Extensions are Indeterminate if Intensions are Classifiers

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

Abstract

In this paper, we explore some consequences of the idea of using classifiers to model intensions of natural language expressions for the notion of extensional meaning, i.e. the idea that the meaning of a word can be modelled as the set of referents in the world (or in a possible world). The upshot is that at least for words referring to observable situations, extensional meaning is derivative of intensional meaning, and modelling meanings as sets of referents is of limited applicability.

1 Introduction

In formal semantics in the Montague tradition (Montague, 1974), which we might refer to as Possible Worlds Semantics (PWS), the meaning of a word such as "dog" is taken to be its extension, i.e. the set of all dogs in the world (or in a possible world). As a research program, PWS has in many ways been quite successful. Within the research area of formal semantics, PWS is still the dominant theoretical framework, and forms the theoretical foundation for a majority of the research published in the major journals and conferences in the field.

Over the last decade or so, work in several neighbouring areas have, in various ways and more or less explicitly, used *classifiers* to model *intensional* meanings. Putting it more precisely, in this work *perceptual meaning* has been modelled intensionally using classifiers. Perceptual meaning is an important aspect of the meaning of linguistic expressions referring to physical objects (such as concrete nouns or noun phrases).

Knowing the perceptual meaning of an expression allows an agent to identify perceived objects and situations falling under the meaning of the expression. For example, knowing the perceptual meaning of "blue" would allow an agent to correctly identify blue objects. Similarly, an agent which is able to compute the perceptual meaning of "a boy hugs a dog" will be able to correctly classify situations where a boy hugs a dog.

In this paper, we will explore some consequences of the idea of using classifiers to model intensions of natural language expressions for the notion of extensional meaning. We first discuss the role of classifiers in natural language processing, in formal semantics, and in TTR, a Type Theory with Records (Cooper, 2012, in progress). We then briefly discuss semantic coordination before moving on to a comparison with related ideas put forward in the framework of possible world semantics. The next section comprises the central part of this paper, namely a three-pronged argument as to why extensional semantics are not well suited for modelling perceptual meanings, especially if the latter are regarded as classifiers, and hence why extensional semantics does not (and indeed cannot) constitute a general account of natural language meaning. We then discuss cases where nevertheless extensional semantics may be adequate, also throwing some light on why extensional semantics may initially appear to be an intuitive theory of natural language meaning. Finally, we briefly discuss the applicability of the argument to non-concrete words, before providing a summary and conclusions.

2 Classifiers and natural language

The usefulness of *classifiers* (Rosenblatt, 1958; Harnad, 1990) in modeling natural language meaning has in recent years found renewed interest, fuelled by progress in image recognition, image captioning and Visual Question Answering using deep learning methods (Antol et al., 2015; You et al., 2016; Monroe et al., 2016). Many models for tackling these tasks rely explicitly or implicitly on the idea of modelling meanings of perceptual words (i.e. words describing perceivable properties of ob-

jects and situations) as classifiers. In the abstract, one can define classifiers as functions that correspond to a natural language phrase or sentence, and that take perceptual input (e.g. from a camera) and decide to what degree the phrase or sentence describes the perceptual input.

$$\textit{Classifier} : \textit{PerceptualData} \rightarrow [0, 1]$$

If the possible degrees are limited to 0 and 1, we have a binary classifier; if it can be any real number between 0 and 1, we have a continuous (e.g. probabilistic) classifier. Following Schlangen et al. (2016), we refer to the idea of modelling linguistic meanings using classifiers as the *words-as-classifiers* approach.

Several approaches have been proposed for modelling meanings of (some) words as classifiers of perceptual (often visual) data (Dobnik, 2009; Schlangen et al., 2016; Monroe et al., 2016; McMahan and Stone, 2015; Larsson, 2015; Fernández and Larsson, 2014; Schlangen et al., 2016; Larsson, 2017). Various types of classifiers have been used, and the output of the classifiers has been rendered and connected to language in various ways (see Larsson (2017) for an overview). This approach has also been used in work on visual question answering (Andreas et al., 2016).

3 Classifiers and formal semantics

Marconi (1997) distinguishes inferential and referential meaning. *Inferential* word meanings enable inferences from uses of the word. Such meanings are sometimes referred to as "high level" or "symbolic", and are typically modelled in formal semantics. *Referential* meaning, on the other hand, allows speakers to identify objects and situations referred to. Referential meaning is sometimes referred to as "low-level" or "subsymbolic". Our working hypothesis is that referential meaning can be modelled using classifiers that output formal representations, thus connecting "high level" formal representations to "low level" perceptual information, and in this we follow Larsson (2011) and Larsson (2015). This is a way of addressing the *symbol grounding problem* put forward by Harnad (1990) in a way that is compatible with formal semantics.

The crucial step in making use of classifiers in formal semantics is to regard them as (parts of) *representations of intensions* of linguistic expressions (Larsson, 2015). Traditionally, the intension of an expression helps determine whether some item belongs to the extension of the expression. Here, this translates to using a classifier to help determine whether some perceptual data derived from some item can be used to classify that item as falling under the expression, i.e., to be included in its extension.

The idea of regarding classifiers as representing intensions is related to proposals by Muskens (2005) and Lappin (2012) to identify the intensions of an expression with an algorithm (implemented using logic programming or a functional programming language, respectively) for determining its extension. The idea of representing referential meanings as classifiers can perhaps be regarded as an application of this general idea to (although we are here using general notation for functions rather than any specific programming language).

4 Classifiers and TTR

Larsson (2020), following Cooper (2019), presents a version of Type Theory with Records which places classifiers at the core of formal semantics, and shows the role of classifiers in deciding whether (or to what extent) an utterance content (an utterance meaning interpreted in context) correctly describes a perceptually available situation (such as a visual scene). See Appendix A for a brief introduction to TTR. The technical details of Larsson (2020) are not important for our present purposes, but are included to make more concrete what we mean when we talk about modelling intensions as classifiers.

The core definition linking meaning to classifiers is

(1) $s : T$ iff $\textit{Clfr}(T)(s) = T$

where s is a situation being classified, T is a ptype (representing the content of an utterance), and *Clfr* is a classifier function associated with T.

As an illustration, Larsson (2011, 2015, 2020) uses a simple dialogue game called the left-or-right (LoR) game. In this game, one agent (A) places objects on a square surface, and the other agent (B) classifies these objects as being to the right or not. In first language acquisition, training of perceptual meanings typically takes place in situations where the referent is in the shared focus of attention and thus perceivable to the dialogue participants. It is assumed that the dialogue participants are able to establish a shared focus of

attention. A (simple) sensor collects some information (sensor input) from the environment and emits a real-valued vector. The sensor is assumed to be oriented towards the object in shared focus of attention.

Larsson formalises the notion of a simple perceptron classifier and provide its TTR type. Whereas a (non probabilistic) classifier normally gives a Boolean output (corresponding to whether the neuron triggers or not), we want as output a ptype (or the negation thereof):

(2) $\pi_{right} : \begin{bmatrix} \text{w}:\mathbb{R}^+ \\ \text{t} :\mathbb{R} \end{bmatrix} \rightarrow \begin{bmatrix} \text{foo}:\mathit{Ind} \\ \text{sr} :\mathbb{R}^+ \end{bmatrix} \rightarrow \mathit{Type}$

such that if

- par: $\begin{bmatrix} \text{w} & : & \mathbb{R}^+ \\ \text{t} & : & \mathbb{R} \end{bmatrix}$ and
- $r : \begin{bmatrix} \text{foo} & : & \mathit{Ind} \\ \text{sr} & : & \mathbb{R}^+ \end{bmatrix}$,

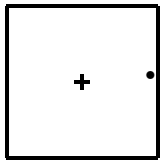
then $\pi_{right}(par, r)$ =

$$\begin{cases} \text{right}(r.\text{foo}) & \text{if } r.sr \cdot par.\text{w} > par.\text{t} \\ \neg\ \text{right}(r.\text{foo}) & \text{otherwise} \end{cases}$$

Here, *par* is a record containing classifier parameters; for a perceptron, a weight vector and a threshold. The second argument to the classifier is the situation to classify, which needs to be of a type specifying a sensor reading (sr) and an object in the focus of attention (foo). Note that the function itself is defined outside TTR. This allows any classifier to used with TTR, no matter how it is implemented.

Assume that an agent A places an object on the surface and says "That one is to the right", or just "Right".

(3)



Agent B watches and gets a position sensor reading $\begin{bmatrix} 0.900 & 0.100 \end{bmatrix}$ which is part of B's take on the current situation (s_1):

B now interprets A's utterance in the context the situation s_1 by computing $[\![\text{right}]\!](s_1)$, which gives the result $[\![\text{right}]\!](s_1) = \text{right}(\mathit{obj}_{45})$.

(4) $s_1 = \begin{bmatrix} \text{sr}_{\text{pos}} & = & \begin{bmatrix} 0.900 & 0.100 \end{bmatrix} \\ \text{foo} & = & \mathit{obj}_{45} \end{bmatrix}$

Next, B decides if A's utterance correctly describes (her take on) the situation, i.e. if

(5) $s_1 : [\![\text{right}]\!](s_1)$, i.e., if $s_1 : \text{right}(\mathit{obj}_{45})$

For

- T=right(obj_{45}),
- Clfr(right(obj_{45}))=π_{right}, and
- par=$\begin{bmatrix} \text{w} = \begin{bmatrix} 0.800 & 0.010 \end{bmatrix} \\ \text{t} = 0.090 \end{bmatrix}$,

we get

(6) $s_1 : \text{right}(\mathit{obj}_{45})$ iff $\pi_{\text{right}}(par)(s)$= right($\mathit{obj}_{45}$)

As shown in the partial derivation in Figure 1 (and in more detail in Larsson (2020)), the RH of this equation holds, and hence B can conclude that s_1 is indeed appropriately described by A's utterance.

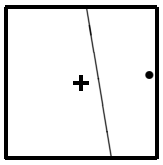
5 Semantic coordination and classifiers

In Larsson (2011) and Larsson (2015), the idea of intensions as classifiers is combined with a notion of semantic coordination — the process of interactively agreeing on the meanings of words and expressions, and (simultaneously) agreeing on which words are appropriate in a given context. Shared meanings (modelled as intensions) are achieved by agents interactively coordinating their respective takes on those meanings, which involves training classifiers based on input from dialogue interaction.

The process of semantic coordination displays a fundamental dialectic between extensional and intensional meaning, in the following way: An individual or situation s is claimed (explicitly or implicitly) by a dialogue participant to be in the extension of (i.e., to be adequately referred to by) an expression e. As a result, the other dialogue participant updates her intensional meaning (classifier) of e based on s. Later, the learner will apply this intensional meaning (classifier) to new situations to determine whether they are in the extension of e.

As an example, we may imagine two rounds of the left-or-right game playing out as in (7).

(7) A: "(the object is to the) right"

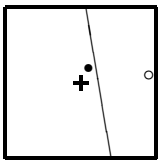


$$\pi_{\text{right}}(par)(s) = \text{right}(obj_{45}) = \pi_{\text{right}}(\begin{bmatrix}\text{w} = \begin{bmatrix}0.800 & 0.010\end{bmatrix} \\ \text{t} = 0.090\end{bmatrix})(\begin{bmatrix}\text{sr}_{\text{pos}} = \begin{bmatrix}0.900 & 0.100\end{bmatrix} \\ \text{foo} = obj_{45}\end{bmatrix}) =$$

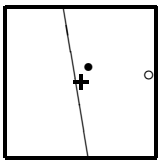
$$(\begin{cases}\text{right}(obj_{45}) & \text{if } \begin{bmatrix}0.900 & 0.100\end{bmatrix} \cdot \begin{bmatrix}0.800 & 0.010\end{bmatrix} > 0.090 \\ \neg\,\text{right}(obj_{45}) & \text{otherwise}\end{cases}) = \text{right}(obj_{45})$$

Figure 1: Example classification derivation

B: "okay"
A: "(the object is to the) right"



B: "aha"



In round 1, B comes to the conclusion that A's utterance matches B's own classification (with the threshold between "left" and "right" shown as a diagonal line). However, in the second round, B's initial classification does not match A's utterance. We imagine that as a result of this, B retrains his classifier that represents the perceptual meaning of "(the object is to the) right" for B. Concretely, this amounts to applying the perceptron training rule to obtain an updated parameterisation (see Larsson (2015) for details) of the classifier, thus updating B's take on the meaning of "right" in the game.

We have now seen a simple example of how perceptual meaning, modelled as a classifier, can be learnt from conversational interaction in a shared perceptual environment. When humans interact they *reciprocally* (Fernández et al., 2011) adapt to each others' language use on multiple levels. Although the left-or-right game models one-sided learning rather than coordination, it could quite easily be altered to illustrate coordination directly. For example, by letting A and B switch roles after each round. In this symmetric left-or-right game, the agents would converge on a meaning of "right" that neither of them may subscribe to initially.

6 Classifiers and Possible Worlds

This view of the nature of linguistic meaning differs in several important respects from the standard view of possible worlds semantics (PWS, as put forward in Montague (1974) and many others):

- Intensions are represented independently from extensions (in the form of classifiers).
- Extensions are *derived* indirectly from intensions by applying classifiers to (takes on) situations, rather than being directly *given*.
- Intensions are *dynamic* and can be revised as a result of interaction (semantic coordination).

We also take it that as a result of semantic coordination, meanings are *intersubjective*, i.e. shared between communities or dyads of speakers, rather than objective in the sense of being independent of individual agents (similar to mathematical concepts).

In Barker (2002) and Lassiter (2011), dynamic extensions are achieved by a notion of possible languages (parallel to the notion of possible worlds), each specifying the extensions of linguistic expressions. By excluding possible languages, one modifies the range of possible extensions and gets successively closer to the actual language. As many others, Barker takes extensions to be at least in part subjective (or else it would not be possible to learn new things about extensions from other agents).

We can perhaps briefly map out a couple of different positions one may take with respect to the nature of linguistic intensions and extensions (see Figure 2. Firstly, extensions may either be considered as static or as dynamic, i.e. up for revision. Second, extensions can be considered as directly given, or as (somehow) indirectly derived from intensions, e.g. by applying intensions (modelled as classifiers) to (takes on) situations or the world. Third, extensions can be regarded as objective (agent-independent) or subjective (agent-dependent). In the subjective category we include intersubjective extensions, i.e. extensions shared among a collective of agents (but not independent of those agents).

7 Indeterminate extensions

We may now ask, if the extensions of at least some natural language expressions are regarded as dy-

Extensions	static/dynamic	given/derived	objective/(inter)subjective
PWS (Montague)	static	given	objective
PWS (Barker)	dynamic	given	(inter)subjective(?)
TTR	dynamic	derived	(inter)subjective

Figure 2: How three different approaches to meaning treat extensions

namic and derived from intensions modelled as classifiers, are they still *determinate* in the sense that *there is at any given time a single extension for each expression* in a language? This, after all, is one of the core assumptions of standard possible worlds semantics, where meanings are modelled as extensions in a model. We would argue that they are not, based on three related arguments. First, however, we want to establish a few more or less trivial facts about classification and coordination.

7.1 Some observations about classification and coordination

First of all, classifiers **generalise over examples**. They are trained on examples and their job is to generalise over these examples so that they can classify previously unseen individuals and/or situations. Normally, the different speakers of a language are exposed to different sets of examples that they train their classifiers on. This means that even if speakers generally agree how to classify a majority of cases, there may be (actual or potential) borderline cases where they (would) make different judgements. In addition, even if two agents are exposed to the same training data, they not make identical generalisations from that data. Another way of putting this is that the classifier is underdetermined by the training data.

Second, classification is **sensitive to noise**. For borderline cases especially, noise coming e.g. from perception may affect the result of classification. In the left-or-right game, for example, noise may affect the perception of the object on the surface and the detection of its precise location, which in borderline cases may determine the outcome of classification.

Third, classification is not abstract, but a concrete process that is physically manifest and **extended in time**.

Finally, since meanings (modelled as classifiers) are coordinated in a community, meanings (and their associated classifiers) are always **provisional and potentially up for revision**. Note that this is a fact about *potential*, so that even a word whose meaning is unchanged for a very long time still has the potential to have its meaning be called into question and become subject to revision.

7.2 The time argument

With this in mind, let us proceed to our first argument, the *time* argument. This relies on two of our assumptions above: classification is a process that takes time, and meanings are always provisional. Given that an agent has classified some situation as being correctly described by an expression e, who is to say whether this is correct or not? If we accept that linguistic meanings are conventional (De Saussure, 1989), the answer is: the speech community that she is part of. The task of coordinating our judgements, i.e. agreeing on how to talk about the world, is therefore vital for the very existence of linguistic meaning, and relies crucially on interactively coordination on meanings (semantic coordination).

Could we then perhaps talk about a determinate extension once a community has come to a final agreement about how to use a word? Well, how could we (or they) tell that they have? Only by successively going through all potentially relevant items (situations or entities) and seeing if they agree. If the notion of extensions in possible worlds (including, presumably, the actual world) is to be taken seriously (which it possibly should not, see Lappin (2012)), this means going through all objects in the universe (which first of all requires individuating all physical matter into discrete objects). Whether the extension is universal or situation-specific, this is a process that will take some time, and what guarantee is there that no single member of the community during this time will modify their take on the meaning of e?

What this argument purports to show is that there is no method that *guarantees* shared extensions, and hence that the idea of meaning as a determinate (objective or intersubjective) extension does not reflect the reality of language use. This does not mean communication is impossible, since humans are quite good at clearing up misunderstandings

and interactively coordinating on meanings that are sufficiently shared for the purposes of their interaction. The argument also does not show that humans can never achieve intersubjectively shared extensions (see also Section 9), only that this is not the general case.

7.3 The generalisation argument

The second argument (based on *generalisation*) goes as follows: given some dataset of entities referred to using an expression e, there are several generalisations that adequately describe it (Quine, 1960). Different agents may agree on all data in training set but have subtly different generalisations. New situations or entities may be encountered which are different from anything in the training set, and in such cases different agents may make make different judgements, even if they agree on everything in the (previous) dataset. This means that we cannot be sure that agents are coordinated on the extension of e unless the dataset contains all relevant entities/situations in the world, which in the worst case is all entities/situations in the world. So the procedure for arriving at a generalisation that ensures a determinate shared extension cannot realistically be executed.

It may be argued that this is just the familiar problem of concept learning, which must be faced by any theory of meaning. We have no problem with that; what we try to show is that *if* one decides (wisely, on our view) to model concrete meanings using classifiers (to account for semantic learning and coordination), then one cannot also keep the notion of determinate extensions of natural language words and expressions.

7.4 The noise argument

The outcome of acts of classification are not decided in advance – if they were, there would be no point in classifying anything. However, the idea of actually classifying everything of a certain category does not seem practically feasible. But if it is not possible to actually classify everything, perhaps we can rely on a counterfactual definition of extensions? Something along the lines of "if all members of the community were to review all entities/situations, they would all agree that the extension is . . .".

This brings us to our third argument, the *noise* argument. Extensions are derived from intensions modelled as classifiers, which take noisy real-world perceptual data as input, which means that noise is an inherent feature of real-world classification procedure. Any counterfactual definition would therefore have to rely on a notion of noise-free classification, which in turn would relies on some capability of deciding what counts as noise, and what does not. However, this leads to an infinite regress since the classification of noise is just another classification problem, sensitive to the same arguments the original classification problem.

The noisiness of classification is related to vagueness, but we would argue vagueness is a separate problem. Even if vagueness is dealt with in an extensional semantics, e.g. by introducing some notion of probabilistic set membership (e.g. fuzzy sets (Hersh and Caramazza, 1976)), there is still a question of whether a certain outcome of observation and classification can even in principle be reliably repeated on repeated classifications, since this requires excluding noise from classification, which in turn requires classifying noise and so on *ad infinitum*.

Another possible counter-argument is that even in physics, many foundational concepts cannot be measured without noise, but this does not disqualify them. However, this misses the point: we are not disqualifying object-level concepts as such (on the grounds of being sensitive to noise, or indeed on any grounds), but the idea that they – in the general case – can be modelled as determinate extensions.

8 What about non-concrete words?

So far, we have talked only about concrete words (i.e. words referring to observable reality). What about non-concrete words? We take it that (following, among others, Carnap (1998)) at least many abstract words are ultimately grounded in experience. Take for example *democracy*, a paradigm abstract term which nevertheless is indirectly observable – as shown by the practice of sending international observers to an election with the task of deciding if and to what degree an election is democratic.

9 When are extensional semantics applicable?

Despite the problems noted above for a purely extensional semantics,from the perspective of a general theory of language use, there may be certain types of activities in which it is nevertheless applicable in the sense that extensions can be treated as given and determinate. These are activities where classification procedures are explicitly oper-

ationalised and agreed upon, such as in the natural sciences when they are in their "normal" (intra-paradigmatic) state of accumulating knowledge according to a more or less fixed conceptualisation. Another example is activities that are sufficiently routinised to be susceptible to formalisation, corresponding to what Dreyfus (1992) refers to as *systematic domains* – a set formal representations that can be used in a e.g. a dialogue system, and that embodies a static, intersubjective and uncontroversial interpretation of the situation in which the system will function.

It may be speculated that it is these and similar special cases of language use that underlie the intuitive attraction of extensional semantics. However, insofar as one is interested in a *general* theory of natural language meaning, it is not sufficient to handle only some particular types of language use. Indeed, even within the natural sciences, semantic change and coordination frequently occurs, as exemplified by Ludlow (2014) using the case of Pluto which was re-classified from being a planet to no longer being a planet. Also in everyday activities, semantic coordination is commonplace (Myrendal, 2015, 2019).

In addition to the cases mentioned, in any specific situation, agents may succeed in jointly classifying all relevant objects (and situations) into shared extensions using their individual classifiers together with protocols for semantic coordination (Larsson, 2018) . As long as these classifications are done sufficiently explicitly to become grounded, the extensions are, at least for all practical purposes, determinate and shared. Regardless of how this is achieved (whether through routinisation, scientific operationalisation or semantic coordination in a specific situation), these determinate extensions could in principle theoretically be treated using a purely extensional semantics.

In TTR such extensions are modelled as *witness caches* that can replace classifiers once (situation-specific) determinate extensions have been achieved (Larsson, 2020). This is in line with the idea of "natural language as a toolkit for building formal languages" (Cooper and Ranta, 2008). Reflecting this, the actual core definition in (Larsson, 2020) is not (1) but (8):

(8) $s : T$ iff $\mathit{Clfr}(T)(s) = T$ or $s \in F(T)$

where $F(T)$ is the witness cache, for type T – a set of situations (in the case of ptypes) previously judged to be of type T. On our present account, the witness cache for a type and an agent could represent mutually (within a community) agreed-upon classifications in a systematic domain. In the event of semantic coordination and updates to classifier parameters, the witness cache would most likely need to be cleared to avoid inconsistencies. To build a new witness cache, the community would again need to go through the process of jointly agreeing on a revised systematic domain, including extensions of central concepts.

10 Misclassification

If the correctness of any classification cannot be determined with reference to some absolute truth, then what does it mean to misclassify something? This question touches on the long-standing problem of rule-following and semantic normativity in the philosophy of language (Wittgenstein, 1953; Kripke, 1982; Boghossian, 1989; Glüer and Pagin, 1998).

In short, our answer is that it depends on what is being classified – specifically, whether there is an agreed-upon systematic domain (providing extensions) or operationalisation agreed upon by experts and deferred to by those not competent or willing to carry out the operationised classification procedure (as in the natural sciences). If there is, misclassification means classifying differently than the generally accepted classification. (However, note that to the extent that counterfactuality is involved here, in the form of assumptions of how something *would* be classified by e.g. an expert, the noise argument may still be applicable but to a lesser extent; operationalisation is no complete guarantee for agreement and noise may still creep in.)

If there is no operationalisation or systematic domain, misclassification (e.g. in the case of many artifacts or social constructions such as democracy) can mean simply classifying differently than others in the community (whose classification prevails). In the case of artifacts, the creator(s) of the artifact (if such a person or group can be singled out) may have more of a say than others concerning the correct classification of the artifact, similar to the case of experts in science. Another possibility is that an individual at time t classifies something x as a but later classifies the same thing as b and comes to regard the latter classification as correct; in this case they can say "I was wrong in classifying x as a, it is actually b. An there are probably more

$$\begin{bmatrix} \ell_1 & = & a_1 \\ \ell_2 & = & a_2 \\ \dots & & \\ \ell_n & = & a_n \\ \dots & & \end{bmatrix} : \begin{bmatrix} \ell_1 & : & T_1 \\ \ell_2 & : & T_2(l_1) \\ \dots & & \\ \ell_n & : & T_n(\ell_1, l_2, \dots, l_{n-1}) \end{bmatrix}$$

Figure 3: Schema of record and record type

$$\begin{bmatrix} \text{ref} & = & \text{obj}_{123} \\ \text{c}_{\text{man}} & = & \text{prf}(\text{man}(\text{obj}_{123})) \\ \text{c}_{\text{run}} & = & \text{prf}(\text{run}(\text{obj}_{123})) \end{bmatrix} : \begin{bmatrix} \text{ref} & : & \text{Ind} \\ \text{c}_{\text{man}} & : & \text{man}(\text{ref}) \\ \text{c}_{\text{run}} & : & \text{run}(\text{ref}) \end{bmatrix}$$

Figure 4: Sample record and record type

variants of misclassification.

11 Summary and conclusion

We explored some consequences of idea of using classifiers to model intensions of natural language expressions for the notion of extensional meaning. Three arguments (the time argument, the generalisation argument, and the noise argument) are offered, on the basis of some observations about classifiers, why extensional semantics is not well suited for modelling perceptual meanings, especially if the latter are regarded as classifiers, and hence why extensional semantics does not (and indeed cannot) provide a general account of natural language meaning. We also discussed cases where nevertheless extensional semantics may be adequate, and the applicability of the argument to non-concrete words.

In future work, we would like to connect the arguments made here more explicitly to the wealth of work on rule-following and semantic normativity (Wittgenstein, 1953; Kripke, 1982; Boghossian, 1989; Glüer and Pagin, 1998). However, the best argument for any theory of natural language semantic is not to be found in polemic debates but in empirical coverage, computational tractability, and overall usefulness in scientific and practical matters. Hence, developing the idea of classifiers in natural language semantics further and showcasing its full potential is and remains our main area of future work.

Acknowledgements

This work was supported by grant 2014-39 from the Swedish Research Council (VR) for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg. Thanks to the anonymous reviewers for useful comments.

A Appendix: TTR

We can here only give a brief and partial introduction to TTR; see also Cooper (2005) and Cooper (2012). To begin with, $s : T$ is a judgment that some s is of type T. To make explicit who is making this judgment, the of-type relation may be subscripted with an agent A, as in $:_A T$. One *basic type* in TTR is Ind, the type of an individual; another basic type is $\mathbb{R}$, the type of real numbers. Given that T_1 and T_2 are types, $T_1 \to T_2$ is a *functional type* whose domain is objects of type T_1 and whose range is objects of type T_2.

Next, we introduce *records* and *record types*. If $a_1 : T_1, a_2 : T_2(a_1), \dots, a_n : T_n(a_1, a_2, \dots, a_{n-1})$, where $T(a_1, \dots, a_n)$ represents a type T which depends on the objects $a_1, \dots, a_n$, the record to the left in Figure 3 is of the record type to the right.

In Figure 3, $\ell_1, \dots \ell_n$ are *labels* which can be used elsewhere to refer to the values associated with them. A sample record and record type is shown in Figure 4.

Types constructed with predicates may be *dependent*. This is represented by the fact that arguments to the predicate may be represented by labels used on the left of the ':' elsewhere in the record type. In Figure 4, the type of c_{man} is dependent on ref (as is c_{run}).

If r is a record and ℓ is a label in r, we can use a *path* $r.\ell$ to refer to the value of ℓ in r. Similarly, if T is a record type and ℓ is a label in T, $T.\ell$ refers to the type of ℓ in T. Records (and record types) can be nested, so that the value of a label is itself a record (or record type). As can be seen in Figure 4, types can be constructed from predicates, e.g.,

"run" or "man". Such types are called *ptypes* and correspond roughly to propositions in first order logic. Given a set of predicates and a set of possible arguments, the set of possible ptypes is $\mathbf{PType}$, thus allowing for polymorphic predicates. The arity of a ptype P is a set of tuple of types $Arity(P)$. For example $Arity(\text{run}) = \{\langle Ind \rangle\}$.

A fundamental type-theoretical intuition is that something of a ptype T is whatever it is that counts as a proof of T. One way of putting this is that "propositions are types of proofs". In Figure 4, we simply use prf(T) as a placeholder for proofs of T; below, we will show how low-level perceptual input can be included in proofs.

Semantic phenomena which have been described using TTR include intensionality and mental attitudes (Cooper, 2005, in progress), dynamic generalised quantifiers (Cooper, 2004), modality (Cooper, in progress), vagueness (Fernández and Larsson, 2014), co-predication and dot types in lexical innovation, frame semantics for temporal reasoning, reasoning in hypothetical contexts (Cooper, 2011), enthymematic reasoning (Breitholtz and Cooper, 2011), clarification requests (Cooper, 2010), various kinds of negation (Cooper and Ginzburg, 2011), and information states in dialogue (Cooper, 1998; Ginzburg, 2012; Cooper, in progress).

References

Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. 2016. Learning to compose neural networks for question answering. *arXiv preprint arXiv:1601.01705*.

Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. 2015. Vqa: Visual question answering. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2425–2433.

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

Paul A Boghossian. 1989. The rule-following considerations. *Mind*, 98(392):507–549.

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

Rudolf Carnap. 1998. *Der logische aufbau der welt*, volume 514. Felix Meiner Verlag.

Robin Cooper. 1998. Information states, attitudes and dependent record types. In *ITALLC98*, pages 85–106.

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

Robin Cooper. 2005. Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3:333–362.

Robin Cooper. 2010. Generalized quantifiers and clarification content. In *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, Poznań. Polish Society for Cognitive Science.

Robin Cooper. 2011. Copredication, quantification and frames. In *LACL*, volume 6736 of *Lecture Notes in Computer Science*, pages 64–79. Springer.

Robin Cooper. 2012. Type theory and semantics in flux. In Ruth Kempson, Nicholas Asher, and Tim Fernando, editors, *Handbook of the Philosophy of Science*, volume 14: Philosophy of Linguistics. Elsevier BV. General editors: Dov M. Gabbay, Paul Thagard and John Woods.

Robin Cooper. 2019. Types as learnable cognitive resources in pyttr. In Cleo Condoravdi and Tracy Holloway King, editors, *Tokens of Meaning: Papers in Honor of Lauri Karttunen*, pages 569–586. CSLI Publications.

Robin Cooper. in progress. *Type theory and language - From perception to linguistic communication*.

Robin Cooper and Jonathan Ginzburg. 2011. Negation in dialogue. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011)*, Los Angeles (USA).

Robin Cooper and Aarne Ranta. 2008. Natural languages as collections of resources. In Robin Cooper and Ruth Kempson, editors, *Language in flux: relating dialogue coordination to language variation, change and evolution*. College Publications, London.

Ferdinand De Saussure. 1989. *Cours de linguistique générale*, volume 1. Otto Harrassowitz Verlag.

Simon Dobnik. 2009. *Teaching mobile robots to use spatial words*. Ph.D. thesis, University of Oxford: Faculty of Linguistics, Philology and Phonetics and The Queen's College, Oxford, United Kingdom.

Hubert Dreyfus. 1992. *What computers still can't do*. The MIT Press.

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

Raquel Fernández, Staffan Larsson, Robin Cooper, Jonathan Ginzburg, and David Schlangen. 2011. Reciprocal learning via dialogue interaction: Challenges and prospects. In *Proceedings of the IJCAI 2011 Workshop on Agents Learning Interactively from Human Teachers (ALIHT)*, Barcelona, Catalonia, Spain.

Jonathan Ginzburg. 2012. *The Interactive Stance*. Oxford University Press, New York.

Kathrin Glüer and Peter Pagin. 1998. Rules of meaning and practical reasoning. *Synthese*, 117(2):207–227.

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

Harry M Hersh and Alfonso Caramazza. 1976. A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General*, 105(3):254.

Saul A. Kripke. 1982. *Wittgenstein on Rules and Private Language : An Elementary Exposition*. Harvard University Press.

Shalom Lappin. 2012. An Operational Approach to Fine-Grained Intensionality. *UCLA Working Papers in Linguistics, Theories of Everything*, 17(2):180–186.

Staffan Larsson. 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).

Staffan Larsson. 2015. Formal semantics for perceptual classification. *Journal of Logic and Computation*, 25(2):335–369. Published online 2013-12-18.

Staffan Larsson. 2017. Compositionality for perceptual classification. In *IWCS 201712th International Conference on Computational SemanticsShort papers*.

Staffan Larsson. 2018. Grounding as a side-effect of grounding. *Topics in cognitive science*, 10(2):389–408.

Staffan Larsson. 2020. Discrete and probabilistic classifier-based semantics. In *Proceedings of PaM, Probability and Meaning*.

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

Peter Ludlow. 2014. *Living Words: Meaning Underdetermination and the Dynamic Lexicon*. Oxford University Press.

Diego Marconi. 1997. *Lexical competence*. MIT press.

Brian McMahan and Matthew Stone. 2015. A bayesian model of grounded color semantics. *Transactions of the Association for Computational Linguistics*, 3:103–115.

Will Monroe, Noah D Goodman, and Christopher Potts. 2016. Learning to generate compositional color descriptions. *arXiv preprint arXiv:1606.03821*.

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

Reinhard Muskens. 2005. Sense and the computation of reference. *Linguistics and Philosophy*, 28(4):473–504.

Jenny Myrendal. 2015. *Word Meaning Negotiation in Online Discussion Forum Communication*. Ph.D. thesis, University of Gothenburg.

Jenny Myrendal. 2019. Negotiating meanings online: Disagreements about word meaning in discussion forum communication. *Discourse Studies*, 21(3):317–339.

Willard Van Orman Quine. 1960. *Word and Object*. MIT Press.

F Rosenblatt. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386–408.

David Schlangen, Sina Zarrie, and Casey 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)*.

Ludwig Wittgenstein. 1953. *Philosophical Investigations*. Basil Blackwell Ltd.

Quanzeng You, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo. 2016. Image captioning with semantic attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4651–4659.