finally, insofar as compositionality in the "opaque functions" approach is limited to intersective compositionality, this means that none of the current approaches fulfill all our desiderata. The question is then if any of the approaches can be improved to satisfy all the desiderata, or if some kind of hybrid approach is needed. We hope to address this question in future work.

Acknowledgements

The author would like to thank to the anonymous IWCS reviewers whose perceptive comments helped improve this paper.

References

Åstbom, A. (2017). How function of objects affects geometry of spatial descriptions. a study of swedish and japanese. bachelors thesis.

Barker, C. (2002). The Dynamics of Vagueness. *Linguistics and Philosophy 25*(1), 1–36.

Breitholtz, E. and R. Cooper (2011). Enthymemes as rhetorical resources. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011)*, Los Angeles (USA), pp. 149–157.

Cooper, R. (2004). Dynamic generalised quantifiers and hypothetical contexts. In *Ursus Philosophicus, a festschrift for Björn Haglund*. Department of Philosophy, University of Gothenburg.

Cooper, R. (2005). Austinian truth, attitudes and type theory. *Research on Language and Computation 3*, 333–362.

Cooper, R. (2011). Copredication, quantification and frames. In S. Pogodalla and J.-P. Prost (Eds.), *LACL*, Volume 6736 of *Lecture Notes in Computer Science*, pp. 64–79. Springer.

Cooper, R. (2012). Type theory and semantics in flux. In R. Kempson, N. Asher, and T. Fernando (Eds.), *Handbook of the Philosophy of Science*, Volume 14: Philosophy of Linguistics. Elsevier BV. General editors: Dov M. Gabbay, Paul Thagard and John Woods.

Cooper, R., S. Dobnik, S. Larsson, and S. Lappin (2015). Probabilistic type theory and natural language semantics. *LiLT (Linguistic Issues in Language Technology) 10*.

Fernández, R. and S. Larsson (2014). Vagueness and learning: A type-theoretic approach. In *Proceedings of the 3rd Joint Conference on Lexical and Computational Semantics (∗SEM 2014)*.

Gapp, K.-P. (1994). Basic meanings of spatial relations: computation and evaluation in 3d space. In *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence*, pp. 1393–1398. AAAI Press.

Krishnamurthy, J. and T. Kollar (2013). Jointly Learning to Parse and Perceive : Connecting Natural Language to the Physical World. *Transactions of the Association for Computational Linguistics 1*, 193–206.

Kumar, A., O. Irsoy, J. Su, J. Bradbury, R. English, B. Pierce, P. Ondruska, I. Gulrajani, and R. Socher (2015). Ask me anything: Dynamic memory networks for natural language processing. *CoRR, abs/1506.07285*.

Larsson, S. (2011). The ttr perceptron: Dynamic perceptual meanings and semantic coordination. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011)*, Los Angeles (USA).

Larsson, S. (2013). Formal semantics for perceptual classification. *Journal of Logic and Computation*.

Lassiter, D. (2011). Vagueness as probabilistic linguistic knowledge. In R. Nowen, R. van Rooij, U. Sauerland, and H. C. Schmitz (Eds.), *Vagueness in Communication*. Springer.

Logan, G. D. and D. D. Sadler (1996). A computational analysis of the apprehension of spatial relations.

Matuszek, C., N. FitzGerald, and L. Zettlemoyer (2012). A joint model of language and perception for grounded attribute learning. In *Proceedings of the 29th International Conference on Machine Learning*.

McMahan, B. and M. Stone (2015). A bayesian model of grounded color semantics. *Transactions of the Association for Computational Linguistics 3*, 103–115.

Monroe, W., N. D. Goodman, and C. Potts (2016). Learning to generate compositional color descriptions. *arXiv preprint arXiv:1606.03821*.

Montague, R. (1974). *Formal Philosophy: Selected Papers of Richard Montague*. New Haven: Yale University Press. ed. and with an introduction by Richmond H. Thomason.

Schlangen, D., S. Zarrie, and C. Kennington (2016). Resolving References to Objects in Photographs using the Words-As-Classifiers Model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*.