Position paper: Proof-Theoretic Semantics as a viable alternative to Model-Theoretic Semantics for natural language

Nissim Francez
Computer Science dept., the Technion-IIT,
Haifa, Israel
(francez@cs.technion.ac.il)

Introduction

– This paper is a response to the last question in the CFP, about alternatives to model-theoretic semantics (MTS).

- I do not present any specific results, just argue for proof-theoretic semantics (PTS) as such an alternative.

– PTS is well-established within Logic (e.g.,Dummet, Prawitz). See SEP for an overview.

- I have extended PTS to fragments of English.

The paper has two parts:

1. a brief exposition of PTS, and
2. criticism of MTS as a theory of meaning and advantages of PTS as such a theory.

The PTS programme In a nutshell - I

– **Sentences:** replace the received MTS approach of taking their meanings as **truth-conditions** (in arbitrary models) by taking their meanings as **canonical derivability conditions** (from suitable assumptions).

- The derivability conditions are formulated in a "dedicated" *meaning-conferring* natural-deduction proof-system. (Francez&Dyckhoff, L&P 2010, Francez&Ben-Avi, jOS 2014).

– In a sense, the proof system should reflect the "use" of the sentences in the considered fragments, and should allow recovering pre-theoretic properties of the meanings of sentences such as entailment, assertability conditions and consequence drawing.

– Dummett introduces an important distinction between *content* and *ingredient sense*.

- The content of a sentence S is the meaning of S "in isolation", on its own.

- The ingredient sense of S is what S contributes to the meaning of an S' in which S occurs as a sub-expression.

- This distinction is incorporated in the PTS.

The PTS programme In a nutshell - II

Subsentential phrases: (down to lexical units): replace their MTS *denotations* (extensions in arbitrary models) as meanings by their **contributions** to the meanings (canonical derivability conditions) of sentences in which they occur.

– Adheres to Frege's *context principle*, made more specific by the incorporation into a type-logical grammar for the fragment considered. This is elaborated upon in Francez, Dyckoff &Ben-Avi, Studia Logica 2009.

- The distinction between contents and ingredient sense is propagated to subsentential phrases.

- A major success of PTS for NL is manifested in Francez&Ben-Avi, JOS 2014, where *conservativity* of all determiners is *proved* rather than stipulated as a universal.

Canonical derivations

Consider a meaning-conferring ND-system N for the NL-fragment, containing I/E-rules for the various constructs in the language.

– *Derivations* $\mathcal{D}$ (in tree-like form), are defined recursively in a standard way.

– canonical derivations play a central role in the definition of the proof-theoretic meaning.

– **canonical derivation from open assumptions:** A N-derivation $\mathcal{D}$ for deriving a conclusion S from (open) assumptions Γ is *canonical* iff it satisfies one of the following two conditions.

1. The last rule applied in $\mathcal{D}$ is an I-rule (for the main operator of ψ).
2. The last rule applied in $\mathcal{D}$ is an assumption-discharging E-rule, the major premise of which is some S' in Γ, and its encompassed sub-derivations $\mathcal{D}_1, \cdots, \mathcal{D}_n$ are all canonical derivations of S.

Denote by $\vdash^c_N$ canonical derivability in N and by $[\![S]\!]^c_\Gamma$ the collection of all (if any) canonical derivations of S from Γ.

Some simple rules

$$\frac{\Gamma, \mathbf{j} \text{ isa } X \vdash S[\mathbf{j}]}{\Gamma \vdash S[(\text{every } X)]}\,(eI) \qquad \frac{\Gamma \vdash S[(\text{every } X)] \quad \Gamma \vdash \mathbf{j} \text{ isa } X}{\Gamma \vdash S[\mathbf{j}]}\,(eE)$$

where $\mathbf{j}$ is fresh for $\Gamma, S[\text{every } X]$ in (eI).

- An instance of (eI): $\dfrac{\Gamma, \mathbf{j} \text{ isa girl} \vdash \mathbf{j} \text{ smiles}}{\Gamma \vdash \text{every girl smiles}}\,(eI)$

$$\frac{\Gamma \vdash \mathbf{j} \text{ isa } X \quad \Gamma \vdash S[\mathbf{j}]}{\Gamma \vdash \mathbf{j} \text{ isa } X \text{ who } S[-]}\,(relI)$$

$$\frac{\Gamma \vdash \mathbf{j} \text{ isa } X \text{ who } S[-] \quad \Gamma, [\mathbf{j} \text{ isa } X]_1, [S[\mathbf{j}]]_2 \vdash S'}{\Gamma \vdash S'}\,(relE^{1,2}),\ \mathbf{j} \text{ fr}$$

$$\frac{\Gamma \vdash \mathbf{j} \text{ isa } X \quad \Gamma \vdash \mathbf{j} \text{ is } A}{\Gamma \vdash \mathbf{j} \text{ isa } A\ X}\,(adjI)$$

$$\frac{\Gamma \vdash \mathbf{j} \text{ isa } A\ X}{\Gamma \vdash \mathbf{j} \text{ isa } X}\,(adjE_1) \quad \frac{\Gamma \vdash \mathbf{j} \text{ isa } A\ X}{\Gamma \vdash \mathbf{j} \text{ is } A}\,(adjE^{1,2})$$

- An instance of $(adjI)$ is $\dfrac{\Gamma \vdash \mathbf{j} \text{ isa girl} \quad \Gamma \vdash \mathbf{j} \text{ is beautiful}}{\Gamma \vdash \mathbf{j} \text{ isa beautiful girl}}\,(adjI)$

Meaning in PTS

– **inferentialism:** I-rules *determine* meanings!

– For a compound S, its *reified proof-theoretic meanings* is $[\![S]\!]^p =^{\text{df.}} \lambda\Gamma.[\![S]\!]^c_\Gamma$

– Note that the "denotational" meaning of S is *a proof-theoretic object*, a function from contexts to the collection of canonical derivations of S from that context.

– The role of canonicity: $\frac{\alpha \quad (\alpha{\rightarrow}(\varphi\wedge\psi))}{\varphi\wedge\psi}\,({\rightarrow}E)$

- a *non-canonical derivation of a conjunction*.

- The conjunction is *not derived according to its meaning!* It could mean anything.!

- The following *canonical* derivation *is* according to the conjunction's meaning.

$$\frac{\dfrac{\alpha \quad \alpha{\rightarrow}\varphi}{\varphi}\,({\rightarrow}E) \qquad \dfrac{\beta \quad \beta{\rightarrow}\psi}{\psi}\,({\rightarrow}E)}{\varphi\wedge\psi}\,(\wedge I)$$

$$\frac{\Gamma\vdash\mathbf{j}\ \text{isa girl} \qquad \Gamma\vdash\text{every girl isa beautiful girl}}{\Gamma\vdash\mathbf{j}\ \text{isa beautiful girl}}\,(eE)$$

- non-canonical, not according to adjectival modification meaning.

Criticism of MTS as theory of meaning: Manifestation

– There is a vast literature with critical arguments against MTS as a theory of meaning.
- I present here briefly only some of the main ones, those pertaining directly to NL.
- Some involve philosophical considerations, and others - not. My personal position is closely related to the latter sort of criticism.
– I. The most famous criticism is Dummett's *manifestation argument*, e.g., associating meaning of a sentence with the understanding of that sentence, manifesting itself as the ability (at least in principle) to verify the sentence as a condition for its assertability.
– Trans-verificational truth is rejected since it is not reflecting a cognitive process (the philosophical position of anti-realism);
- Rejection of *bivalence*, where every sentence is either true or false, independently of any ability to verify what that value is.
There are undecidable sentences!
- Contrasts the common situation where derivability in proof-systems is algorithmically decidable, due to the availability of terminating proof-search procedures.

Criticism of MTS as theory of meaning: Explanatory Power

– II. Another kind of criticism of MTS questions its *explanatory power*. The received wisdom regards MTS as a formalization of the relationship between language and the world.
- Quine relates to this view as "the museum myth": NL expressions are stuck on objects like labels in a great museum.
- The claim is that no theory can succeed in directly relating language to the world. At most, language is related to some meta-language (e.g., some set-theoretical language), used to specify models and truth-conditions in them.
- This is particularly relevant to the case of NL, which is its own ultimate meta-language.

– Since I find this criticism a very compelling one, independent of philosophical stand on metaphysical issues, I want to elaborate more on it.

Criticism of MTS as theory of meaning: Explanatory Power II

– Consider the usual MTS definition of conjunction ‘and’, using the usual models:

$M \models S_1 \text{ and } S_2 \text{ iff } M \models S_1 \textbf{ and } M \models S_2$.

- How does such a clause define the meaning of ‘and’?

- Unless the meaning of ‘**and**’ (in the meta-language, here English) is *already* known, this does not define *meaning* at all!

Otherwise, a similarly structured definition of a connective ‘blob’ would be equally well-defined by

$M \models S_1 \text{ blob } S_2 \text{ iff } M \models S_1 \textbf{ blob } M \models S_2$

Criticism of MTS as theory of meaning: Ontological Commitment

– III. One may feel some dissatisfaction with the *ontological commitment* accompanying MTS, relating to various entities populating models:
- possible-worlds, events (and their participants), properties, times, locations, degrees, kinds and many more.

– As emphasized by Paoli, when adhering to PTS, the definition of meaning need not appeal to any external apparatus; it can use the (syntactic!) resources provided by the rules of the underlying deductive system, which are artefacts of this system, devoid of any ontological commitment.

- A related issue, associated with entities in models, is the problematic possibility of quantifying over "absolutely everything", accompanying MTS (cf. Rayo and Uzquiano).

Criticism of MTS as theory of meaning: Granularity of Meaning

– IV. There is a criticisms of MTS as a theory of meaning, pointing to an advantages of PTS as such a theory, which is independent of cognitive and/or epistemic considerations, as well as from metaphysical ones.
– A notorious problem of MTS is its coarse granularity of meaning, where logically equivalent propositions, which have the same truth-conditions, are assigned the *same* meaning.
– Example: in propositional Classical Logic, we have the equivalence $\varphi \wedge \psi \equiv \neg(\neg\varphi \vee \neg\psi)$.
- Both sides of the equivalence are assigned the same meaning (here, same truth-table).
- However, those two proposition do differ in several aspects involving meaning, most notable in inference.

Criticism of MTS as theory of meaning: Granularity of Meaning - II

- It is fairly natural to regard a transition from $\varphi \wedge \psi$ to φ as "elementary"; a transition from $\neg(\neg\varphi \vee \neg\psi)$ to φ, while valid, cannot count as elementary, and its validity needs explanation by means of decomposition to more elementary steps.
- A fortiori, the same applies to more complicated, less transparent logical equivalences.
– In particular, in mTS all logical validities are assigned the same meaning.
- However, $[\![\varphi \rightarrow \varphi]\!]^p \neq [\![^p \varphi \vee \neg\varphi]\!]$.
– In natural language this discrepancy is even more salient. Identifying the meanings of every girl is a girl with that of every flower is a flower, and even with that of no bank is a non-bank, is clearly inadequate.
In PTS, directly appealing to inferential roles for conferring meaning, finer-grained meanings are obtained, not suffering from this deficiency.