Applications of Epsilon Terms in Natural Language Processing

Christoph Roschger Gergö Barany

VU Epsilon Calculus, May 2009

Abstract

Not concrete.

1 Introduction: Semantics of Language

Any application that aims to process natural language semantically needs some format in which to represent and reason about semantics. The central theorem for that purpose stating that "There is no essential difference between the semantics of natural languages (like English) and formal languages (like predicate logic)" was elobareted by R. Montague in the 1960s and early 1970s.

There are various kinds of formalisms used to represent the semantics of natural language nowadays. One especially used in linguistics are discourse representation structures, but here we will focus on predicate logic and its extension by ϵ-terms proving itself as a useful choice.

For example, the semantics of phrase (1) can be represented in first-order logic as (2).

(1) A man walks. He talks.

(2) $\exists x.(M(x) \wedge W(x) \wedge T(x))$

Given this formalization, the sentence could be evaluated for its truth value in some situation (represented by some other set of formulas); it could be stored in some database; or it could be used to derive new formulas. The difficulty of semantic modeling lies in finding a way to arrive at the formula, given a parse of the phrase.

Several large problem areas will be discussed in this paper:

Compositionality is the problem of finding meaningful semantic representations for fragments of text; these semantic fragments can then be combined to represent the semantics of the entire text by composing the parts according to syntactic structure. The logical formula for a sentence is thus typically expected to have a structure that is comparable to the syntactic structure of the sentence—but this is not the case for the example, in which the existential quantifier extends over the semantics of two sentences. The semantics of the part 'he talks' cannot be identified in this formula, as the variable x occurs free if only the fragment $T(x)$ is considered.

The problem of *coreference* is that of determining that the anaphoric pronoun 'he' refers to the man named in the sentence before. To resolve this, some information on the previously established context must be used in the correct way.

The problem of *definiteness* is that of determining what a definite noun phrase may refer to, if that object is not unique. For instance, the phrase 'the island' is not unique as there are many islands in the world; however, in some given situation, it may be entirely clear which island is meant.

2 Applications

This section presents a number of interesting applications of ϵ-terms in representing natural language semantics.

2.1 Anaphora resolution

An anaphoric expression in general is an instance of an expression referring to another one. For example, in sentence (1), *He* is an anaphoric pronoun referring to *A man*. As already mentioned, anaphora together with the requirement of composability pose a major problem when trying to represent a sentence (or a sequence of sentences) as a first order formula.

One possible solution presented by Meyer Viol [Mey95] is to use ϵ-terms in order to represent the semantics of a sentence. Consider the following sentence:

(3) A man loved a woman.

Possible translations to first order logic are:

(4) $\exists x.[M(x) \wedge \exists y.[W(y) \wedge L(x,y)]]$

(5) $\exists y.[W(y) \wedge \exists x.[M(x) \wedge L(x,y)]]$

Making use of ϵ-terms one can instead find the following representation for sentence (3):

(6) $\exists x.M(x) \wedge \exists y.W(y) \wedge L(\epsilon x.M(x), \epsilon y.W(y))$

Meyer Viol shows [Mey95] that formula (6) has exactly the same ϵ-free consequences as formulas (4) and (5), meaning that both these formulas can be derived from formula (6) and that every model of formula (4) or formula (5) can be supplied with a choice function to verify formula (6). Thus, it is indeed an adequate representation of sentence (3). An interesting property of this representation is the fact that both of the quantifiers in $\exists x.M(x)$ and $\exists y.W(y)$ do not range over the term $L(\epsilon x.M(x), \epsilon y.W(y))$.

Using ϵ-terms this way, anaphoric references can be systematically resolved [Mey95]. The basic idea is that the anaphoric expression is represented by exactly the ϵ-term containing the representation of the referent. As an example, sentence (1) is formalised as:

(7) $\exists x.[M(x) \wedge W(x)] \wedge T(\epsilon x.[M(x) \wedge W(x)])$

in the same sense of having the same ϵ-free consequences as formula (2).

Formally, the introduction of the ϵ term works as follows. First, the structure of the semantic representation is constructed, with *schematic terms* denoted ν taking the place of anaphoric references. These schematic terms are placeholders for future replacement by ϵ terms. The schematic formula for sentence (1) is (8). Note that this formula is compositional.

(8) $\exists x.[M(x) \wedge W(x)] \wedge T(\nu)$

The basic idea of finding the correct ϵ term for a given schematic term is to identify an existentially quantified formula which describes the object that is referenced anaphorically. Finding the most appropriate existentially quantified formula uses the concept of *extended scope* (E-scope) of logical formulas, defined by the following rules:

1. In $\varphi \wedge \psi$ both φ and ψ have E-scope over $\varphi \wedge \psi$.
2. If φ has E-scope over $\psi \# \chi$, $\# \in \{\wedge, \vee, \rightarrow, \leftrightarrow\}$, then it has E-scope over ψ and χ unless it is identical.
3. In $\varphi \rightarrow \psi$, φ has E-scope over ψ.
4. If φ has E-scope over ψ and ψ has E-scope over χ then φ has E-scope over χ.

In contrast to more traditional definitions of 'scope', the E-scope of a formula does not only extend to its subformulas and subterms. Rather, in conjunctions and implications, E-scope of operands is 'lifted' to the whole formula. This is intended to match the fact that in phrases such as (1), objects are referenced across sentence boundaries, which are most naturally modeled as conjunctions.

The substitution rule for schematic terms is simple: If all occurrences of a schematic term ν in some formula ψ lie in the E-scope of some formula $\exists x.\chi$, then ψ may be instantiated to $\psi[\epsilon x.\chi/\nu]$.

In the example (8), each conjunct has E-scope over the whole formula. In particular, ν is in the E-scope of $\exists x.[M(x) \wedge W(x)]$, thus it can be replaced by $\epsilon x.[M(x) \wedge W(x)]$. This instantiation yields the desired result, formula (7), which is still a compositional representation of the intended semantics of the example phrase.

This approach can be refined further to handle more complex sentences including the quantifiers *every* and *some* and the proper representation of plurals. This is (partly) achieved by Meyer Viol [Mey95] by giving a grammar defining a fragment of the english language and providing rules who to construct a proper representation using ϵ-terms.

2.2 Dynamic semantics

Peregrin and von Heusinger [PvH95] propose a different way of resolving anaphora: Their representation uses special epsilon-like terms, and their semantics interprets sentences as relations between sets of choice functions. This approach interprets sentences by their potential to change choice functions and builds upon Groenendijk and Stockhof's Dynamic Predicate Logic [GS91], in which the assignment function from variables to domain objects is updated dynamically in order to represent anaphora.

For example, in (1), the phrase *a man* in the first sentence will 'introduce' a new object by modifying the current choice function to refer to this new man. The reference *he* in the second sentence simply evaluates the current choice function for the set 'man', which in the context of the first sentence will refer to just that currently most salient man. Both sentences are relations between sets of choice functions; their relational semantics can be derived independently, and their combined semantics is defined simply by the usual composition of relations.

The language of this dynamic approach is a variant of the language of classical first-order logic. It does not have the classical quantifiers $\forall$ and $\exists$, but rather special quantifiers *some* and *every*; terms are either constants or built from predicates using the determiners *a* or *the*. Conjunction is written as ';'. The semantics of (1) is given by (9).

(9) $W(a(Man)); T(he)$

The semantics of $the(P)$ can be understood as $\epsilon x.P(x)$, where the *the* term is always interpreted by the 'current choice function' as the 'currently most salient' object of kind P. The semantics of $a(P)$ is just to introduce a new object and corresponding choice function for kind P, such that later occurrences of $the(P)$ reference just this object.

The semantics $\|E\|$ of a term or formula E is given as a relation between choice functions, i. e., a set of pairs $\langle e, e'\rangle$ of choice functions. The notation e^S denotes a choice function that is just like e, except that $e^S(S) \neq e(S)$. That is, e^S assigns a new representative to set S. The equations defining $\|\cdot\|$, as they are relevant to the example, are:

$$\begin{aligned} \|a(P)\| &= \{\langle e, e'\rangle \mid e' = e^{\|P\|}\} \\ \|the(P)\| &= \{\langle e, e'\rangle \mid e' = e \text{ and } e'(\|P\|) \text{ is defined}\} \\ \|he\| &= the(Man) \\ \|S_1; S_2\| &= \{\langle e, e''\rangle \mid \langle e, e'\rangle \in \|S_1\|, \langle e', e''\rangle \in \|S_2\| \text{ for some } e'\} \end{aligned}$$

Thus, evaluating $W(a(Man))$ in (9) gives a relation between choice functions where the 'new' choice functions map to a new representative for the *Man* predicate. The relation for $T(he)$ does not change any choice function, it merely insists that there be a defined representative for *Man*. These relations can be derived independently, in a fully compositional way. Joining them with the ; operator ensures that *he* will refer to the representative created for the previous occurrence of $a(Man)$. The conjoined relation itself can be joined with semantic relations for other parts of a larger text.

The approach also provides for universal and existential quantifiers. The semantics of the universal *every* is given by:

$$\begin{aligned} \|every(S_1, S_2)\| = \{\langle e, e'\rangle \mid &\text{ for every } e_1, \text{ if } \langle e, e_1\rangle \in \|S_1\|, \\ &\text{then } \langle e, e_2\rangle \in \|S_2\| \text{ for some } e_2\} \end{aligned}$$

So-called *donkey sentences* serve as an important touchstone for any theory about anaphora. They are of the form

(10) Every farmer who owns a donkey beats it.

The difficulty here is to resolve *it* to refer to *a donkey*, although *a donkey* is bound inside the *every*-quantifier. Using this dynamic approach the sentence can be formalized as

(11) $every(Own(a(Farmer), a(Donkey)), Beat(he, it))$

which indeed gets the intended interpretation due to the semantic constaints on the choice function relations of the subformulas:

$$\begin{aligned} &\|every(Own(a(Farmer), a(Donkey)), Beat(he, it))\| = \ldots = \\ &\quad \{\langle e, e'\rangle \mid e = e' \text{ and for every } e_1, \text{ if } e_1 = e^{\|Farmer\|, \|Donkey\|} \text{ and} \\ &\quad\quad \langle e_1(\|Farmer\|), e_1(\|Donkey\|)\rangle \in \|Own\| \\ &\quad\quad \text{then } \langle e_1(\|Farmer\|), e_1(\|Donkey\|)\rangle \in \|Beat\|\} \end{aligned}$$

There are several directions the dynamic semantics presented can be extended. For example, one could introduce *categories*, and when updating a choice function e for a set S, also update $e(S')$ for all categories that S belongs to. This way, for sentences like

(12) A boxer walks. The sportsman whistles.

the anaphoric expression *the sportsman* can be linked to *a boxer*, provided $\|boxer\|$ belongs to the category $\|sportsman\|$.

2.3 Situation-dependent descriptions

Egli and von Heusinger [EvH95] discuss the role of the epsilon operator in the representation of definite descriptions. Their focus is on cases where the description is not unique because more than one object may be categorized by it.

Consider the difference between the two phrases (13) and (14).

(13) the center of mass of the solar system

(14) the island on Lake Constance

There is a unique center of mass of the solar system, but there are three islands on Lake Constance. If these phrases are represented by terms (15) and (16), respectively, the choice function for (15) therefore is uniquely determined, while there are three possible choice functions to interpret (16).

(15) $\epsilon x.C(x)$

(16) $\epsilon x.I(x)$

The problem, then, is to determine the correct choice function, considering factors such as the situation and the speaker. The solution proposed by Egli and von Heusinger is to assign different choice functions to different situations and speakers; in the semantics formulas, these are represented by indexed epsilon operators.

The second contribution of their work is to view a choice function for some set S not as a mapping from S to some element of S (if there is one), but rather as a mapping from S to a well-ordering on S. A term $\epsilon x.I(x)$ is then interpreted as the first island (i. e., the least element) in the order of islands assigned by the choice function. This formalization can be used to express more complex definite phrases; for instance, the phrase (17) can be formalized as (18).

(17) the other island on Lake Constance

(18) $\epsilon y.[I(y) \wedge y \neq \epsilon x.I(x)]$

That is, the semantics of *other* can be expressed by referring to the second object in the ordering.

2.4 Semantics for impossible objects

Egli and von Heusinger [EvH95] present a further interesting application of epsilon terms. They mention sentences that refer to nonexistent or, stronger, impossible objects. Although such sentences are seemingly (logically) contradictory, they are sometimes used in "real world" language and can be understood in an informal way. Semantics for these can exploit the property of epsilon terms that they assign arbitrary objects to the empty set.

Consider (19), which they claim 'can be uttered without contradiction'.

(19) The ghost that is noisy in the attic is not a ghost.

(20) $\neg G(\epsilon x.[G(x) \wedge N(x)])$

A corresponding formalisation using an epsilon term is (20). It can be read as 'that thing that you call a ghost, which makes noise, is not a ghost'.

Here, if there are no ghosts, the set denoted by $G(x) \wedge N(x)$ will be empty. Thus, the epsilon operator may denote an arbitrary object—for example, it could be a cat making noise in the attic. In this sense, the rhetoric figure of (19) can be seen to be represented by the formula (20).

Moreover, from (20) one can directly infer $\neg(\exists x.[G(x) \wedge N(x)])$ using the tautology $\neg p \rightarrow \neg(p \wedge q)$ and the first Hilbert rule which captures the intended meaning of (19).

Another example demonstrating that impossible objects can be represented by ϵ-terms as well is the following:

(21) (Even) the nappy that is not dry is dry.

(22) $D(\epsilon x.[N(x) \wedge \neg D(x)])$

A possible interpretation for this sentence is that one wants to state that all nappies are dry and, again, one can easily infer $\forall x.[N(x) \rightarrow D(x)]$ from (22).

References

[EvH95] U. Egli and K. von Heusinger. The epsilon operator and e-type pronouns. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, pages 121–141, 1995.

[GS91] Jeroen Groenendijk and Martin Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14(1):39–100, 1991.

[Mey95] W. Meyer Viol. Instantial Logic. *Utrecht: PhD dissertation*, 218, 1995.

[PvH95] J. Peregrin and K. von Heusinger. Dynamic semantics with choice functions. *Choice Functions in Natural Language Semantics. Arbeitspapier*, 71:329–353, 1995.