

# Towards a Type-Theoretic Formalisation of Semantic Prototypes

Jonathan Tan Jiayi

May 13, 2025

## Abstract

In this paper, I present a novel application of a Martin-Löf-style dependent type theory as a formal realisation of the linguistic concept of semantic prototypes. Together with a new approach to formal semantics that shifts the focus away from terms almost entirely and abandons considerations of objective truth, the paper also proposes a new theory of meaning for natural languages that formalises the plurality of interpretations in terms of semantic prototypes more completely. In doing so, it allows for a greater appreciation for the connection between the field of linguistics with formal semantics, metalogic, and type theory.

## 1 Introduction

Humanity has long since had several attempts to analyse meaning; out of the confluence of linguistics, psychology, philosophy, mathematics, and logic, several views have risen to prominence. However, while there are a plenitude of different approaches in our conquest to understand meaning, there are also opposing views and conceptual differences in equal number that can be challenging to unite into one cohesive understanding. One such example is the cognitive linguists' view that traditional takes on semantics based on objective truth are too restrictive and not powerful enough to properly express all of the complexities of language. Traditional views typically presuppose that meaning takes on a singular whole and talk about truth values, proofs, and models that can fail to capture the more subjective aspects of meaning and language.

In this same vein, more formal approaches often struggle to represent the subjectivity of language due to the inconsistencies it can introduce that are hard to reconcile. It is easier, and often more germane, to idealise the denotation of an expression as some composition of the denotation of its parts that each have some fixed meaning. Given that formal analyses are often more interested in these compositions as a representation for the grammar of natural language, many find the venture to formalise the specifics of the meaning of each part unwarranted.

This paper aims to bridge this gap between formal and cognitive semantics, which has largely remained overlooked even by prior attempts to unite formal and cognitive semantics, so as to allow the techniques and analyses of formal semantics to be applied even while taking into consideration the true subjectivity of language.

## 2 Motivation

Most orthodox approaches to formal semantics tend to find meaning within the alethic value of terms themselves; be it propositions and predicates in model-theoretic, proof-theoretic, and truth-value semantics, or even possible worlds and Kripke frames. Even for approaches involving the theory of types such as Montague grammar, which makes use of Kripke models together with Church's simply-

typed lambda calculus, meaning continues to be interpreted through the relationship of alethic values in terms.

One “shortcoming” of this approach to semantics when applied to linguistics is that it presupposes a mutual understanding of what constituent parts<sup>1</sup> mean between the individual who performed the analysis and the individual reading it. For instance, the common use of predicates in statements such as  $PERSON(x) \wedge APPLE(y) \wedge EAT(x, y)$  to represent someone eating an apple, semi-informal notation like in  $(\lambda xy.x \text{ eats } y)(person)(apple)$ , or some combination of the two. As discussed in the section below, this is often because the precision of rigorously defining what each part represents is not germane to the scope of the analyses.

The former example is explicitly a statement of truth-hood regarding whether some  $x$  is a person, some  $y$  is an apple, and whether  $y$  is being eaten by  $x$  with predicates. The oversight in systems such as these is that ultimately, the author and the reader may not have the same understanding of what  $APPLE(y)$  might be (e.g. whether  $y$  can be any kind of apple, or only stereotypical red apples; whether it includes a condition on  $y$  that it should appear to, or actually be edible given the context, or if the statement also includes the situation where someone is eating an apple rotten beyond recognition), resulting in a meta-hermeneutical conundrum. In a broader lens, this also highlights an issue with trying to formalise semantics considering the pluralism of interpretations.

## 2.1 A Tacit Oversight

This “shortcoming” is referred to in quotes mainly because the above points usually do not impede the analysis’s purpose, which is often limited to showing the relation of semantic ideas to one another or to show how concepts are composited with a certain verbiage, even in modelling natural languages as having a formal grammar (such as with Montague Grammar). However, to establish a theory of meaning, it would be remiss to gloss over these points as it admits some degree of incapacity to explain fully<sup>2</sup> how meaning is found naturally in statements such as “it is an apple” at all levels beyond structure alone. Moreover, by focusing on the alethic value of statements as the carriers of meaning, it introduces several paradoxes and inconsistencies that such calculi may not be powerful enough to reconcile.

To elaborate on the issue, consider the following two illustrations:

**Illustration 1** First, suppose there is an individual who has grown up in an environment that has indoctrinated in them the understanding that all crows are black, and another who has grown up in an environment that has indoctrinated in them that all crows are white. If they were to read a warning on a sign that reads “there are crows in the backyard”. How would one formally analyse the semantics of that statement in a way that properly represents what both individuals might understand from it?

One possibility that current approaches might advocate for is to establish that  $CROW(x)$  is simply any one individual’s understanding of what a crow is. This might entail breaking the statement into cases of  $BLACK_{CROW}(x)$  and  $WHITE_{CROW}(x)$ , or even defining  $CROW(x)$  to be some expression in terms of  $WHITE(x)$  and  $BLACK(x)$ . Now, introduce more people who each have different understandings of what a crow is (consider that the general population is unable to agree on even the difference between a crow and a raven for a myriad reasons). It becomes evident that it is no longer feasible to consider every possible interpretation of that statement to analyse it formally in this way, given that enumerating through every possible configuration to exhaustion is not a valid technique when generalising beyond a finite number.

<sup>1</sup>The imprecise term “part” is used intentionally to avoid loose use of terms such as “referent” due to issues of nuance regarding different semantic theories such as Referent Semantics.

<sup>2</sup>There is an argument to be made regarding semiotics from the point of view that signs (such as words) simply signify a reference to some object or concept, although this line of justification encounters the same issues highlighted for the purposes of formalising meaning.

Furthermore, if a model of a natural language should suggest that such a statement be parsed into some logical form, consider that upon arriving at the backyard, at least one of the two original individuals will experience a category mistake. This presents a paradoxical inconsistency where one logical statement can have contradicting alethic values at the same time following orthodox approaches to model meaning.

**Illustration 2** Second, consider what objects the word “fish” immediately brings to mind. In practice, it is likely that the objects include a salmon, a tuna, or a similar fish for most, due to one’s culture and experiences making them stereotypical “fish”. It is also likely that the word “pet” brings to mind a dog or a cat. An interesting behaviour occurs when the two are juxtaposed as the phrase “pet fish”; it is just as likely that the objects brought to mind are neither those of “pet” nor of “fish”, like a guppy or a goldfish. This additional consideration of interpretation introduces a conundrum when trying to formally express the composition of concepts as predicates. For instance, the traditional analysis of “pet fish” as  $PET(x) \wedge FISH(x)$  would suggest that one’s interpretations of “pet fish” should be the intersection of that of “pet” and “fish” (which is only true when idealising understanding to only be facts that are true to the world).

While a substantial amount of work has been done on the linguistics end to reconcile these problems (see Section 3.2), most formal semantic representations maintain the traditional view and idealise natural languages to be able to fit those analyses, due to the difficulty in attempting to represent something so subjective and informal formally.

## 3 Current Work

### 3.1 Neo-Davidsonian Discourse Representation Theory

Worth mentioning is that this paper is not alone in attempting to marry cognitive linguistics with formal linguistics, nor is it unique in endeavouring to do so using types. A paper by Hamm, Kamp, and Lambalgen [3] attempted to show that the validity of Discourse Representation Theory (DRT), where the interpretation of language is influenced by the context it is in, is not precluded by formal methods of analysis. The paper does this by relating DRT to a variation of neo-Davidsonian event semantics, which uses a combination of predicates and an event calculus to formally analyse sentences by relating them to events, adapted to make use of a set of axioms established in the paper.

However, while the paper addresses the complexity of sentences as having multiple possible interpretations based on context, it continues to employ arbitrarily assigned predicates in its realisation of Discourse Representation Structures. In this way, the paper does not directly address the issues highlighted above, albeit by virtue of not having any intention to do so in the first place.

### 3.2 Semantic Prototype Theory

The above criticisms of classical methods of analysis (excluding the second illustration) are similar to prior criticisms of classical theories of categorisation [9], which are accompanied with several studies demonstrating the extent that attempting to put categorical labels on concepts (such as making use of arbitrary predicates and terms) often fall victim to large variances of interpretation rather than absolute membership in any one category [4, 5, 7, 8].

Semantic Prototype Theory (SRT) departs from traditional view of categorisation which categorises based on fulfilment of necessary conditions, and instead categorises based on prototypical members. Operating on the principle that every concept has an example that best represents it, called a prototype, SRT considers a category<sup>3</sup> to be defined by the set of entities within it that represent a

---

<sup>3</sup>The term “category” here and throughout the rest of the paper refer to linguistic categories and not the eponymous concept of category theory.

prototypical member, which can be thought of as members that are most often associated with a category [9].

This paper makes use of this concept to legitimise differences in interpretation by similarly focusing on prototypical members of categories of concepts. However, one slight difference is that the paper places focus on the personal level of interpretation rather than the “average” of interpretations. Specifically, it considers a prototype to be the members of a category that *any one person* considers to represent the category best, and a category to be defined based on the plurality of all prototypes of the category rather than just the most common. Additionally, it abandons notions of set theory and Boolean logic, which the paper later demonstrates to reconcile one formal argument against SRT [2, 6] mentioned in Illustration 2 (termed the *PET-FISH problem*).

## 4 Type-Verification Semantics

To address the issues highlighted in Section 2.1, we can make use of types. Types shift the focus from truth-hood in the classical sense to whether a statement is well-typed, and allow us to formally model semantic prototypes. For instance, we can replace  $CROW(x)$  in the first illustration with  $x : crow$  instead, where  $crow$  is some type constructed from a set of base types and type constructors (elaborated in a later section). In this way, we assert that  $x$  matches some semantic prototype of a crow instead of asserting that “ $x$  is a crow” by some arbitrary definition.

Following the first illustration’s depiction of the coexistence of differing semantic prototypes of a crow, we can represent each of the individuals’ understandings as  $x : crow_{black}$  and  $x : crow_{white}$  (where the two are types that have been constructed in some way). Then, it is possible to generalise them as  $x : crow_i$  where  $crow_i : crow$ , such that  $crow$  is a universe which represents the category of crows.

Moreover, in considering the second illustration, it follows from using (dependent) type theory that terms of a type cannot also be terms of another type. Thus, it is natural to represent how our brains compose “pet” and “fish” in a statement such as “pet fish” as some type construction with *pet* and *fish*, which avoids the paradox of  $x \models PET(x) \wedge FISH(x)$  while  $x \not\models PET(x) \vee FISH(x)$  since  $x : \tau \not\models x : pet$ ;  $x : \tau \not\models x : fish$  given any type  $\tau$  composed of *pet* and *fish*.

### 4.1 Evidentiality as Dependent Types

Diving deeper into the first illustration, we see that the type representing the individuals’ understandings (notated above as  $crow_i$ ) depends on the individual who processed the information. In a dependent type theory, this can be represented as a type-valued function  $\phi_{crow} : perceiver \rightarrow crow$ , where any concept of someone who has perceived has the base type *perceiver*. This way, we can type every term more precisely as a dependent pair type  $(t, x) : \Sigma_{t:perceiver} \phi_{crow}(t)$ . Thus, for the two individuals<sup>4</sup>  $a, b : perceiver$ , we can define  $\phi_{crow}(a) \equiv crow_{black}$ ;  $\phi_{crow}(b) \equiv crow_{white}$  and both their understandings as  $(a, x)$  and  $(b, x)$ . This allows us to formally express the generalised case described by the illustration and the base case as a specific form of the generalised case.

Generalising beyond just crows, we arrive at  $\phi : perceiver \rightarrow \mathcal{U}$  and  $(t, x) : \Sigma_{t:perceiver} \phi(t)$  where  $\mathcal{U}$  represents any category. With this, we have a way to rigorously generalise all interpretations of a concept tagged with the individual to whom the interpretation belongs. For convenience, we will use the general notation  $\phi$  even for specific cases unless distinction is necessary.

It follows from this implementation that we are also able to make analyses of evidentiality. Using the same premise as the first illustration, we can model the difference in meaning between person A saying “there is a crow” from first-hand experience and reporting what the other had said through a

<sup>4</sup>More precisely, we refer not to the concept of them as humans, which would have a more complex type, but as sources of perception.

difference in tagging. The first case is synonymous with representing their interpretation of the same statement, hence their understanding of “crow” in the statement is  $(a, x)$ . However, in the second case, the understanding of “crow” is in fact person A’s understanding of person B’s understanding of a crow, as person A is reporting the presence of a crow as person B first perceived it.

Expanding the definition to  $(t_1, \dots, t_n, x) : \Sigma_{t:\Pi_{n:I} \text{perceiver}} \phi(t)$  where  $\phi : (\Pi_{n:I} \text{perceiver}) \rightarrow \mathcal{U}$ , assuming person A has a perfect understanding of person B’s understanding of crows<sup>5</sup>, we can define  $\phi(a, b) \equiv \phi(b) \equiv \text{crow}_{\text{white}}$  such that  $(a, b, x : \text{crow}_{\text{white}}) : \Sigma_{t:\Pi_{n:I} \text{perceiver}} \phi(t)$ . If person A instead wrongly assumes person B has the same understanding of crows as they do, we can then define  $\phi(a, b) \equiv \phi(a) \equiv \text{crow}_{\text{black}}$ , such that  $(a, b, x : \text{crow}_{\text{black}}) : \Sigma_{t:\Pi_{n:I} \text{perceiver}} \phi(t)$ .

## 4.2 A New Theory of Meaning

The crux of this system is its logic and how it forms a theory of meaning. Cognitive linguistics and neuroscientific research [1] have posited that the way we understand is directly related to how we perceive; in this way, understanding is intrinsically found on being able to construct meaning using our experiences of perception. By making use of a Martin-Löf-style dependent type theory, we find an almost direct parallel between constructing meaning in this manner with constructive logic. This then allows us to use the logic inherent to our type theory as a formal representation of meaning, as illustrated in the examples below.

Define a base type *nociception* representing any concept of a physiological sensory mechanism in response to potentially harmful stimuli (such as the triggering of nociceptors). Consider the English word “pain”, which we can attempt to represent as the universe *pain*. For some individual<sup>6</sup>  $i : \text{perceiver}$ , we can then define  $\phi(i) \equiv \text{nociception}$ , which we would have described as having some individual understand pain to refer to any instance of nociception in accordance with the explanations in the earlier section.

However, what does this entail precisely? Behind the scenes of mentally parsing the concept of pain, one subconsciously imagines themselves feeling pain, or otherwise remembers what it feels like to be in pain. For the individual above, they may understand pain by subconsciously thinking about the sensation they felt when they fell off a bicycle or simulating a similar sensation on the body from their experiences with it. In this way they construct an example  $y$  such that  $y : \text{nociception}$  and therefore  $(i, y : \text{nociception}) : \Sigma_{t:\Pi_{n:I} \text{perceiver}} \phi(t)$  are well typed. Thus, the concept of pain has meaning to them since their understanding of pain is an inhabited type.

Conversely, consider some other individual  $j : \text{perceiver}$  whose understanding of pain can be expressed as some type construction  $\tau$  such that  $\phi(j) \equiv \tau$ . In trying to parse the concept of pain, they can think of no instance of them feeling pain and are unable to imagine themselves feeling pain due to having had no perceptive experience tagged as pain. In this way it is not possible to construct an example  $z$  such that  $z : \tau$  or  $(j, z : \tau) : \Sigma_{t:\Pi_{n:I} \text{perceiver}} \phi(t)$  are well typed. Thus, this in turn suggests that  $\tau \equiv \mathbf{0}$  and is uninhabited, representing formally that the individual cannot understand the concept of pain.

Recall that terms of type  $\text{perceiver} \rightarrow \mathcal{U}$  represent associations between some individual and their respective understanding of a concept. Generalising this meta-understanding, terms of type  $\tau_1 \rightarrow \tau_2$  can also be said to represent mental associations of some instance of a concept to some instance of another concept<sup>7</sup>. Likewise, terms of type  $\tau \rightarrow \text{perceiver}$  can be said to represent an association between some instance of a concept to some individual who is perceiving it (an act of perception).

With this, we can then also represent one understanding of the concept of causing pain as  $\phi(i) \equiv$

<sup>5</sup>This assumption is only for illustration; in a more realistic definition  $\phi(a, b) \neq \phi(b)$ .

<sup>6</sup>For this and further mentions of individuals with *perceiver* as a type, see footnote 4

<sup>7</sup>It is an important distinction to make that our methodology does not arbitrarily prescribe formal expressions to represent meta-concepts such as these. Instead, these explanations are merely informal attempts to approximately explain to the reader what the formalisms are able to represent in the context of linguistics.

$perceiver \rightarrow \phi_{pain}(i) \rightarrow perceiver$ , which can be chiefly described as something which associates a perceiver with the perception of pain. Assuming that the concept of pain is meaningful for the individual such that there exists a  $\hat{a} : \phi_{pain}(i)$ , the process of understanding causing pain can then be described as being given a  $\hat{b} : perceiver$  and constructing an example of perceiving pain  $\phi_{pain}(i) \rightarrow perceiver$ . In other words, the individual proves to themselves that there exists some way to “give”  $\hat{a}$  to  $\hat{b}$  such that this process becomes a witness for  $perceiver \rightarrow \phi_{pain}(i) \rightarrow perceiver$  and gives it meaning.

A simplistic example supposes that the individual understands the concept of being hit such that there exists a  $c : perceiver \rightarrow \phi_{hit}(i) \rightarrow perceiver$  and the concept of hitting being perceived as nociception such that there exists a  $d : (\phi_{hit}(i) \rightarrow perceiver) \rightarrow (nociception \rightarrow perceiver)$ . The individual can then pass  $\hat{a}$  as an argument to  $c$  giving an element of  $\phi_{hit}(i) \rightarrow perceiver$ , which represents imagining the perceiver being hit. Then passing as an argument to  $d$ , representing imagining the perceiver feeling pain following being hit, giving an element of  $nociception \rightarrow perceiver \equiv \phi_{pain} \rightarrow perceiver$  as desired, proving that there exists at least one example of an act of causing pain.

In this way, we can formally describe the process of understanding any concept as being able to prove it constructively.

## 5 Conclusion

We see that with this novel theory of meaning and application of type theory, we are able to make formal analyses while continuing to respect the plurality of interpretations that can be had for each semantic category. Nonetheless, this paper only covers a subset of possibilities and only seeks to demonstrate the theoretical feasibility for use as a theory of meaning and model of semantics. Further research is required to fully flesh out the base types and type constructors required for the type theory to be powerful enough to theoretically express all concepts and provide additional formal explication for the definition of categories from base types.

## References

- [1] Seana Coulson. Constructing Meaning. *Metaphor and Symbol*, 21(4):245–266, October 2006.
- [2] Jerry Fodor and Ernest Lepore. The red herring and the pet fish: why concepts still can't be prototypes. *Cognition*, 58(2):253–270, February 1996.
- [3] Fritz Hamm, Hans Kamp, and Michiel Van Lambalgen. There is no opposition between Formal and Cognitive Semantics. *Theoretical Linguistics*, 32(1):1–40, September 2006.
- [4] Willett Kempton. *The Folk Classification of Ceramics: A Study of Cognitive Prototypes*. Academic Press, New York, 1981.
- [5] William Labov. The boundaries of words and their meanings. In Charles-James N. Bailey and Roger W. Shuy, editors, *New Ways of Analyzing Variation in English*, pages 340–373. Georgetown University Press, 1973.
- [6] Daniel N. Osherson and Edward E. Smith. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9(1):35–58, January 1981.
- [7] Eleanor Rosch. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3):192–233, 1975. Place: US Publisher: American Psychological Association.
- [8] Eleanor H. Rosch. On the internal structure of perceptual and semantic categories. In *Cognitive development and the acquisition of language*, pages xii, 308–xii, 308. Academic Press, Oxford, England, 1973.
- [9] John R. Taylor. *Linguistic Categorization*. Oxford Textbooks in Linguistics. Oxford University Press, Oxford, New York, third edition, third edition edition, November 2003.