I-Languages and T-sentences

Paul M. Pietroski

This paper is about the relevance of Liar Paradoxes for truth-theoretic accounts of linguistic meaning. In "Framing Event Variables," I develop an independent objection to such accounts. But in both essays, the first numbered sentence is the one displayed below.

(1) The first numbered sentence in "Framing Event Variables" is false.

Let 'Lari' be a name for the first numbered sentence in that other paper, Pietroski (2015). Since the definite description in (1) can be used to describe Lari, the grammatical subject of (1) can be used to talk about that very sentence. So (1) can be used to say that (1) is false. Likewise, (2)

(2) The second numbered sentence in "I-Languages and T-sentences" is not true.

can be used to say that it isn't true. In my view, this isn't paradoxical: it's true that (2) isn't true, and it's false that (1) is false. But such examples do tell against the Davidsonian thesis (D);

(D) for each Human Language **H**, some Tarski-style theory of truth for **H**
is the core of a correct theory of meaning for **H**

where a Human Language is a spoken or signed language that a biologically normal human child can acquire, given an ordinary course of experience. In section one, I say more about thesis (D). But details aside, many Human Languages permit sentences like (1) and (2), in which a word like 'true' is combined with devices that can be used to describe a linguistic expression and express negation. These sentences are grammatically akin to (3) and (4).

(3) The largest frog in Walden Pond is green.

(4) Kermit is not red.

Though as discussed below, it seems that any *truth theory* that accommodates (1) and (2) will be too sophisticated to serve as the core of a plausible *meaning theory* that applies to (3) and (4); and this concern is amplified if Human Languages are I-languages in Chomsky's (1986) sense.

In short, sentences like Lari highlight a deep tension between (D) and an independently plausible conception of Human Languages. I think the best response is to deny that (1) and (2) have truth conditions, and then admit that (3) and (4) also fail to have truth conditions. While sentences of a Human Language are often used to express truth-evaluable thoughts, it remains a hypothesis that these sentences are themselves truth-evaluable. And this hypothesis has implausible consequences; indeed, it leads to contradiction, given very plausible background assumptions. In my view, this tells against the idea that sentences of a Human Language have truth conditions, even if theorists can invent ways of evading or embracing the *reductio*.

1. The Bold Conjecture

Before returning to Liar Sentences, I want to set the stage with some discussion of Davidson's (1967b, 1984) influential proposal about how meaning is related to truth. In my view, thesis (D)

(D) for each Human Language **H**, some Tarski-style theory of truth for **H**
is the core of a correct theory of meaning for **H**

turns out to be about as plausible as the simpler thesis (D′).

(D′) there are correct Tarski-style theories of truth for Human Languages

Prima facie, (D′) is quite implausible. But one can try to motivate and defend (D′) as part of a package that also includes the following conditional claim: if (D′), then (D). By itself, this conditional is plausible; and (D) offers an attractive conception of linguistic meaning. Though as I'll argue in sections two and three, the possibility of Liar Sentences—i.e., the fact that Human Languages can generate such sentences—tells against conjoining (D′) with a conditional that leads from (D′) to (D). The net result, I claim, is that both theses remain implausible.

1.1 Truth Theories for Human Languages

Let Θ be a Tarski-style theory of truth for a given language if each sentence *S* of the language is such that Θ has a theorem of the following form: True(*S*, **c**) ≡ F(**c**); where '**c**' ranges over contexts, and '≡' indicates the material biconditional. The languages that Tarski discussed can be viewed as special cases, with vacuous relativization to contexts, allowing for theorems of a simpler form: True(*S*) ≡ p.[1] Correlatively, if we idealize away from context sensitivity, then a Tarski-style truth theory for English—or better, for any version of English whose sentences include (5) and (6)—will have boundlessly many theorems like (5T) and (6T).

(5) Ernie snores. (5T) True('Ernie snores.') ≡ Snores(Ernie)

(6) Bert yells. (6T) True('Bert yells.') ≡ Yells(Bert)

If at least one such theory is correct, at least in the sense that each of its theorems is true, then one might hope that some such theory can serve as the core of a theory that correctly specifies the semantic properties of all English expressions. But one might suspect that any Tarski-style theory of truth for English will generate incorrect theorems, even restricting attention to declarative sentences. And it is important to distinguish two kinds of skepticism.

One might concede that many English sentences have truth conditions, but suspect that certain constructions introduce special complications, with the result that some sentences either fail to have truth conditions or have truth conditions that preclude a correct Tarski-style theory. Alternatively, one might think that no English sentences have truth conditions, but that some constructions are especially useful for purposes of making arguments that apply more generally. My skepticism is of the latter sort. For example, I think (7) and (8) fail to have truth conditions.

(7) Bert thinks that Hesperus rose at dusk.

(8) Bert thinks that Phosphorus rose at dusk.

In this respect, I think (7) and (8) are like (1) and (2).

(1) The first numbered sentence in "Framing Event Variables" is false.

(2) The second numbered sentence in "I-Languages and T-sentences" is not true.

But my claim is not that (D) is false because of special words like 'think' and 'true'.

(D) for each Human Language **H**, some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

In my view, truth-theoretic semantics is wrong even for "fragments" of Human Languages. My aim is to argue for this general conclusion while bracketing doubts about specific examples. But (7) and (8) can be helpful as illustrations of initial points that carry over to "simpler" sentences.

It's easy to provide implausible theories. For example, given that Hesperus is Phosphorus, it's easy to provide theories according to which (7) is true whenever (8) is true. But prima facie, any such theory is wrong. This can make it tempting to say that for any context **c**, (7) is true relative to **c** if and only if Bert bears a certain relation **R** to sentence (9) relative to **c**,

(9) Hesperus rose at dusk.

and (8) is true relative to **c** if and only if Bert bears **R** to sentence (10) relative to **c**.

(10) Phosphorus rose at dusk.

The familiar idea is that depending on the context, a thinker might bear **R** to (9) but not (10)—or to (10) but not (9), or to both sentences, or to neither. But depending on how the alleged relation

---

[1] For these languages, truth can also be characterized in terms of satisfaction, without treating sentences and predicates as devices that denote truth values (⊤ or ⊥) and functions; see §2.3 below. But for other languages, a Tarski-style theory may characterize truth in terms of truth values and context-sensitive denotation conditions: True(*S*, **c**) ≡ Denotes(*S*, ⊤, **c**); cp. Frege (1892a), Church (1941), Montague (1974). And for these purposes, let's be generous about what contexts can be: sequences of entities, possible worlds, centered worlds, situations, etc.

is described, and how contexts are specified, the corresponding instances of 'True(*S*, **c**) ≡ F(**c**)' may be hard to evaluate—and in that sense, neither implausible nor plausible.

Let's focus on (7) and (9), repeated below, and consider a specific proposal:

(7) Bert thinks that Hesperus rose at dusk.

(9) Hesperus rose at dusk.

(7) is true relative to **c** if and only if Bert endorses the thought expressed with (9) relative to **c**. This biconditional, which employs technical notions, is not a truism.[2] Its "left side" concerns the truth of (7); its "right side" concerns Bert and a thought allegedly expressed with (9). To make an informed judgment about whether the biconditional is true, we need independent measures of whether each side is true, given various contexts. But how do we tell whether or not Bert endorses the thought expressed with (9) relative to a given context?

It's hard enough to say whether or not (7) is true relative to the contexts associated with certain episodes of communication, once the relevant notion of context becomes technical. But we can gather evidence about the semantic properties of (7) by describing scenarios, and asking competent speakers whether or not Bert thinks that Hesperus rose at dusk. For we can treat the responses—'yes', 'no', 'maybe', 'I don't know', etc.—as evidence of whether or not, in the conversational settings we have created, sentence (7) can be used to describe the scenarios correctly. So given some ancillary assumptions about contexts, and how to control for various pragmatic effects, we can draw conclusions about whether (7) is true relative to various contexts. Let's assume, for now, that such conclusions can be reliable. Indeed, let's pretend that for each context **c**, we know whether or not (7) is true relative to **c**. That still leaves the harder question: is (7) is true relative to **c** if and only if Bert endorses the thought expressed with (9) relative to **c**?

If we can't tell—because we we're not sure what it is for sentences to express thoughts relative to contexts—then providing a Tarski-style theory that assigns the alleged truth condition to (7) doesn't show that some such theory accommodates (7) in a plausible way. It's good for theories to avoid the implication that (7) is true if and only if Bert thinks that Phosphorus rose at dusk. But one wants a theory that can be *confirmed* by noting that its theorems seem to be *true*. And for these purposes, it won't help to exchange 'endorses the thought expressed with (9)' for 'thinks what (9) says' or 'samethinks with (9)' or 'thinks* (9)'; cp. Davidson (1967a), Segal (1989), Larson and Ludlow (1993). Instances of 'True(*S*, **c**) ≡ F(**c**)' invite a picture in which speakers of the relevant language can evaluate the left side (even if they don't have the technical knowledge required to evaluate the right side), and theorists can evaluate the right side (even if they don't have linguistic knowledge required to evaluate the left side); see Davidson (1967b, 1984). We assume that the speakers in question can understand the object language sentence on the left and make reliable judgments about whether it can be used to describe various scenarios correctly. But likewise, theorists need to understand the metalanguage sentence on the right and be able to make reliable judgments about whether it is true in those scenarios.

This doesn't show that any particular proposal is wrong, much less that "attitude reports" are counterexamples to all Tarski-style theories of truth for Human Languages. But given the range of mental states that (7) can be used to describe, and the many ways in which the truth of corresponding assertions can depend on details of the conversation, why think that (7) has a stable truth-theoretic "character" that maps contexts to truth or falsity? And if there are reasons for doubting that (7) has a truth condition, then absent countervailing considerations, why think that any unconfirmed claim of the form '(9) is true relative to **c** if and only if …**c**…' is true?

[2] Compare: Sadie is a mare if and only if Sadie is a mature female horse.

As discussed below, one can't just insist that relative to every context, (7)

(7) Bert thinks that Hesperus rose at dusk.

is true if and only if Bert thinks Hesperus rose at dusk. Disquoting doesn't ensure truth. If (7) doesn't have a truth condition, then (11) isn't true, not even relative to particular contexts.

(11) 'Bert thinks that Hesperus rose at dusk.' is true if and only if Bert thinks that Hesperus rose at dusk.

If it was clear that "simpler" sentences have truth conditions, then it might be reasonable to bracket doubts about propositional attitude reports, and conclude that (7) has a truth condition that is just harder to specify. But upon reflection, it isn't clear how to specify a correct instance of 'True(*S*, **c**) ≡ F(**c**)' for any sentence *S* of a Human Language.

Following Davidson (1967a) and many others, one might suggest that (9)

(9) Hesperus rose at dusk.

is true relative to a context **c** if and only if ∃e[Rise(e, Hesperus) & At-Dusk(e) & Past(e, **c**)]; where 'Rise(e, Hesperus)' indicates that e is a rising of Hesperus, and 'Past(e, **c**)' indicates that e occurred prior to the time of **c**. But to evaluate any such biconditional for truth or falsity, given a context, one needs to evaluate its right side. So in particular, 'Rise(e, Hesperus)' must be clear enough that theorists—who know that the sun doesn't rise, but that bread can rise—can say whether or not the world contains events that satisfy this open sentence. And given any particular construal of '∃e[Rise(e, Hesperus) & At-Dusk(e) & Past(e, **c**)]', one needs to *argue* that this invented sentence is true relative to **c** if and only if (9) is.[3] Similar remarks apply to (12).

(12) The sky is blue.

And while (6) is an even simpler sentence, it exhibits at least two significant complications.

(6) Bert yells.

One is that 'yells' suggests a habit of yelling.[4] This raises the question of whether a Tarski-style theory can plausibly accommodate both 'yells' in (6) and 'yell' in 'heard Bert yell' or 'make Bert want to yell'. Characterizing the relevant notion of habitual (or "generic") properties is notoriously difficult, however these alleged properties are related to meanings; see Leslie (2007) for discussion of examples like (13).

(13) Snowflakes are white, and mosquitoes carry diseases.

But even if we characterize a notion of being a chronic yeller, and then stipulate that the right side of (6T) is true if and only if the entity denoted by 'Bert' is a chronic yeller,

(6T) True('Bert yells.') ≡ Yells(Bert)

the right side of this biconditional still cannot be evaluated for truth or falsity, absent a further stipulation about which entity is denoted by the *metalanguage* expression 'Bert'. There are many Berts, none of which is *the* bearer of the proper noun 'Bert' that appears on left side of (6T); see, e.g., Burge (1973). So whatever 'Bert' denotes on the right, (6T) mischaracterizes (6), which is not about any particular Bert. Examples like (14) highlight other questions about names;

(14) France is hexagonal, and France is a republic.

cp. Austin (1961, 1962), Chomsky (1977, 2000), Pietroski (2005b). So instances of 'True(*S*, **c**) ≡ F(**c**)' are likely to be controversial if the object language sentence *S* contains a proper noun.

Of course, theorists can posit further context parameters—and perhaps an 'in the story' operator that would allow for correct uses of 'yells' in contexts where the relevant individual is a

[3] One can grant that verbs are associated with "eventish" variables without granting that verbs have Tarski-style satisfaction conditions; see Pietroski (2015, forthcoming).

[4] For simplicity, let's ignore "reporting uses" of the present tense, which introduces other complications.

muppet or some other object that cannot really yell. My claim is not that any particular sentences are clear counterexamples to the hypothesis that there a correct Tarski-style theory of truth for English. My claim is that sentences like (6) and (9) do not confirm this hypothesis,

(6) Bert yells.

(9) Hesperus rose at dusk.

are so they are not parade cases that motivate the hypothesis despite examples like (7) and (12).

(7) Bert thinks that Hesperus rose at dusk.

(12) The sky is blue.

Davidson (1967b) initiated a long and fruitful line of work. But so far as I know, there are still no parade cases of 'True(*S*, **c**) ≡ F(**c**)' where *S* is a sentence of a Human Language. Every proposed instance of this schema seems like a promissory note, at least to some extent. For many purposes, that's fine. But we shouldn't pretend that there are good inductive reasons for thinking that for every sentence of a Human Language has a truth condition. There is no model that yields plausible instances of 'True(*S*, **c**) ≡ F(**c**)' for elementary sentences, and that theorists have been progressively applying to more complicated constructions.

That said, induction is not the only way to motivate a generalization. If (D′) is true,

(D′) there are correct Tarski-style theories of truth for Human Languages

then as discussed in §1.2, (D) is at least plausible.

(D) for each Human Language **H**, some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

And if we assume (D), then (D′) becomes a lot more plausible. Put another way, while (D′) seems grossly implausible if assessed solely in terms of its implications for particular sentences, one can try to motivate and defend (D′) as part of a larger conception of linguistic meaning.

Given this project, it's reasonable to accept various promissory notes as theorists try to formulate truth theories that yield increasingly better instances of 'True(*S*, **c**) ≡ F(**c**)'. If an advocate of (D′) concedes that each sentence of a Human Language presents difficulties for this hypothesis, but claims that (D′) is still part of an attractive account of linguistic meaning, then the question is whether the difficulties outweigh the attractions of using truth theories as meaning theories. When addressing this question, it's reasonable to ask skeptics for concessions with regard to specific examples. But then one can't assume that (D′) is independently plausible, on empirical grounds, when faced with objections to the idea that truth theories can serve as meaning theories. In sections two and three, I will be pressing the following question: is any Tarski-style theory of truth for English sophisticated enough to be correct, given considerations that are easily illustrated with Liar Sentences, yet naïve enough to be the core of a correct theory of meaning for a language that children can acquire? In defending an affirmative answer, advocates of (D′) need not provide theories that yield plausible instances of 'True(*S*, **c**) ≡ F(**c**)' for specific examples. But neither can they assume that sentences of a Human Language *have* the truth conditions assigned by some Tarski-style theory.

I'll return to the idea defending (D) by rejecting (D′) and saying that an *incorrect* theory of truth for **H** can be the core of a correct theory of meaning for **H**; cp. Eklund (2002). But for now, let's bracket this idea, which comes with the burden of saying which false theorems of the form 'True(*S*, **c**) ≡ F(**c**)' are tolerable. I think that getting clear about *why* conjoining (D) and (D′) is implausible, even setting aside objections to (D′) that are based on specific examples, undercuts much of the motivation for (D). By being concessive about the difficulties that particular sentences present for (D′), I hope to show that (D) faces a general difficulty that gets manifested in many ways.

1.2 Truth Theories as Meaning Theories

As thesis (D) suggests, a mere truth theory is not yet a theory of meaning.

(D) for each Human Language **H**, some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

Ancillary assumptions are required; and not every truth theory for a Human Language can be supplemented in the requisite way. Davidson (1967b, 1984) held that only *suitably formulated* truth theories that meet certain *empirical constraints* can play this role. There are at least two clusters of reasons for saying this. The first, and perhaps most obvious, is that not all sentences are declarative. The second, and perhaps more important, was stressed by Foster (1976): given one correct truth theory for a Human Language **H**, there are boundlessly many such theories none of which is the core of a correct theory of meaning for **H**.

If a truth theory Θ is to serve as the core of a meaning theory for English, then Θ needs to be formulated so that it can be supplemented in ways that accommodate expressions like (15-20).

(15) Ernie snores? (16) Does Bert yell?
(17) When does Ernie snore? (18) Does Bert yell when Ernie snores?
(19) Please color Kermit green. (20) Don't color Kermit red!

One obvious thought, which can be developed in various ways, is that each sentence of a Human Language inflects a sentential "radical" with a grammatical mood; where the radical has a truth condition, and the mood defeasibly indicates the kind of speech act being performed.[5] The idea is that (15) and (5) share a *part* that is true (relative to a context **c**) if and only if Ernie snores (in **c**);

(5) Ernie snores.

while the difference in pronunciation reflects a difference in mood that is significant. Many theories go farther, treating interrogatives and declaratives as expressions that denote things of different sorts—e.g., questions and propositions. Perhaps a truth-theoretic conception of meaning must take some such form in order to adequately describe the systematic relations exhibited by matrix questions like (17), relative clauses as in (18), and embedded questions as in (21); see, e.g., Hamblin (1973), Karttunen (1977), Higginbotham and May (1981).

(21) Does Bert know when/where/why/whether Ernie snores?

But let's not fuss about this. Given a truth theory that can serve as a decent theory of meaning for English declaratives, it would be churlish to complain if the theory doesn't itself explain everything we would like to explain about the meanings of other sentences.

Still, a correct theory of meaning for a "declarative fragment" of a Human Language must be extendable to nondeclarative sentences. We can agree to set (15-21) aside for many purposes, just as we can to set (2) and (7) aside.

(2) The second numbered sentence in "I-Languages and T-sentences" is not true.
(7) Bert thinks that Hesperus rose at dusk.

But for any grammatical mood, to accommodate sentences with that mood in the way required by thesis (D), a truth theory must be formulated so that it can be embedded in a larger theory that accommodates sentences with other moods. Likewise, a correct theory of meaning for any "fragment of a declarative fragment" that excludes words like 'true' and 'think' must be formulated so that the theory can be extended to handle sentences that include such words. This turns out to be a nontrivial constraint, for reasons connected with "Foster's Problem."

Suppose that Ernie snores and Bert yells, and that the invented sentence (22) is thus true.

[5] See Dummett (1976) on the need for a theory of "content" that meshes with a theory of "force," regardless of how the relevant notion of content is related to classical notions of truth; see also Segal (1991), Ludwig & Lepore (2007), Lohndal & Pietroski (2011), and references there.

(22) Snores(Ernie) & Yells(Bert)

If we ignore context sensitivity, and pretend that the natural sentences (5) and (6) are also true,

(5) Ernie snores.

(6) Bert yells.

then the invented sentences (23) and (24) are true instances of ‘True(*S*) ≡ p’.

(23) True(‘Ernie snores.’) ≡ Yells(Bert)

(24) True(‘Bert yells.’) ≡ Snores(Ernie)

But (23) and (24) are *uninterpretive* in the following sense: the metalanguage sentence, used on the right side, is not a good translation of the object language sentence mentioned on the left. So a theory that generates (23) and (24), rather than (5T) and (6T),

(5T) True(‘Ernie snores.’) ≡ Snores(Ernie)

(6T) True(‘Bert yells.’) ≡ Yells(Bert)

is presumably not the core of a correct theory of meaning of English. Similarly, since neither (25) nor (26) is true, (27) is a true but uninterpretive instance of ‘True(*S*) ≡ p’.

(25) Every number is even.

(26) Five precedes two.

(27) True(‘Every number is even.’) ≡ True(‘Five precedes two.’)

And if the invented sentence (28) is false, then (29) is also true but uninterpretive.

(28) <(5, 2)

(29) True(‘Every number is even.’) ≡ <(5, 2)

Given any interpretive instance of ‘True(*S*) ≡ p’, there will endlessly many uninterpretive analogs that replace the instance of ‘p’ with a sentence having the same truth value. This is unsurprising, since ‘≡’ indicates a truth function, and sentence meanings are individuated more finely than truth values. But it raises the question of how any Tarski-style truth theory, even one whose theorems happen to be interpretive, can serve as a meaning theory for a human language. What would make any such theory more than a mere specification of truth/falsity for sentences of the object language?

As (25-29) suggest, it doesn’t help to stipulate that a truth theory is *modally correct* only if its theorems are true at every possible world. Moreover, even if examples involving mathematical/logical necessity are set aside as special cases, sentence meanings seem to be individuated more finely than sets of possible worlds; see Kripke (1982). If Hesperus is Phosphorus at every world where Venus exists, then (30) is true but uninterpretive.[6]

(30) True(‘Hesperus is Phosphorus.’) ≡ Identical(Venus, Venus)

If there are no worlds where Plato went back in time and killed all his ancestors, and also none where Edison invented a perpetual motion machine, then (31) is true but interpretive.

(31) True(‘Plato went back in time and killed all his ancestors.’) ≡
Edison invented a perpetual motion machine

Advocates of thesis (D) can posit worlds where (30) and (31) are false; see, e.g., Lewis (1986).

(D) for each Human Language **H**, some Tarski-style theory of truth for **H**
is the core of a correct theory of meaning for **H**

But to defend the idea that a modally correct truth theory will not have uninterpretive theorems, one needs to provide *independent* arguments for positing the requisite space of possible worlds.

[6] We can individuate expressions semantically, so that (30) is not obviously false at worlds where the pronunciation of ‘Hesperus is Phosphorus.’ has the meaning of (26). And we can say that each side of (30) has the same *content*; cp. Stalnaker (1984). But then it seems that meanings are not contents; see Pietroski (2006).

A different idea is that a truth theory with the right axioms will generate only interpretive theorems; see Davidson (1976). Stressing derivability, as opposed to modality, can seem more promising. And by aiming for axioms that generate "T-sentences" like (5T), as opposed to (23),

(5T) True('Ernie snores.') ≡ Snores(Ernie)

(23) True('Ernie snores.') ≡ Yells(Bert)

one might hope to specify theories that speakers naturally but tacitly deploy; see §2.3 below. But this strategy highlights three further questions for advocates of (D).

First, are there limits to *how technical* the right side of an interpretive T-sentence can be? Compare (5T) with (32), which is genuinely disquotational.

(32) 'Ernie snores.' is true if and only if Ernie snores.

The sentence used on the right side of (32) *is* the sentence mentioned on the left side; and this sentence of the object language—viz., (5)—is a good translation of itself.

(5) Ernie snores.

So (32) is interpretive. But one wants to know far T-sentences can depart from the disquotational paradigm, yet still be interpretive, given the need for significant *formalism* in the metalanguage. The importance of this question becomes obvious given examples that involve context sensitivity and quantification, as illustrated with (33), which isn't true if and only if I chased something.

(33) I chased something.

In the spirit of concession, let's grant that for any context **c**: True('I chased something.', **c**) ≡ ∃e∃x[See(e, Speaker[**c**], x) & Past(e, **c**)]; where the right side of this biconditional is an invented sentence that has a certain (stipulated) interpretation. And let's grant that some such T-sentence for (33) can be derived from plausible axioms, given appropriate *structural descriptions* of English expressions. But a theorist who says that his "formalized" T-sentences are interpretive, and thus relevantly like (32) as opposed to (23), owes some account of what distinguishes *interpretive but not disquotational* T-sentences from mere specifications of truth/falsity for sentences of the object language; see Lepore and Ludwig (2007) for further discussion.

One can reply that right sides of derivable T-sentences can exhibit "logical forms," so long as they don't employ notions that are "foreign" to the object language sentences, whose meanings may be obvious to competent speakers but not manifest in representations like (32). Though as we'll see, it is far from obvious that this reply is available to advocates of thesis (D).

(D) for each Human Language **H**, some Tarski-style theory of truth for **H**
is the core of a correct theory of meaning for **H**

For even if a truth theory does not generate implausible T-sentences, it may employ notions that are foreign to the languages that children naturally acquire.

This point is related to a second question: does any truth theory for a Human Language have enough "deductive power" to generate interpretive T-sentences, but not enough to generate uninterpretive analogs? Tarski envisioned a truth theory as an addition to a nonsemantic theory—e.g., of arithmetic—already formulated in a language that is governed by a logic, which licenses deduction of theorems from axioms. From this perspective, it isn't a problem if (29)

(29) True('Every number is even.') ≡ <(5, 2)

can be derived from the enlarged theory. On the contrary, it would be more worrisome if some analog of (29) *wasn't* derivable from a theory according to which: the right side is false; the left side is true if and only if every number is even; and it's not the case that every number is even. More generally, uninterpretive theorems seem unavoidable once a truth theory is embedded in a theory governed by a logic that licenses some inferences that preserve truth but not meaning. So one might wonder how any plausible theory can generate only interpretive T-sentences.

Perhaps the right response is that a truth theory for a Human Language has its own limited subject matter, and that at least for purposes of distinguishing interpretive T-sentences from uninterpretive analogs, a "pure" truth theory *isn't* part of a larger theory-of-the-world. I think it's odd to view theories of *truth* as segregated, in this way, from theories of the things that words and phrases are allegedly true of.[7] But let's not worry here about whether truth theories for Human Languages can be isolated from theories of numbers, spacetime, matter/energy, life, animal psychology, Human Language syntax, etc. In particular, let's grant that a truth theory for English need not be interwoven with premises that would support a derivation of (34) from (5T).

(34) True('Ernie snores.') ≡ Snores(Ernie) & Precedes(Two, Five)

(5T) True('Ernie snores.') ≡ Snores(Ernie)

Even this much concession still leaves the question of whether any theory Θ that generates interpretive T-sentences like (5T) can fail to generate uninterpretive T-sentences of the forms shown in (35) or (36); where **Γ** is itself a theorem of Θ.

(35) True('Ernie snores.') ≡ Snores(Ernie) & **Γ**

(36) True('Ernie snores.') ≡ **Γ** ⊃ Snores(Ernie)

The scope of this point depends on the relevant background logic. But consider a Fregean (second-order) logic that licenses instances of (37); where **A** is a theorem of arithmetic,

(37) HP ⊃ **A**

and 'HP' stands for (the conjunction of suitable Fregean definitions and) the Humean Principle that some things correspond one-to-one with some things if and only if the former have the same cardinality as the latter; see Heck (2011) for discussion. Replacing **Γ** in (35-36) with instances of (37) yields T-sentences that are wildly uninterpretive. One can try to formulate theories of truth against the background of a weak logic that does not generate interesting theorems of its own. But a truth theory that has (5T) and (6T) as theorems will also have (38) as a theorem

(5T) True('Ernie snores.') ≡ Snores(Ernie)

(6T) True('Bert yells.') ≡ Yells(Bert)

(38) True('Ernie snores.') ≡ Snores(Ernie) & [True('Bert yells.') ≡ Yells(Bert)]

if the theory permits replacement of a sentence on the right side of '≡' with the conjunction of that sentence and a theorem. So the background logic needs to be *very* weak. I'll return to this point in section two. But note that the requisite isolation of a truth theory is radical. Given some elementary principles of logic, a truth theory will generate uninterpretive T-sentences.

We are, recall, reviewing some questions raised by the following idea: if a truth theory for English has no uninterpretive theorems, then it can serve as the core of a theory of meaning for English. The first two questions concerned the plausibility of this conditional's antecedent. But as Foster (1976) discussed, one can ask a third question about the conditional itself.

Given a theory that generates (5T), one can easily construct an *extensionally equivalent* theory whose theorems include (39), where 'TRUTH' stands for any truth you like.

(39) True('Ernie snores.') ≡ Snores(Ernie) & TRUTH

For illustration, suppose that (5T) follows from a theory that has (40) and (41) as axioms,

(40) ∀x[TrueOf('snores', x) ≡ Snores(x)]

(41) ∀x[TrueOf('Ernie', x) ≡ (x = Ernie)]

along with some schema like (42), whose instances include (42a).

[7] It's even odder to then say, as Davidson (1986, p. 446) does, 'there is no boundary between knowing a language and knowing our way around in the world'. If uninterpretive T-sentences can be distinguished from theorems of meaning theories, then *pace* Quine (1951), there seems to be a theoretically interesting analytic/synthetic distinction.

(42) True(*NAME VERB.*) ≡ ∃x[TrueOf(*NAME*, x) & TrueOf(*VERB*, x)]

(42a) True('Ernie snores.') ≡ ∃x[TrueOf('Ernie', x) & TrueOf('snores', x)]

Let's not worry about the formal gap between (5T) and the conjunction of (40-42). The point is that however this gap is filled—so that (5T) is actually *derivable* from (40-42) via some suitably general principle—one can replace (40) with (43), and then derive (44).

(43) ∀x[TrueOf('snores', x) ≡ Snores(x) & Yells(Bert)]

(44) True('Ernie snores.') ≡ Snores(Ernie) & Yells(Bert)

Alternatively, one can replace (42) with (45), and then derive (44).

(45) True(*NAME VERB.*) ≡
∃x[TrueOf(*NAME*, x) & TrueOf(*VERB*, x)] & Yells(Bert)

So even if *some* truth theory for English has (5T) as a theorem,

(5T) True('Ernie snores.') ≡ Snores(Ernie)

this doesn't yet explain why (5) means what it does,

(5) Ernie snores.

since *another* truth theory has (44) as a theorem. At least one of these truth theories cannot be the core of a correct theory of meaning, since (5) and (46) differ in meaning.

(46) Ernie snores and Bert yells.

Both theories may generate (47). But the point remains that (44) is uninterpretive.

(47) True('Ernie snores and Bert yells.') ≡ Snores(Ernie) & Yells(Bert)

One can say that (43) and (45) are not interpretive axioms. Indeed, that's one way of putting Foster's point. But in my view, it's no defense of (D)

(D) for each Human Language **H**, some Tarski-style theory of truth for **H**
is the core of a correct theory of meaning for **H**

to say that a theory of truth can serve as a theory of meaning so long its axioms are interpretive. If a truth theory is to explain why Human Language expressions mean what they do, then we need some account of what makes (40-42) better than alternatives like (43) and (45), as *theoretical descriptions of* the semantic properties exhibited by the English expressions.

Davidson (1976) evidently agreed, since he suggested that uninterpretive axioms fail to meet certain *empirical* constraints, which he tried to specify in terms of what an idealized interpreter would ascribe to speakers on the basis of certain evidence. But many authors have objected to the verificationistic conception of meaning and language acquisition that Davidson adopted by appealing to what a "Radical Interpreter" would say; see Fodor and Lepore (1992), Pietroski (2005b), Ludwig and Lepore (2007). In my view, unabashed cognitivism is a better option. One can hypothesize that speakers of **H** have a language of thought in which they encode truth-theoretic axioms, from which consequences can be extracted in a constrained way. The idea is that axioms like (40-42) are better than (43) and (45) because competent speakers *encode and deploy* the former but not the latter; cp. Evans (1982).

Larson and Segal's (1995) system for deriving T-theorems is an explicit paradigm. It permits instantiation of axiomatic schemata that specify hypothesized contributions of phrasal syntax to phrasal meaning. But the rules for deriving further consequences, given axioms that specify the hypothesized contributions of lexical items, only permit replacements of *established* equivalents: *if* $\vdash P \equiv Q$ *and* $\vdash Q \equiv R$, *then* $\vdash P \equiv R$; and *if* $\vdash \alpha = \beta$, *then* $\vdash \Phi(\alpha) \equiv \Phi(\beta)$. So even if $\mathbf{\Gamma}$ is a theorem, the inference from (5T) to (35) is not licenced.

(5T) True('Ernie snores.') ≡ Snores(Ernie)

(35) True('Ernie snores.') ≡ Snores(Ernie) & $\mathbf{\Gamma}$

This reflects an interesting psychological hypothesis about the human capacity to apply semantic competence to particular expressions. (Heim and Kratzer's [1998] system can be spelled out in an analogous way.) One worry, developed in section two, is that (48) and (49)

(48) True('Lari is false.') ≡ False(Lari)

(49) True('Linus is not true.') ≡ ~True(Linus)

will be as derivable as (5T), regardless of what 'Lari' and 'Linus' name. But if the psychological hypothesis is correct, one can try to explicitly formulate the theories that speakers somehow encode and tacitly deploy. And as theorists progress towards this goal, they can say that for any Human Language **H**, a proposed truth theory Θ is *more interpretive* than alternative proposals if Θ *better* reflects how speakers of **H** represent the semantic properties of **H**-expressions.

The appeal to a language of thought abandons certain reductive ambitions for theories of meaning; cp. Dummett (1976). And by focusing on how speakers actually represent the alleged truth conditions of English sentences, one goes beyond the task of specifying a theory such that knowledge of it would suffice for knowing that English sentences have those truth conditions; cp. Davidson (1976), Foster (1976). But these are not yet objections to a cognitivist gloss of (D).[8]

(D) for each Human Language **H**, some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

So absent a better response to the questions about how a truth theory could ever serve as the core of meaning theory, let's grant that a cognitivist gloss of (D) provides a way of distinguishing interpretive from uninterpretive T-sentences.

At this point, we have conceded a lot. We're supposing that some theory Θ, yet to be explicitly formulated, is such that: Θ assigns truth conditions to English declaratives (modulo Liar Sentences) in a descriptively plausible way that accommodates the respects in which the alleged truth conditions are context sensitive; Θ can be supplemented with ancillary hypotheses in a way that accommodates nondeclarative sentences; and because Θ meets certain formal and empirical constraints, its theorems are interpretive. But (D) goes farther, and not merely because of the generalization to other languages. While an interpretive truth theory *represents* truth conditions in a special way, it still assigns *truth conditions* to sentences. And sentences of a Human Language, along with their constituent expressions, may be significant in other ways.

Davidson was inspired by invented languages whose expressions have no interpretive properties apart from the semantic properties characterized by truth theories. By stipulation, the expressions of such a language have no meanings that go beyond the truth-theoretic properties specified by some Tarski-style theory. But the expressions of a Human Language may—perhaps as devices for accessing and combining concepts in certain ways—have meanings that are *prior to* and often *presupposed by* their varied uses, including uses as tools for making truth-evaluable claims in contexts; see Pietroski (2005b, 2008, 2010, 2015 forthcoming). I think Liar Sentences point in this direction. They reveal a fundamental problem for the project of characterizing linguistic meanings in terms of truth.

---

[8] Some ambitions are unattainable, and some tasks are unduly modest. It may be hard to *find out* how speakers represent truth conditions. But there may be evidence of various kinds; see Evans (1982), Peacocke (1986) and Davies (1987), Lidz et.al. (2011). Heck (2007) offers a different cognitivist gloss on (D), drawing on Higginbotham (1991) to suggest that one overt truth theory for **H** is more interpretive than another if the former better reflects how speakers of **H** use their semantic competence in communication. I have doubts about emphasizing communication if Human Languages are I-languages in Chomsky's (1986) sense. But this debate is intramural. Like Higginbotham and Heck, I think the best defense of (D) is via some cognitivist gloss. Though in the end, cognitivism about meaning may be more plausible than (D).

2. Lies, Damned Lies, and Semantics

Ordinary humans have the capacities required to understand expressions of a Human Language. So if some truth theory Θ is the core of a correct theory of meaning for **H**, then ordinary humans can—and speakers of **H** do—understand expressions of **H** *as* expressions that have the properties specified by Θ. Liar Sentences remind us how hard it would be to understand expressions of a language like English as expressions that have the properties specified by a plausible truth theory. In §2.1, I review a simple version of the concern, ignoring context sensitivity and quantification. As we'll see in §2.3, sentences like (33) heighten the basic concern.

(33) I chased something.

2.1 The Problem, First Pass

Recall that (1) is also the first numbered sentence in the paper "Framing Event Variables,"

(1) The first numbered sentence in "Framing Event Variables" is false.

and that 'Lari' is a name for this sentence. Given plausible assumptions, Lari is false if Lari is true, and Lari is true if Lari is false. So prima facie, Lari is neither true nor false.[9]

This conclusion is not paradoxical. Many things are neither true nor false: frogs, numbers, rocks, stars, rock stars, etc. We can also invent languages whose sentences exhibit three truth values—⊤ for true, ⊥ for false, **N** for neither—where a sentence is false if and only if its negation is true; see Kleene (1950). But the *truth-evaluable* things, which may include many acts of judgment/assertion, may not include sentences of Human Languages.

Another possibility is that such sentences are typically true or false, relative to contexts, but that (1) and (50) are among some special cases that are neither true nor false.

(50) Lari is false.

If (50) is neither true nor false, then both sides of (48) are false, and so (48) is true.

(48) True('Lari is false.') ≡ False(Lari)

One might think that (51) is also true, along with its right side, and hence that (52) is true.

(51) True('Lari is not true.') ≡ ~True(Lari)

(52) Lari is not true.

So one might conclude that a truth theory for English can and should have theorems like (48) and (51). I disagree. In my view, (51) is false because (52) is neither true nor false; the right side of (51) is true, but its left side is false. More importantly, I don't think a theory of meaning for English should generate theorems like (48) or (51). Such a theory will go wrong in many ways.

Consider "strengthened" Liar Sentences like (2).

(2) The second numbered sentence in "I-languages and T-sentences" is not true.

If (2) is true, then (2) is not true. So (2) is not true. But if sentences of the form '… is not true' are among the truth-evaluable things, then presumably, (2) is true; and if a correct theory of meaning has theorems like (51), then presumably, sentences of the form '… is not true' are among the truth-evaluable things. I conclude that a correct theory of meaning for English should not have theorems like (51). But it is worth going slowly and being explicit about the relevance of these points for the idea that truth theories can serve as meaning theories.

---

[9] Let 'FEV' be a name for that other paper, and let 'iff' abbreviate 'if and only if'. If Lari is either true or false, then Lari is true iff the first numbered sentence in FEV is false. So *if Lari is true*: the first numbered sentence in FEV is false; and since Lari is that sentence, *Lari is false*. But *if Lari is false*: the first numbered sentence in FEV is false; and since Lari is true iff the first numbered sentence in FEV is false, *Lari is true*. So if Lari is true or false, then Lari is true and false. One can say that some claims are true yet contradictory; see Priest (1979, 2006). But even if that is so, it isn't yet a reason for concluding that Lari is both true and false. The hypothesis that Lari is true or false is much less plausible than the implied contradiction.

Suppose that Kermit, the largest frog in Walden Pond, is neither red nor blue. Then modulo context sensitivity, (4) is true if any sentences containing the word 'red' are true.

(4) Kermit is not red.

And if (53) is true, along with its right side, then (4) is true.

(53) True('Kermit is not red.') ≡ ~Red(Kermit)

So one might conclude that (4) is true, along with both sides of (53). Likewise, if Linus is a frog, then Linus is neither true nor false. In which case, (54) is true if (49) is true.

(54) Linus is not true.

(49) True('Linus is not true.') ≡ ~True(Linus)

And if Linus is a frog—or even if Linus is a sentence that isn't true—one might be tempted to say that both sides of (49) are true. But suppose that Linus is sentence (2).[10]

(2) The second numbered sentence in "I-languages and T-sentences" is not true.

Again, if Linus is true, then Linus is not true. So like Lari, Linus is not true. But since Linus is the second numbered sentence in this very paper, (54) is true if and only if (2) is true; and so (54) is true if and only if Linus is true. Hence, *(54) is not true*. No surprise there. But given who Linus is, (49) implies that the sentence 'Linus is not true.'—a.k.a. (54)—is true if and only if (2) is not true. And (2) is not true. So while (49) looks plausible, it implies that *(54) is true*. Hence, (49) is not true, and so no true theory has (49) as a theorem.

That's not a paradox. That's an argument against any theory that has (49) as a theorem. The right side of (49) is true, since Linus is not true. But the left side of (49) is false, since (54) is not true. That's OK: (49) is an *invented* sentence of a metalanguage that lets us express certain *theoretical* claims about Human Languages; and theoretical claims are often false.

Perhaps (55) is true, and (54) is not TruthEvaluable in the relevant sense. That's also OK.

(55) TruthEvaluable('Linus is not true.') ⊃
True('Linus is not true.') ≡ ~True(Linus)

One can say that (54) fails to be true in the way that Kermit fails to be true—viz., by not being the sort of the thing that has a truth condition. If (54) doesn't have a truth condition any more then Kermit does, then we need not (and should not) say that (54) is true if and only if Linus isn't true. You can say that (56) is true, since both sides of the biconditional are false.

(56) True('Linus is not true.') ≡ Precedes(Five, Two)

But don't replace the right side with a truth, as in (49), and try to maintain the biconditional.

(49) True('Linus is not true.') ≡ ~True(Linus)

Of course, if a theory of meaning has (51) as a theorem,

(51) True('Lari is not true.') ≡ ~True(Lari)

then it's hard to see how the theory could fail to have (49) as a theorem. So evidently, a theory of meaning had better not have (51) as a theorem. Likewise, even though (48) is true,

(48) True('Lari is false.') ≡ False(Lari)

it seems that a theory of meaning had better not have (48) as a theorem, on pain of having (51) and (49) as theorems. It doesn't *follow* that (51) is false. Theorists can still appeal to three-valued logics and say that Lari—a.k.a. (1)—has the third truth value **N**.

(1) The first numbered sentence in "Framing Event Variables" is false.

[10] There are many ways in which 'Linus' could end up being a name for (2). But let's explicitly introduce 'Linus' as a name for the second numbered sentence in this very paper, "I-Languages and T-sentences." Compare the introduction of 'Julius' as a name for whoever invented the zipper; see Evans (1982). For many purposes, we could use '(2)' instead of introducing of a name whose pronunciation does not connote the thing named. But the issue here concerns Human Languages. And '(2)', an invented numeric description, is not a name in any natural sense.

But absent reason for thinking that (51) follows from a good theory of meaning, why think (51) is true? Indeed, one might suspect the left side of (51) is just as false as the left side of (49). For if (54) doesn't have a truth condition any more than Kermit does, why think (52) is different?

(54) Linus is not true.

(52) Lari is not true.

These points ramify. If (52) doesn't have a truth condition, why think (50) is different?

(50) Lari is false.

Upon reflection, the truth of (48) provides no reason for thinking that (50) has a truth condition.

(48) True('Lari is false.') ≡ False(Lari)

Moreover, if a proposed theory of meaning has (57) as a theorem,

(57) True('Kermit is red.') ≡ Red(Kermit)

then it's hard to see how the theory could fail to have (48) as theorem. Likewise, if the theory has (53) as a theorem, then it's hard to see how the theory could fail to have (49) as theorem.

(53) True('Kermit is not red.') ≡ ~Red(Kermit)

(49) True('Linus is not true.') ≡ ~True(Linus)

This is an indirect argument that no correct theory of meaning for English has theorems like (53) and (57). It would be tendentious to *assume* that the left sides of these biconditionals are false, like the left sides of (48) and (49). But even if one can consistently maintain (53) and (57), it doesn't follow that one should do so. Perhaps the left sides of (53) and (57) are false—though not as obviously false as the left side of (49)—because (4) and (58) don't have truth conditions.

(4) Kermit is not red.

(58) Kermit is red.

Correlatively, we need not say that (59) is true.

(59) 'Kermit is red.' is true if and only if Kermit is red.

For we can say that (59) doesn't have a truth condition any more than (58) does.

Let me stress this last point. If (59) is a Human Language sentence—and so not confined to a written form governed by philosophical conventions—then it's not clear that (59) is *true*, as opposed to a sentence that can be *used to report* the fact that (58) can be used to make a *claim* that is true if and only if the Kermit in question is red in the relevant sense. It's often convenient to simplify, and say that (59) is true, perhaps modulo a little context sensitivity. But this doesn't warrant the assumption that instances of (60) are true, modulo context sensitivity;

(60) *S* is true if and only if S.

where 'S' is replaced by an English declarative and '*S*' is replaced with that declarative quoted. For even if 'modulo context sensitivity' can be cashed out plausibly for examples like (61),

(61) 'I am hereby quoting him.' is true if and only if I am hereby quoting him.

saying that instances of (60) are *true*, modulo context sensitivity, is tantamount to adopting (D).

(D) for each Human Language **H**, some Tarski-style theory of truth for **H**
is the core of a correct theory of meaning for **H**

Davidson and others may be right that *if* there are endlessly many true instances of some context-sensitive version of (60), then the obvious and perhaps only explanation for all those truths is that they follow from a finite number of more basic principles that constitute a truth theory for English. But the antecedent of this conditional is as much in question as (D) itself. Note that (60) has endlessly many instances like (62),

(62) 'Bert thinks that Ernie snores.' is true if and only if Bert thinks Ernie snores.

whose quoted sentence can be used to report (some of) what Bert thinks. *If* all these English biconditionals are true, they presumably follow from a finite number of basic principles that constitute a truth theory for English. But that's a terrible *argument* for the hypothesis that some theory of truth for English accommodates belief ascriptions. The argument takes immediate consequences of the hypothesis as premises, when those consequences are also at issue.

So while it might seem that (D) explains the "fact" that instances of (60) are true,

(60) *S* is true if and only if S.

modulo context sensitivity, this description of the explananda is tendentious. Liar sentences highlight this general point. For consider (63).

(63) 'Linus is not true.' is true if and only if Linus is not true.

This instance of (60) might seem to be true if Linus is a frog, or a sentence that has the third truth value **N**. But if Linus is sentence (2), then (63) looks like a counterexample to (60).

(2) The second numbered sentence in "I-languages and T-sentences" is not true.

2.2 Complicating Complicates

This makes it tempting to look for a theory with *analog* theorems like (49L) and (53L);

(49L) Legit('Linus is not true') ⊃ [True('Linus is not true.') ≡ ~True(Linus)]

(53L) Legit('Kermit is not red') ⊃ [True('Kermit is not red.') ≡ ~Red(Kermit)]

where both conditionals are true, but the antecedent of (49L) is false, and the antecedent of (53L) is true. Let's not worry here about how the technical notion 'Legit' gets spelled out—e.g., via fixed points (Kripke 1975), revisions (Gupta & Belnap 1997), or whatever. For even if all the troublemakers can be filtered out, two related concerns remain if the goal is to defend thesis (D).

First, it seems that any such filtering will yield consistency at the cost of explanation. If endlessly many expressions of the object language **H** are such that a proffered theory of meaning for **H** fails to assign interpretations to those expressions, then we seem to be left with the same kind of explananda that we started with. With regard to Human Languages, the theoretical task is to explain how humans understand linguistic expressions, not merely to explain how a finite mind could understand boundlessly many expressions. We want theories of how we understand the (boundlessly many) expressions that we do understand. If a theory of meaning accommodates 'chased every cow' and 'cow that saw brown dogs' but not 'every dog that chased brown cows', then it does not accommodate the shorter phrases in the right way. And if a truth-theoretic account of meaning classifies (4) but not (54) as truth-evaluable,

(4) Kermit is not red.

(54) Linus is not true.

that is worrisome, since (54) is as meaningful and comprehensible as (4) and (50).

(50) Lari is false.

The second concern is that theorems like (49L) and (53L) don't seem to be interpretive. One can *say* that competent speakers recognize the antecedent of (53L) as true, that they discharge it to obtain (53), and that (54) is somehow understood by analogical extension.

(53) True('Kermit is not red.') ≡ ~Red(Kermit)

(54) Linus is not true.

But if speakers can discharge the antecedent of (53L), then we need to revisit the concession that speakers tacitly deploy some truth theory Θ that is governed by a very weak logic that does not generate uninterpretive theorems like (36);

(36) True('Ernie snores.') ≡ **Γ** ⊃ Snores(Ernie)

where **Γ** can be a theorem of Θ. In this regard, it is important to remember that troublemakers like (54) cannot be filtered out in terms of their *grammatical* properties; see Kripke (1975), Parsons (1974).

Adapting an example from Kripke, suppose that each of nine people write something on a bit of paper, which is placed in an otherwise empty box called 'Bo'. Four of the people write (64), and four others write (65). The ninth person considers three options: (64), (65) and (66).

(64) 2 + 2 = 4

(65) 2 + 2 = 5

(66) Five bits of paper in Bo carry inscriptions of a sentence that is not true.

Once the last bit of paper is deposited, someone utters (66). Unlike the arithmetic sentences (64) and (65), (66) is context sensitive if only because it is tensed. But it's hard to see how any context sensitive element of (66) could track which sentence the ninth person chose to inscribe. Yet the *utterance* of (66) is false, true, or neither depending on this choice. In my view, (66) does not itself have a truth condition—not even relative to a context that determines the relevant bits of paper, and hence the relevant inscribed sentence. If the ninth person inscribed (64) or (65), it might seem that (66) is straightforwardly false or true. But if the ninth person inscribed (66), the problem becomes manifest. One can say that (66) is Legit in many but not all contexts. But if one offers a theory of meaning that generates theorems like (5L), with some context relativization,

(5L) Legit('Ernie snores.') ⊃ [True('Ernie snores.') ≡ Snores(Ernie)]

one needs to say how and under what conditions speakers discharge the antecedents of such conditionals without deploying knowledge that would also yield uninterpretive instances of (36).

(36) True('Ernie snores.') ≡ **Γ** ⊃ Snores(Ernie)

Lycan (2013) recognizes that there is no hope of providing a syntactic characterization of the Human Language sentences that are paradox-inducing (given uncontroversial assumptions) if these sentences have truth conditions. But in defense of (D), he offers a clever variant proposal.

(D) for each Human Language **H**, some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

According to Lycan, English is not an instance of **H**; "English" is the sum of unboundedly many sublanguages, each with a proprietary truth predicate. The idea is that in acquiring English, one acquires a core competence that can be extended with a kind of metacapacity. Think of the core competence as a capacity to generate (use, and comprehend) the expressions of a basic language **B** that is like one of the many versions of English, except for having no semantic vocabulary. The meta-capacity lets one deploy a capacity to generate the expressions of a given language λ in a way that generates them as expressions of a distinct language λ′ that contains a language-relative truth predicate 'T_{λ}', which applies to all and only the sentences that are true-in-**λ**.[11]

If sentences of λ′ cannot be true-in-λ, then no sentences of λ′ satisfy 'T_{λ}'. However, sentences of λ′ that are true-in-λ′ satisfy the language-relative truth predicate '$T_{\lambda'}$' that appears in the distinct language λ″. And so on. According to Lycan, the many truth predicates that are available to "English speakers" cannot be grammatically substituted for each other, even if they are all adjectives. But neither are these truth predicates mere homophones, since a common conceptual thread runs through the hierarchy. In this sense, Lycan adopts a version of the idea that (54) is understood by a kind of analogical extension of how (4) is understood.

(54) Linus is not true.

(4) Kermit is not red.

[11] Cp. Davidson's (1986) description of communication in terms of shared "passing" theories that speakers deploy, in contexts, by adjusting "prior" theories that may never be used without adjustments; see also Ludlow (2011).

This may not preserve the idea that each language in the hierarchy has a *truth* predicate. But advocates of (D) may not care, if only a few truth predicates are ever used in ordinary discourse.

A deeper concern is that on Lycan's view, sentences of λ can be *sentences of* λ′, but sentences of λ′ containing a word like 'true' cannot be true-in-λ. As discussed in section three, I think this conflicts with the idea that Human Languages are generative *procedures*. Chomsky (1986) spoke of "I-languages" in part to distinguish these procedures, characterized intensionally, from sets of expressions or external manifestations of internalized generative rules. Though as Lycan's discussion reminds us, Davidson held that a single *utterance* might be transcribed with /empedɔkliyz liypt/, classified by one interpreter as an utterance of the English sentence (67), and classified by another interpreter as an utterance of the German sentence (68).

(67) Empedocles leaped.

(68) Empedocles liebt.

If an utterance *u* can be true-in-English iff Empedocles leaped *and* true-in-German iff Empedocles loved, then *u* can be true-in-English and true-in-German. So perhaps an utterance of (67) can be an utterance of sentences of distinct languages, and likewise for an utterance of (54).

(54) Linus is not true.

But it isn't clear that an expression of *one* of my I-languages can also be an expression of *another one* of my I-languages. And it isn't clear that Lycan's λ and λ′ are distinct Human Languages, rather than arbitrarily distinguished sets of expressions. But let me return to this point after briefly discussing context sensitivity and the idea that truth theories for Human Languages will have theorems concerning *utterances*.

2.3 Sequence Relativity and Context Sensitivity

One might hope that the right account of context sensitivity and sentences like (33)

(33) I chased something.

will somehow help in dealing with sentences like (54) and (66).

(54) Linus is not true.

(66) Five bits of paper in Bo carry inscriptions of a sentence that is not true.

But as we'll see, the two main strategies for accommodating sentences like (33)—relativizing the truth of sentences to contexts, or specifying truth conditions for sentential utterances—*heighten* the worry that no truth theory for a Human Language **H** is the core of a correct meaning theory for **H**. Focusing on truth favors the second strategy, while focusing on meaning favors the first. But either way, the net result is not good.

We sometimes speak as if 'true' might itself be the central predicate in a theory of meaning for a Human Language, or at least for an invented language that provides a decent model of a Human Language. But this is a simplification, as Tarski's own semantics for the first-order predicate calculus reminds us. One can invent a language in which (69-71) are sentences,

(69) $\forall x \exists x' Pxx'$

(70) $\exists x \forall x' Pxx'$

(71) Pxx'

and then provide a semantics according to which: (69) is true if everything precedes something, and otherwise (69) is false; (70) is true if something precedes everything, and otherwise (70) is false; (71) is neither true nor false, but not because it has a neutral truth value. In a *propositional* calculus, the truth or falsity of atomic sentences determines the truth or falsity of complex sentences. But Tarski showed how to characterize truth, for a *predicate* calculus, in terms of a more basic notion of satisfaction; where closed sentences like (69) and (70) have satisfaction conditions that are specified in terms of the satisfaction conditions of open sentences like (71).

Truth is satisfaction by all sequences. So one T-sentence for (69) is simply (69a).

(69a) True(∀x∃x′Pxx′) ≡ for each sequence σ, Satisfies(σ, ∀x∃x′Pxx′)

If you like, add the syntactic requirement that only closed sentences are true. But the work of the "truth" theory is done by the axioms that license derivations of "S-theorems" like (69b);

(69b) Satisfies(σ, ∀x∃x′Pxx′) ≡ each sequence σ* such that $\sigma^* \approx_x \sigma$ is such that some sequence σ** such that $\sigma^{**} \approx_{x'} \sigma^*$ is such that σ**(x) precedes σ**(x′)

where for any sequence F and variable *i*, F(*i*) is the *i*th element of σ for each variable *i*, and for each sequence F*, $F^* \approx_i F$ if and only if F* differs from F at most with respect to the *i*th element. Let variables be characterized recursively: 'x' is the first; and for each variable α, α′ is the next.

One might wonder if any analogous S-theorems for sentences like (72) are interpretive.

(72) Every dog chased a cat.

Even if (72a) is true, its right side may employ notions that are "foreign" to (72).

(72a) True(72) ≡ for each sequence σ, each sequence σ* such that $\sigma^* \approx_x \sigma$ and σ*(x) is a dog is such that some sequence σ** such that $\sigma^{**} \approx_{x'} \sigma^*$ and σ**(x′) is a cat is such that σ**(x) chased σ**(x′)

And given sentences like (73-75), whose meanings are not "firstorderizable,"

(73) Most of the dogs chased at least half of the cats.

(74) For every paper that was accepted, nine were rejected.

(75) Some critics admire only one another.

specifying satisfaction conditions requires more technicalia; see Rescher (1962), Wiggins (1980), Boolos (1998). But let's continue to assume that some theory of truth will yield interpretive theorems, perhaps because speakers tacitly deploy a truth theory that is encoded in a language of thought whose predicates include SATISFIES(σ, S) and TRUE(S); where small capitals indicate mental analogs of the overt technical vocabulary.

As Heck (2004) notes, 'true' may express "the very concept of *truth* that plays a central role in the semantic theory we tacitly know (p. 343);" cp. Larson and Segal (1995). Heck also shows that even if the mental language is the relevant metalanguage, very modest assumptions about TRUE(S) still lead to paradox. So writing truth theories in mentalese is not a panacea. But let's grant that humans have a mentalese predicate TRUE(S), often expressed with 'true', and that our linguistic competence includes a mentalese predicate like SATISFIES(σ, S).

That still leaves the question of *how* to accommodate sentences like (76) and (33).

(76) It is here.

(33) I chased something.

But such sentences can be modeled with a Kaplanian extension of a Tarskian language; see Kaplan (1978a, 1978b, 1989). Add three indices—*s*, *p*, *t*—and a pointer π such that for each pointer α, α′ is also a pointer. Then extend each Tarskian sequence by adding a "zeroth" element that is an ordered triple of domain entities. Each extended-sequence, or *assignment* of values to variables, is of the form: $<<e^s, e^p, e^t>$, e, e′, e″, ...>. For each assignment **A**: **A**(*s*)/**A**(*p*)/**A**(*t*) is the first/second/third element of **A**'s zeroth element; **A**(π) = **A**(x), **A**(π′) = **A**(x′), etc. This allows for sentences like 'Lπ*pt*' and the S-theorems indicated in (76K) and (33K).

(76K) Satisfies(**A**, Lπ*pt*) ≡ **A**(π) is located in **A**(*p*) at **A**(*t*)

(33K) Satisfies(**A**, ∃x(Bx*t* & ∃x′(Cxsx′)) ≡ some **A*** such that $\mathbf{A}^* \approx_x \mathbf{A}$ is such that **A***(x) occurs before **A***(*t*), *and* some **A**** such that $\mathbf{A}^{**} \approx_{x'} \mathbf{A}$ is such that **A****(x) is a chase by **A****(*s*) of **A****(x′)

As desired, sentences like 'Lπpt' are not true or false simpliciter, since they are satisfied by some but not all assignments. But one can introduce a notion of truth relative to contexts that *select* assignments. One can say that a context **c** selects assignment **A** if and only if: **A**(*s/p/t*) is the speaker/place/time of **c**; and for each pointer π^i, $\mathbf{A}(\pi^i)$ is the *i*th thing demonstrated in **c**. That makes room for (77), according to which 'Lπpt' is true relative to some but not all contexts.

(77) True(*S*, **c**) ≡ for each assignment **A** such that Selects(**c**, **A**), Satisfies(**A**, *S*)

On this view, T-sentences still concern the context-relative truth conditions of *sentences*. But following Davidson (1967a), one might think that theories of truth for Human Languages will have theorems concerning *utterances*, perhaps along the lines of (76D) and (33D);

(76D) True(*u*, Lπpt) ≡ the thing demonstrated with the (first) deictic act in *u* is located in the place of *u* at the time of *u*

(33D) True(*u*, $\exists$x(B*xt* & $\exists$x′(Cxsx′)) ≡ some **A*** such that $\mathbf{A^*} \approx_{x} \mathbf{A}$ is such that **A***(x) occurs before the time of *u*, and some **A**** such that $\mathbf{A^{**}} \approx_{x'} \mathbf{A}$ is such that **A****(x) is a chase of **A****(x′) by the speaker of *u*

see Burge (1974), Larson and Segal (1995), Lepore and Ludwig (2005). For many purposes, the two approaches are interchangeable. But here, I want to stress a respect in which the K(aplan)-strategy and the D(avidson)-strategy reflect fundamentally different ways of thinking about how Human Languages and their users are related to truth and contexts.

On the K-strategy, a theory of truth for a language is a theory of a *language* in a sense familiar to logicians and grammarians: a *set* of well-formed formulae, specified by some (typically recursive) procedure; or a physically realizable *procedure* that generates expressions of some kind from finitely many lexical items and combinatorial principles. As discussed below, I follow Chomsky (1986) in taking Human Languages to procedures rather than sets. But the important point is here that on the K-strategy, theories of truth are theories of linguistic objects. Since the use of a sentence on a particular occasion determines which assignments are licensed on that occasion, truth depends on use. But the context-sensitive truth conditions of sentences are specified without reference to use. So a K-style theory ascribes truth conditions to *sentences*.

By contrast, on the D-strategy, a truth theory for a language turns out to be a theory of certain *actions of using* the language. Such theories ascribe truth conditions to spatiotemporally located *utterances of* generable sentences. In my view, this strategy is misguided, and not only because it is hard to formulate axioms from which biconditionals concerning utterances *follow*, given a weak logic that avoids uninterpretive theorems. (While Larson and Segal [1995] endorse a D-strategy in their text, their derivations of theorems usually employ the K-strategy.) Given the independent need for sequence relativization, one might hope for a theory that delivers theorems like (76K) and (33K) above; these are hard enough to derive. But more importantly, specifying meanings in terms of utterances seems to confuse linguistic competence with language use.

That said, the D-strategy can seem reasonable in other respects. With regard to Human Languages, it may well be that truth is more plausibly ascribed to certain *actions of using* sentences than to sentence-context pairs. Indeed, "Contingent Liar Sentences" like (66)

(66) Five bits of paper in Bo carry inscriptions of a sentence that is not true.

illustrate an important general point: falsity requires more than grammaticality, sincerity, and absence of truth. Making a false claim requires a significant kind of success, in that the world must conform to certain presumptions. Since attempts to name or demonstrate can fail, (78)

(78) Vulcan is bigger than that.

may not be truth-evaluable in each context of use; cp. Strawson (1950), Evans (1982). And there are other kinds of "presupposition failure." (A speaker who says that Harry is bald may wrongly assume that Harry is not a vague case.) Perhaps linguistic *expressions* carry presuppositions, and many derivable specifications of truth conditions are correspondingly conditional. But while some presuppositions of truth/falsity are grammatically constrained, it is far from obvious that all of them can or should be captured by a theory of linguistic meaning; see Schlenker (2009).

Prima facie, some constraints on truth-evaluability are constraints on *actions of using* sentences in certain ways. And examples like (66) suggest that even if these limits can be systematically described, they are not principles governing how humans *understand* linguistic expressions. Rather, it seems that the true/false distinction—like some other right/wrong distinctions—targets actions that meet the conditions for being evaluable a certain way. The actions may be uses of sentences. But the sentences used need not be truth-evaluable. For some purposes, it does no harm to simplify and say that Human Language sentences are true or false relative to contexts. We can also speak of "propositions" that are unrelativizedly true or false; see, e.g., Cartwright (1962). Given episodes of thinking, we can speak of thoughts thought; given episodes of asserting/endorsing/judging, we can speak of propositions asserted/endorsed/judged. But it doesn't follow that truth-evaluable things are truth-evaluable by virtue of having propositional contents, much less that theories of meaning should specify such contents.

If truth is downstream from meaning—in that certain uses of meaningful expressions are *candidates,* subject to review, for being true or false—then good theories of meaning for Human Languages may not deploy the notion of truth even if humans enjoy a mental predicate TRUE(_) that applies to certain events or mental sentences. That leaves the hard task of describing the context sensitive conditions on being true or false, and explaining why these are conditions on truth/falsity; though see Glanzberg (2004, forthcoming) for a good start. In any case, one can share Davidson's suspicion that any plausible theory of *truth* for a Human Language will have theorems that specify properties of spatiotemporally located actions of using expressions, while also suspecting that a plausible theory of *meaning* for a Human Language will have theorems that specify use-independent properties of expressions; where these properties constrain how the expressions can be understood, and hence how the expressions can be naturally used.

3. Human Languages as I-Languages

If we want to compare Human Languages with invented formal systems, or systems of animal communication, it is useful to begin with a generous conception of language. Then we can ask what kind(s) of languages children naturally acquire. This highlights the need to distinguish conventionally determined sets of expressions, or utterances thereof, from biologically implemented procedures that generate expressions. And if Human Languages are biologically implemented procedures, as Chomsky (1986) urges, then (D) must be evaluated accordingly.

(D) for each Human Language **H**, some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

3.1 I Before E

Let's be very generous, and say that a language associates signals of some kind with interpretations of some kind. An expression of a language pairs a signal (type) with an interpretation (type). An expression might pair a certain sound or gesture or inscription with a certain object or property or concept. Sounds, of a sort that can be produced by a mezzo soprano or a basso profundo, might be paired with concepts via some algorithm. Or expressions might pair instructions for how to generate complex gestures with instructions for how to generate complex concepts, thereby pairing (instruction-relative) gesture types with concept types.[12] This

[12] For these purposes, instructions include strings of '1's and '0's used in a von Neumann machine to access other such strings and perform certain operations on them: '0100110101' might be executed by performing *operation two*

allows for bee languages, languages of thought, mathematical languages with gothic scripts, and various conceptions of human phonology/semantics. It also allows for languages of various ontological sorts: *sets* of expressions; *procedures* that generate expressions; physical *implementations* of such procedures; classes of similar sets, procedures, or implementations; etc.

A language can have finitely many expressions. But a Human Language has endlessly many complex expressions that pair certain sounds or gestures—let's call them pronunciations—with certain interpretations that we can call meanings. The pronunciation-types are abstract, as opposed to types characterized acoustically or in terms of perceptible bodily motions. Meanings seem even less tied to physically characterizable spatiotemporal particulars. This suggests that Human Language expressions link phonological representations to semantic representations, even if represented signals and thereby paired with represented aspects of the environment. But in any case, a child who acquires a Human Language evidently acquires a *procedure that generates* unboundedly many expressions, each of which pairs a pronunciation with a meaning.

These procedures allow for homophony, but in constrained ways. In spoken English, the sound of 'bear' has more than one meaning, as do the sounds of 'duck' and (79).

(79) The duck was ready to eat.

But (79) is homophonous even if we hold the meaning of 'duck' fixed. It can mean that duck was ready to dine, or that the duck was *prêt-a-manger*. By contrast, (80)

(80) The guest was eager to please.

has the meaning indicated with (80a) *but not* the equally coherent meaning indicated with (80b).

(80a) The guest was eager that he please relevant parties.

(80b) #The guest was eager for relevant parties to please him.

Moreover, (81) fails to have the meaning indicated with (81a).[13]

(81) The guest was easy to please.

(81a) #It was easy for the guest to please relevant parties.

(81a) It was easy for relevant parties to please the guest.

And while (82) is ambiguous, it fails to have the third meaning indicated with (82c).

(82) The doctor called the lawyer from Las Vegas.

(82a) The doctor called the lawyer, and the lawyer was from Las Vegas.

(82b) The doctor called the lawyer, and the call was from Las Vegas.

(82c) #The doctor called the lawyer, and the doctor was from Las Vegas.

In general, each expression of a Human Language **H** pairs a meaning μ with a pronunciation π such that for some number n, **H** pairs π with n but not n+1 meanings. So a child who acquires **H** acquires a procedure that generates unboundedly many expressions *in a certain constrained way*.

Following Chomsky (1986), I think Human Languages *are* generative procedures that humans can acquire and use. The targets of inquiry in this vicinity—the "natural kinds," as opposed to objects of stipulation—seem to be the expression-generating procedures that children can naturally acquire, and more deeply, the distinctly human capacity to acquire and use such procedures given a relatively brief and limited course of ordinary human experience. So I take Human Languages to be biologically implementable generative procedures, as opposed to other things that are somehow related to expression-generators; cp. Katz (1981), Soames (1984). But

(010) on the *number* (0) *fifty-three* (110101), while '1101011010' calls for *operation six* (110) on the *number stored in* (1) register *twenty-six* (011010). And one can imagine operations like conjunction on accessible/generable concepts. For further discussion, see Pietroski (2013), from which some of this subsection is adapted.

[13] See Chomsky (1965), Higginbotham (1985). For extended discussion in the context of broader issues concerning language acquisition and the nature of Human Languages, see Crain and Pietroski (2001), Berwick et.al. (2011).

let me briefly address the contrasting idea that children acquire *sets of* expressions, which pair pronunciations with meanings in conventionally governed ways; see, e.g., Lewis (1975).

Even if each child acquires an infinite set of expressions *by* acquiring a generative procedure, and children converge on similar procedures despite variance in experience, one might prefer to focus on generated expressions (or utterances thereof). For one might be more interested in the possibility of communication than in Human Language acquisition; and it is logically possible that other minds could come to pair each English pronunciation with its meaning(s) via some inhuman procedure that still allows for successful communication. It is also logically possible that two "speakers of English" generate the same pronunciation/meaning pairs in different ways. So one might also hope to describe possible interpretations that any rational communicators could agree on, abstracting from how humans assign these publicly available construals to pronunciations. But this hope is not an argument that human linguistic communication is supported by shared (languages that are) *sets of expressions* as opposed to *generative procedures*. One can stipulate that two people "share a language" if and only if they can communicate linguistically. But then positing shared languages does not *explain* successful communication; and appeal to Human Languages, which may not be languages in the stipulated sense, may be required to explain how humans can share languages in the stipulated sense.

Moreover, children acquire expression-generating procedures that exhibit striking commonalities *not* due to the environment; see note 13. So one needs reasons for identifying Human Languages with any alleged sets of expressions that might be generated in various ways. In thinking about whether such reasons are likely to emerge, it is useful to follow Chomsky (1986) in applying Church's (1941) intensional/extensional distinction—regarding mappings from inputs to outputs (cp. Frege [1892b])—to the study of Human Languages and the aspects of cognition that support acquisition of these languages.

We can think of functions as *procedures* (intensions) that determine outputs given inputs, or as *sets* (extensions) of input-output pairs. Consider the set of ordered pairs $<x, y>$ such that x is a whole number, and y is the absolute value of x–1: {... (−2, 3), (−1, 2), (0, 1), (1, 0), (2, 1) ...}. This set can be characterized in many ways. Using the notion of absolute value, one can say that $F(x) = |x - 1|$. One can also use the notion of a positive square root: $F(x) = {}^{+}\sqrt{}(x^2 - 2x + 1)$. These *descriptions* of the set correspond to distinct procedures for computing values given arguments. And a mind might be able to execute one algorithm but not the other. In Church's idiom: we can use lambda-expressions to indicate functions-in-intension, with $\lambda x.|x - 1|$ and $\lambda x.{}^{+}\sqrt{}(x^2 - 2x + 1)$ as distinct but extensionally equivalent procedures; or we can use lambda-expressions to indicate functions-in-extension, saying that $\lambda x.|x - 1|$ is the same set as $\lambda x.{}^{+}\sqrt{}(x^2 - 2x + 1)$. Though as Church stressed, specifying the space of *computable* functions requires talk of procedures.[14]

Echoing this point, Chomsky contrasted I-languages with E-languages: an I-language is a procedure that pairs signals with interpretations; an E-language is a language in any other sense—e.g., a set of signal-interpretation pairs, or a cluster of dispositions to make certain utterances. Note that to even *specify* a set with endlessly many elements, one must somehow

[14] For Frege (1892b), functions as procedures are logically prior to the more set-like *courses of values of* functions. Marr (1982) likewise distinguished "Level One" descriptions, of functions computed, from "Level Two" descriptions of the algorithms employed to compute those functions. And while Level One descriptions can have a certain primacy in the order of discovery, making it fruitful to ask what a system does before worrying about how the system does it, a function cannot be computed without being computed in some way. So one must not confuse the methodological value of Level One descriptions, in characterizing certain cognitive systems, with any suggestion that the corresponding extensions are themselves targets of inquiry.

specify a procedure that determines the set. (In the previous paragraph, I was able to talk about *the* infinite set only because the context included reference to a determining procedure.) And while many procedures determine sets—arithmetic procedures defined over numbers being paradigm cases, along with invented procedures that generate sets of well-formed formulae—a biological system might pair pronunciations with meanings, yet *not* determine any set of expressions, if only for lack of a fixed domain of inputs. Consider (83).

(83) *The child seems sleeping.

Speakers of English know that this string of words is defective, but meaningful. In particular, speakers hear (83) as having the interpretation of (83a) *and not* (83b);

(83a) The child seems to be sleeping.

(83b) #The child seems sleepy.

see Chomsky (1965), Higginbotham (1985). The defect does not preclude *understanding* (83), which is neither word salad like (84)

(84) *be seems child to sleeping the

nor a grammatical expression like (85) that is understood as expressing a bizarre thought.

(85) Colorless green ideas sleep furiously.

So a Human Language **H** can, as an implemented procedure used in comprehension, assign a meaning to the pronunciation of (83). But it doesn't follow that (83) is an expression of **H**.

We can and perhaps should introduce a graded notion of expressionhood. This does not threaten the idea that Human Languages are generative procedures, which can be used in many ways. But it does challenge the idea that such procedures determine sets of expressions, in any theoretically interesting sense. If (83) is a second-class expression of my I-language, is (84) an especially degenerate expression, or not an expression at all? Are translations of (85), in Japanese or Walpiri, terrible expressions of my idiolect? The point here is not merely that the "expression of" relation is *vague* for Human Languages. It's rather that there is no way specify—absent arbitrary stipulations—what it is to *be* an expression of a Human Language **H**, to any given degree, without relying on a *prior* notion of **H** as a generative procedure that somehow interfaces with other cognitive systems that can deal with pronunciations and meanings (perhaps by treating them as instructions for how to produce certain signals and concepts).

I have stressed this conception of what Human Languages *are*, because if Chomsky's hypothesis (C) is correct, then advocates of (D) need to defend (CD).

(C) each Human Language is an I-language

(CD) each Human Language **H** is an I-language such that some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

And in terms of the discussion in §2.3, it seems implausible (to me) that a theory of truth *for an I-language* will have theorems concerning *utterances*. This seems to confuse the distinction that Chomsky (1986) was emphasizing when he contrasted questions concerning the linguistic knowledge/competence that children acquire with questions concerning how that knowledge/competence is put to use. Given (C), it also strikes me as optimistic to think that Human (I-)Language sentences—expressions that children can generate and comprehend—are themselves truth-evaluable. Like many philosophers in the Rationalist tradition, I find it amazing that humans can make *any* claims that rise to the level of falsity: once a thinker achieves this level of clarity and contact with her environment, *truth* is a mere negation sign away. Obviously, humans can often use their I-languages to make truth-evaluable assertions and judgments. That's a wonderful thing, but it may require many ancillary cognitive capacities (and in many domains, hard cognitive work); see Pietroski (2010, forthcoming).

One can hypothesize that humans enjoy an ancillary system that specifies truth conditions of *uses of* I-language expressions that *already* pair pronunciations with meanings, which are to be specified truth-theoretically via some version of the Kaplan-strategy for accommodating context sensitivity. This interesting hypothesis embraces (C). But on this view, one can't deal with Liar Sentences by saying that truth and falsity are really evaluative properties of spatiotemporlly located assertions/judgments, and that theories of truth for Human Languages assign semantic properties to utterances *rather than* sentences relativized to contexts.

3.2 Back to Troublemakers

Recall Lycan's (2013) suggestion that a sentence like (4)

(4) Kermit is not red.

is a sentence of both a "core" language λ that contains no semantic vocabulary and an expanded language λ′ that contains a predicate satisfied by all and only the sentences that are true-in-λ. A more expanded language λ′′ contains a predicate satisfied by all and only the sentences that are true-in-λ′. And so on. The idea is that (4) can be a true sentence of each of these languages. Given that Kermit is not red, (4) is true-in-λ, true-in-λ′, etc. By contrast, on Lycan's view, (54a)

(54a) Linus is not true(-in-λ).

is not an expression of λ, and hence is not true-in-λ.[15] Though given that Linus is not true, or least not true in λ or λ′, (54a) is true-in-λ′, true-in-λ′′, etc. If λ and λ′ are *sets* of expressions, with λ as a proper subset of λ′, this proposal is coherent: whatever (4) is, it can be an element of both sets while (54a) is an element of λ′ but not λ. But if λ and λ′ are Human *I-Languages*, then it isn't clear how the suggestion is supposed to work.

Suppose that (4) and (54a) are generable expressions—certain pronunciation/meaning pairs—and that these expressions are generated by some procedures but not others. Let λ be a procedure that generates (4) but not (54a). Let λ′ be a procedure that generates both. So far, this is just an intensional version of the idea that λ is a proper subset of λ′. But if children who "acquire English" acquire both λ and λ′, then it is hard to see what λ can *be* apart from an arbitrarily restricted variant of λ′: the same generative procedure, apart from some lexical items included in the specification of λ′. And it is very hard to see how λ′ can be an I-language distinct from λ′′, which allegedly includes a truth predicate that is pronounced 'true' but not a lexical item of λ′. (It is very very hard to see how λ′′ can be an I-language distinct from λ′′′, and so on.)

Lycan is surely right that 'English' is not a name for any particular Human Language. Even a single speaker of Standard American English may have acquired various I-languages that are similar but not identical for communicative or sociopolitical purposes. A person can be monolingual in the ordinary sense, yet have several overlapping I-languages. Likewise, as a typical American child acquires a mature version of English, she acquires various I-languages that differ from those of local adults. For a while, her I-language(s) will differ grammatically from "adult English" in ways that may be striking. But soon enough, the child will acquire at least one I-language that corresponds to adult competence, apart from the absence of certain (open-class) lexical items. It is unclear when children acquire semantic vocabulary, in part because one can have lexical items without articulating them. But it is also unclear what counts as semantic vocabulary. Do words like 'correct' count? What about "factive" verbs like 'know'?

[15] Lycan speaks in terms of grammar. But even if 'Linus is not true(-in-λ′)' is ungrammatical in λ′, whose only truth predicate is 'true-in-λ', it does not *follow* that λ′ does not assign the truth condition of (54a) to the sound of 'Linus is not true(-in-λ′)'; cp. 'seems sleeping'.

But suppose we had a characterization of the trouble-making words that must excluded from the lexicon of a Lycan-style "core" language that does not itself generate Liar Sentences.

It still doesn't follow—and it isn't plausible—that speakers who understand (4) and (54),

(4) Kermit is not red.

(54) Linus is not true.

have one I-language that generates only (4) *and another* I-language that generates (4) and a suitably subscripted analog of (54). We can imagine a child who hasn't acquired the word 'true', but who does have an I-language λ that generates (4). Later, the child may have an I-language λ′ that also generates (54). We might describe this change by saying that the child added some mental analog of (86) to an expandable list of semantic axioms.

(86) ∀x[TrueOf('true', x) ≡ True(x)]

Or we might say that the child replaced λ with a extended language λ′. But why think that when the child has λ′, she *also* has λ? Once the child has λ′, it seems that (for her) λ is an arbitrarily defined object. Indeed, λ seems gruesome: λ′ except for lexical items acquired after some time *t*.

Theorists are free to describe a lexicon **L** and a combinatorial system **C** as the result of adding one or more lexical items to an I-language specified in terms of **C** and a slightly smaller lexicon $\mathbf{L}^{-}$. One can then introduce, by stipulation, extensionally distinct notions like true-in-**CL** and true-in-$\mathbf{CL}^{-}$. But whatever semanticists choose to do, and however they choose to talk, the empirical questions remain: is there a Tarski-style theory of truth for the generative procedure **CL**; and if so, is some such theory the core of a correct theory of meaning for **CL**? Certain arguments for a negative answer will not apply to $\mathbf{CL}^{-}$, since the "smaller" language does not generate the apparent counterexamples. But that does not make the restriction principled. (Clauses like 'thinks he snores' present concerns even if some languages have no such clauses.)

Adding 'true' to an I-language seems no different than adding 'snores' or 'Linus'. And if λ′ is a generative procedure, then prima facie, a truth theory for λ′ is also a truth theory for any generative procedure λ that is just like λ′ except for having a slightly smaller lexicon. So to repeat, why think that when our imagined child acquires the word 'true', she both acquires a slightly more productive procedure and retains the slightly less productive procedure? One needs reasons, independent of (D), for thinking that Human I-languages exhibit Lycan's hierarchy.

(D) for each Human Language **H**, some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

Nonetheless, Lycan offers a serious attempt to engage with the difficulty that Liar Sentences present for (D). He also admits that his proposal is, in certain respects, *ad hoc*. Lycan's aim was to offer an alternative Ludwig and Lepore (2005), who offer a different kind of proposal. On their view, a theory whose theorems include (63) or some analog like (63a)

(63) 'Linus is not true.' is true if and only if Linus is not true.

(63a) True('Linus is not true.') ≡ ~True(Linus)

can be a correct theory of meaning for English. The idea is that (63) and the axioms it follows from can be interpretive, even if the axioms—and endlessly many theorems like (63)—are false. In which case, a truth theory Θ doesn't have to be *true* to be a correct theory of meaning for a Human Language **H**. It's good enough if Θ generates an interpretive T-theorem for each sentence of **H**, given a background logic that does not also yield uninterpretive theorems; cp. Eklund (2002). I am suspicious of the idea that false theories of truth can be correct theories of meaning. In this respect, I appreciate the motivations for Lycan's strategy, even if my conclusion is that (D) is the trouble-making thesis that we should reject. But given my objections to Lycan's strategy, let me briefly address Ludwig and Lepore's defense of (D).

While they don't explicitly reject thesis (C),

(C) each Human Language is an I-language

my sense is that Ludwig and Lepore have some other conception of what Human Languages *are*. Or perhaps they hope to remain neutral on this score. But in my view, (C) is more plausible than any extant defense of (D). Advocates of (D) may not want to assume (C). But I see no reason not to assume (C) when evaluating (D). So let's consider thesis the very bold thesis (CDF),

(CDF) each Human Language **H** is an I-language such that some *false* Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

which invites the question of how can a false theory of truth be a correct theory of meaning.

I'll return to a cognitivist answer. But on my reading of Lepore and Ludwig, they are not offering the empirical hypothesis that competent speakers of **H** mentally represent the axioms of a certain truth theory whose axioms are false. According to them, a truth theory can serve as a meaning theory so long as its theorems are interpretive, but theorems don't have to be true to be interpretative. It's enough, Lepore and Ludwig suggest, if the right sides of T-sentences like (53)

(53) True('Kermit is not red.') ≡ ~Red(Kermit)

provide the overall best translations of the object language sentences, given certain restrictions on the metalanguage. One can argue about how to characterize the relevant notion of translation (or "samesaying"). But my concern is more basic. Absent an independently motivated condition on being interpretive, I don't see why a set of *inconsistent* axioms should be regarded as a correct theory of meaning for a Human Language just because the axioms are better than *other* truth-theoretic axioms with regard to yielding theorems that can be used to *translate* sentences that a child could understand *into* sentences of an invented metalanguage.

Given a consistent theory that recursively pairs each English expression E with a plausible S-theorem of the form shown in (87),

(87) $\forall\sigma[\text{Satisfies}(\sigma, \text{E}) \equiv \Phi(\sigma)]$

I can see why one might say that the theory is meaning-specifying. Even if the theorems involve considerable formalism, one can argue that the apparent "material adequacy" of the axioms is best explained by the hypothesis that the axioms do indeed capture the core semantic properties of English expressions—at least in the absence of any alternative explanation for why the theorems seem to be *true*. Correlatively, one might think that if at least one (suitably formulated) theory of truth for English is true, that itself is an argument that the core semantic explananda concern Tarski-style relations that linguistic expressions bear to the environments that speakers share. But one needs independent reasons for thinking that *false* theorems capture explananda. So even if the false theorems (of a false theory) are "interpretive" in some technical sense, why take that to be evidence that the theorems (and the axioms) are meaning-specifying? Perhaps the Lepore and Ludwig defense of (D) is the best one available. But if so, that may tell against (D).

3.3 Last Refuge?

At this point, one might combine (CDF)

(CDF) each Human Language **H** is an I-language such that some *false* Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

with the idea that speakers of **H** represent the axioms of a truth theory for **H**. Perhaps speakers of English tacitly endorse a *false* theory whose theorems include Liar T-sentences like (63a).

(63a) True('Linus is not true.') ≡ ~True(Linus)

On this cognitivist view, an interpretive truth theory is not merely a procedure that pairs object language expressions with good formal translations. The hypothesis is that the relevant metalanguage is a mental language that the (one or more) speakers of **H** use to represent expressions as having certain truth-theoretic properties; cp. Heck (2004, 2007), Eklund (2002).

Let's grant that if speakers use a truth theory Θ to understand sentences like (4) and (54),

(4) Kermit is not red.

(54) Linus is not true.

then Θ is plausibly viewed as a theory of meaning, even if it has some false axioms/theorems. If the best theory of *truth* for **H** differs from the best theory of *understanding* for **H**, then the latter may well be the better candidate theory of meaning for **H**. Indeed, one might think that (D)

(D) for each Human Language **H**, some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

is attractive only in so far as it remains plausible that a truth theory for a Human Language can serve as a theory of how expressions of that language are understood by competent speakers; see Dummett (1976). And given Liar Sentences, perhaps the best way to defend (D) is by defending a cognitivist version of (CDF). But this strategy comes with a justificatory burden. Why retain the idea that correct theories of meaning/understanding for Human Languages will take the form of *truth* theories if the best candidate theories have false axioms/theorems?

One has to ask what explanatory role the appeal to truth—or a mental truth predicate—plays in a theory according to which axioms like (40) and (86)

(40) ∀x[TrueOf('snores', x) ≡ Snores(x)]

(86) ∀x[TrueOf('true', x) ≡ True(x)]

are interpretive because speakers have mental analogs of these axioms and thereby *mis*represent linguistic expressions as having certain truth-theoretic properties. If a truth theory Θ implies that certain expressions of **H** are true of certain things, but (on pain of contradiction) the expressions are not true of those things, why think that speakers of **H** tacitly endorse Θ?

One can posit "error" theories. But note that (D) is usually embedded within a broader view of the subject matter and methodology of semantics: complex expressions of a human language *have* the properties ascribed by a certain truth theory; competent speakers of the language often *recognize* that the expressions have these properties; and clever theorists—who suitably control for various complications when creating settings in which speakers are asked to assent or dissent, given a sentence and a situation—can cite *judgments as evidence* for or against proposed truth theories, because in suitably controlled settings, competent speakers are reliable judges of whether or not the truth conditions *that sentences have* are met in various situations.

Advocates of (D) who reject this broader view, and grant that sentences of a Human Language are *not* conditionally true or false as specified by a correct Tarski-style theory, thereby concede much to critics of (D). So what justifies the residual appeal to *truth* in (CDF)? Perhaps in the end, (CDF) will be part of the best account of how humans understand expressions, and how our judgments can be evidence for or against proposed theories of understanding. But recall that in order to focus on sentences like (54), we set aside *many* other concerns about (D).

(54) Linus is not true.

So if the response to (54) is that false truth theories can still serve as correct meaning theories, one wants to hear more about how proposals regarding specific constructions like (7) and (12-14)

(7) Bert thinks that Hesperus rose at dusk.

(12) The sky is blue.

(13) Snowflakes are white, and mosquitoes carry diseases.

(14) France is hexagonal, and France is a republic.

are to be evaluated. If proposals regarding belief ascription can be cast in terms of (CDF), is that a licence to adopt a theory some of whose theorems are (perhaps demonstrably) false?

If speakers' judgments are symptoms of mental representations, but not evidence that (7) and (12-14) have the properties ascribed by a Tarski-style truth theory, then advocates of (D) need to say how the data that semanticists use is related theories of truth/meaning/understanding. In section two, I noted that advocates of (D) cannot assume that sentences like (59) are true.

(59) 'Kermit is red.' is true if and only if Kermit is red.

If the reply is that (59) isn't true, but that (CDF) is, the next question is obvious: why retain (D)? At this point, what motivates the alleged connection between understanding and truth?

I don't doubt that some theories of meaning are better than others. But the question is how advocates of (CDF) can adjudicate between alternative theories, yet plausibly maintain that the cited evidence is evidence that favors one *truth* theory over another. This is not the place to explore other conceptions of meaning. But axioms like (40)

(40) ∀x[TrueOf('snores', x) ≡ Snores(x)]

might be replaced with overtly mentalistic alternatives according to which the meaning of 'snores' is a mental representation—e.g., the concept SNORES(_)—or perhaps an instruction for how to assemble a mental representation of a certain sort; where executing such an instruction might involve accessing a "root concept" that is stored at a certain "lexical memory address" that associated with the pronunciation of 'snore' (see Pietroski 2010, forthcoming).

Lewis (1972) famously rejected Katz and Fodor's (1963) mentalistic suggestion that the generable Chomsky-style syntactic structures of a Human Language are meaningful by virtue of some computational procedure that relates those structures to independently significant (and often truth-evaluable) expressions of "Markerese." I don't want to defend the Katz-Fodor proposal. But note that advocates of (CDF) cannot offer Lewis' reply to psychological conceptions of meaning. Lewis held that Katz and Fodor did not offer a genuine *semantic* theory, absent a specification of how their posited representations are semantically related to the world that determines the truth or falsity of truth-evaluable thoughts. As he put it, "Semantics with no truth conditions is no semantics." But this slogan is not an argument; see Harman (1974).

One can view Lewis' slogan as a stipulation regarding 'semantics'.[16] But then the question is whether a good theory of *meaning* for a Human Language will be a semantics in this stipulated sense; cp. Chomsky (2000). So let's read Lewis as offering the hypothesis that correct theories of meaning for Human Languages can and should take the form of truth theories. On this view, English bears a certain Tarskian relation to the world; the semanticist's job is to characterize this relation; and correct theories are *true*. Given this unpsychological conception of the project, explicit in Montague (1974), a suitable algorithm that relates syntactic structures to truth-in-a-model conditions (via the lambda-calculus, which lets one represent the space of computable functions) is at least a candidate for being a correct semantics for a language. But it is unclear how the unpsychological project is related to the natural phenomenon of linguistic understanding, apart from the latter providing some data points that may interest some modelers. By contrast, while a Katz-Fodor algorithm is not a candidate semantic theory for Lewis, it is at least a candidate component of linguistic understanding.

Advocates of (CDF) are welcome to join those of us who reject Lewis-style rejections of mentalistic conceptions of meaning. But having conceded so much, advocates of (CDF) need to say why their error theory is better than an overtly mentalistic conception of meaning. It may be that once we adopt a cognitivist conception of meaning theories, we should not insist that the content of the cognition is a *true* theory of truth. But then why think that the content of the

---

[16] Lewis followed Tarski in this respect; see Burgess (200x), who notes that Tarski's use of 'semantics' was also stipulative and nonstandard.

cognition is *any* theory of *truth*? Liar Sentences are not direct counterexamples to (CDF). But if the question is whether (D) is plausible, and Liar Sentences drive advocates of (D) to (CDF), then when comparing (CDF) with alternatives—including accounts of meaning that reject (D)—we need to set aside any alleged virtues of (D) that are not preserved by (CDF). If only for this reason, I think that Liar Sentences bear importantly on the study of human linguistic meaning.

4. Final Thoughts

When theorizing about a cognitive competence whose application to cases is subject to logically contingent constraints, we need to distinguish two notions of implication. Given any theory, we want to know what it implies given the rest of science, logic and mathematics included. We don't want theories whose implications are false, much less contradictory. But in giving a theory of meaning for a Human Language, we also need to posit a more parochial notion of implication, corresponding to the *constrained* capacities of speakers to deploy semantic competence in ways that support the comprehension of boundlessly many expressions. Children somehow acquire knowledge of word meanings, and of how these meanings can be systematically combined, in a way that lets children generate and comprehensible novel expressions. Linguistic competence evidently involves a capacity to extract, on demand, various consequences of lexical and compositional "axioms." But as with other natural competences, deployment is constrained by the relevant representational forms, which are biologically implemented. So the knowledge in question is not closed under the kind of deduction that we use to evaluate theories.

In doing science—and more generally, when trying to find out whether a particular claim is true—one wants a background logic that is far more powerful than the propositional calculus. Ideally, one wants to make *all* the implications of a claim manifest, whatever implications turn out to be; see, e.g., Frege (1879, 1884, 1893). This requires an interesting logic, and presumably one that licenses many inferences not licensed by the first-order predicate calculus. And for purposes of figuring out what follows from a theory, mathematical implications usually count. But not even arithmetic is reducible to "pure logic;" analysis, geometry, and topology seem to go well beyond logic. Moreover, we often extend the notion of implication to include inferences that are not justifiable a priori, so long as the tacit premises are beyond reasonable doubt for the purposes at hand. In any case, as theorists, we can infer from (5T) to (34).

(5T) True('Ernie snores.') ≡ Snores(Ernie)

(34) True('Ernie snores.') ≡ Snores(Ernie) & Precedes(Two, Five)

By contrast, if ordinary speakers of English can extract an analog of (5T) from their linguistic knowledge, and this capacity is the source of why (5) means what it does,

(5) Ernie snores.

then this capacity cannot support extraction of (34) or other instances of (35);

(35) True('Ernie snores.') ≡ Snores(Ernie) & **Γ**

where **Γ** is itself an extractable theorem. But as we've seen, this requires a simple (and naïve) notion of extraction that seems to be odds with the sophistication required to keep semantic paradoxes at bay. In short, there is a tension between (i) characterizing knowledge of meaning in terms of truth and (ii) supposing that this knowledge is, unlike truth, closed only under very weak deductive principles.

In retrospect, it seems that Davidson (1967a) was remarkably brief about paradoxes.

> The semantic paradoxes arise when the range of the quantifiers in the object language is too generous in certain ways. But it is not really clear how unfair to Urdu or to Wendish it would be to view the range of their quantifiers as insufficient to yield an explicit definition of 'true-in-Urdo' or 'true-in-Wendish'....In any case, most of the problems of general philosophical interest arise within a fragment of the natural language that may be conceived as containing very little set theory (pp. 28-29).

A correspondingly brief reply is that deeper issues, concerning linguistic *meaning*, arise in many ways that Davidson did not address. For example, if Linus is sentence (2),

(2) The second numbered sentence in "I-Languages and T-sentences" is not true.

then (49) is false. So prima facie, a theory of meaning for English should not imply (49).

(49) True('Linus is not true.') ≡ ~True(Linus)

This leads to problems for thesis (D) that are not addressed by restricting the range of quantifiers.

(D) for each Human Language **H**, some Tarski-style theory of truth for **H** is the core of a correct theory of meaning for **H**

I'm not sure how Davidson was counting "problems of general philosophical interest." But one question of interest to many philosophers, among others, is whether (D) is true. So we should ask if (D) is plausible in light of sentences like (54).

(54) Linus is not true.

If not, then (D) may not be plausible for any Human Language that generates (4) or (5).

(4) Kermit is not red.

(5) Ernie snores.

According to (D), Human Language expressions are meaningful by virtue of being related to the world in the way a Tarskian language is related to its domain, but talking about this relation somehow leads to paradox. In my view, expressions have meanings that are largely independent of how they are related to language-independent things. Humans can use meaningful expressions to form and express boundlessly many judgments that are systematically related and truth-evaluable. And here lies the hard work: describing, in a consistent way, how meaning is related to judgment; where judgments are often—and perhaps always aim to be—true or false, like the sentences of an invented *Begriffsschrift*; and meaningful Human Language expressions are, without being true of things, essential to the human capacity for judgment.

Providing such description requires a coherent conception of truth. In this paper, I have said very little about which things *are* truth-evaluable, and nothing about how to avoid the problems that arise in thinking about such things. I have merely argued that sentences like (54) illustrate a deep difficulty for truth-theoretic conceptions of meaning for Human (I-)Languages, and that we should look for a different conception according to which expressions of these languages are not among the truth-evaluable things. Providing theories of meaning/understanding is hard enough without requiring them to also serve as theories of truth.

By tying meaning tightly to truth, (D) offers a quasi-reductive conception of meaning that can initially seem attractive. But one needs to factor in the costs of eschewing meanings as "middle men" that mediate the complex relations that pronounceable syntactic structures bear to the language-independent things towards which truth-evaluable judgments are directed. Liar Sentences remind us that these relations are intricate enough to permit *proofs* that some conceptions of these relations are too simplistic. And since this wouldn't be the first time that a thesis with initial attractions turned out to be wrong, for reasons illustrated with a *reductio*, we shouldn't be too surprised if (D) suffers this fate.

References

Austin, J. (1961). *Philosophical Papers*. Oxford: Oxford University Press.

Austin, J. (1962). *How to Do Things with Words*. Oxford: Oxford University Press.

Berwick, R. et.al. (2011). Poverty of the Stimulus Revisited. *Cognitive Science* 35: 1207-42.

Boolos, G. (1998). *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press.

Burge, T. (1973): Reference and Proper Names. *Journal of Philosophy* 70:425-39.

Burgess, J. (2008). Tarski's Tort. In *Mathematics, Models, and Modality*. Cambridge: Cambridge University Press.

Cartwright, R. (1962). Propositions. In R. Butler (ed.), *Analytical Philosophy*, 1st series, Oxford: Basil Blackwell, 81–103.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1977). *Essays on Form and Interpretation*. New York: North Holland.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Church, A. (1941). *The Calculi of Lambda Conversion*. Princeton: Princeton University Press.

Crain, S. and Pietroski, P. (2001). Nature, Nurture, and Universal Grammar. *Linguistics and Philosophy* 24:139-186.

Davidson, D. (1967a). The Logical Form of Action Sentences. Reprinted in Davidson 1980.

Davidson, D. (1967b). Truth and Meaning. Reprinted in Davidson 1984.

Davidson, D. (1968). On Saying That. Reprinted in Davidson 1984.

Davidson, D. (1969). The Individuation of Events. Reprinted in Davidson 1980.

Davidson, D. (1976). Reply to Foster. Reprinted in Davidson 1980.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.

Davidson, D. (1984). *Essays on Truth and Interpretation*. Oxford: Oxford University Press.

Davidson, D. (1986). A Nice Derangement of Epitaphs.

Davies, M. (1987). Tacit knowledge and semantic theory: Can a five per cent difference matter? *Mind*, *96*, 441–462.

Donnellan, K. (1966). Reference and Definite Descriptions. *Philosophical Review* 75: 281-304.

Dummett, M. (1976). What Is a Theory of Meaning? In Evans and McDowell (1976), 67–137.

Eklund, M. (2002). Inconsistent Languages. *Philosophy and Phenomenological Research* 64:251-275

Etchemendy, J. (1990). The Concept of Logical Consequence. Cambridge: Harvard University Press.

Evans, G. (1982). Varieties of Reference (Oxford: Oxford University Press).

Evans, G. and McDowell, J., eds. (1976). *Truth and Meaning*. Oxford: Oxford University Press.

Fodor, J. and Lepore, E. (1992). *Holism: A Shopper's Guide*. Oxford: Blackwell.

Foster, J. (1976). In Evans and McDowell (1976).

Frege, G. (1879). *Begriffsschrift*. Halle: Louis Nebert. English translation in *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931,* J. vanHeijenoort (ed). Cambridge, MA: Harvard University Press, 1967.

Frege, G. (1884). *Die Grundlagen der Arithmetik*. Breslau: Wilhelm Koebner. English translation in *The Foundations of Arithmetic*, J. L. Austin (trans). Oxford: Basil Blackwell, 1974.

Frege, G. (1892a). Function and Concept. In Geach & M. Black (trans). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell, 1980.

Frege, G. (1892b). On Sense and Reference.In Geach & M. Black (trans). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell, 1980.
Frege, G. (1893/1903). *Grundgesetze der Arithmetik*, Jena: Verlag Hermann Pohle, Band I/II. Partial translation of Band I, *The Basic Laws of Arithmetic*, by M. Furth, Berkeley: U. of California Press, 1964.
Glanzberg, M. (2004). A Contextual-Hierarchical Approach to Truth and the Liar Paradox, *Journal of Philosophical Logic* 33: 27-88.
Glanzberg, M. (forthcoming). Logical Consequence and Natural Language. In C. Caret and O. Hjortland, eds., *Foundations of Logical Consequence* (Oxford: Oxford University Press).
Gupta, A., and Belnap, N. (1993). *The Revision Theory of Truth*, Cambridge, MA: MIT Press.
Hamblin, C. (1958). Questions. *The Australian Journal of Philosophy* 159–168.
Hamblin, C. (1973). Questions in Montague English. *Foundations of Language* 10: 41–53.
Harman, G. (1974). Meaning and Semantics. In K. Munitz & P. Unger (eds.), *Semantics and Philosophy*. New York: New York University Press (1974)
Heck, R. (2004). Truth and Disquotation. *Synthese* 142: 317–352.
Heck, R. (2007). Meaning and Truth-conditions. In D. Greimann and G. Siegwart, *eds*., *Truth and Speech Acts: Studies in the Philosophy of Language* (New York: Routledge).
Heck, R. (2011). *Frege's Theorem*. Oxford: Oxford University Press.
Higginbotham, J. (1985). On Semantics. *Linguistic Inquiry* 16: 547–93.
Higginbotham, J. (1991). Truth and understanding. *Iyyun* 40:271-288.
Higginbotham, J. and May, R. (1981). Questions, Quantifiers, and Crossing. *The Linguistic Review* 1: 47-80.
Kaplan, D. (1978a). Dthat. In P. Cole, ed., *Syntax and Semantics* (New York: Academic Press).
Kaplan, D. (1978b). On the Logic of Demonstratives. *Journal of Philosophical Logic* 8: 81–98.
Kaplan, D. (1989). 'Demonstratives', in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan* (New York: Oxford University Press).
Karttunen, L. (1977). Syntax and Semantics of Questions. *Linguistics and Philosophy* 1: 3–44.
Katz, Jerrold (1981). *Language and Other Abstract Objects*. Rowman and Littlefield: Totowa, NJ.
Katz and Fodor (1963). The Structure of a Semantic Theory. *Language* 39: 170–210.
Kleene, S. (1950). *Introduction to Metamathematics*. D. Van Nostrand, Princeton, NJ.
Kripke, S. (1975). Outline of a theory of truth, *Journal of Philosophy*, 72: 690–716.
Kripke, S. (1979). A Puzzle about Belief. In A. Margalit (ed.) *Meaning and Use* (Dordrecht: Reidel).
Kripke, S. (1980). *Naming and Necessity.* Cambridge, MA: Harvard University Press.
Larson, R. and Segal, G. (1995). *Knowledge of Meaning*. Cambridge, MA: MIT Press.
Larson, R. and Ludlow, P. (199x). Interpreted Logical Forms. *Synthese* 95:305-55.
Lepore, E., ed. (1986). *Truth and Interpretation*. Oxford: Blackwell.
Lepore, E., and K. Ludwig (2005). *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford: Clarendon Press.
Lepore, E. and Ludwig, K. (2007). *Donald Davidson's Truth-Theoretic Semantics*. Oxford: Oxford University Press.
Leslie, S. (2007). Generics and the Structure of the Mind. *Philosophical Perspectives* 21: 375-403.
Lewis, D. (1972). General Semantics. In D. Davidson and G. Harman (eds.), *Semantics of Natural Language* (Dordrecht: Reidel).

Lewis, D. (1975). Languages and Language. In K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, Volume 7, Minneapolis: University of Minnesota Press, pp. 3–35.
Lewis, (1986). On the Plurality of Worlds (Oxford: Blackwell).
Lidz, J. et.al. (2011). Interface Transparency and the Psychosemantics of 'Most'. *Natural Language Semantics* 19: 227-56.
Lohndal, T. and Pietroski, P. (2011). Interrogatives, Instructions, and I-languages: An I-Semantics for Questions. *Linguistic Analysis* 37:459-510.
Ludlow, P. (2011). *The Philosophy of Generative Linguistics*. Oxford: OUP.
Lycan, W. (2012). A Truth Predicate in the Object Language. In G. Preyer (ed.), *Donald Davidson on Truth, Meaning and the Mental* (Oxford: Oxford University Press).
Marr, D. (1982). *Vision.* San Francisco: Freeman and Company.
Montague, R. (1974). *Formal Philosophy*. New Haven: Yale University Press.
Parsons, C. (1974). The Liar paradox. *Journal of Philosophical Logic* 3: 381–412.
Peacocke, C. (1986). Explanation in computational psychology: Language, perception and level 1.5. *Mind and Language* 1:101–123.
Pietroski, P. (199x). Fregean Innocence.
Pietroski, P. (2005a). *Events and Semantic Architecture*. Oxford: Oxford University Press.
Pietroski, P. (2005b). Meaning Before Truth. In *Contextualism in Philosophy* (G. Preyer and G. Peters, eds.). Oxford: Oxford University Press.
Pietroski, P. (2006). Character Before Content. In *Content and Modality: Themes from the Philosophy of Robert Stalnaker* (J. Thomson and A. Byrne, eds.), Oxford University Press.
Pietroski, P. (2008). Minimalist Meaning, Internalist Interpretation. *Biolinguistics* 4:317-41.
Pietroski, P. (2010). Concepts, Meanings, and Truth: First Nature, Second Nature and Hard Work. *Mind and Language* 25: 247-78.
Pietroski, P. (2011). Minimal Semantic Instructions. In the *Oxford Handbook on Linguistic Minimalism* (C. Boeckx, ed.), Oxford University Press.
Pietroski, P. (2015). Framing Event Variables. *Erkenntnis* 80:31-60.
Pietroski, P. (forthcoming). *Conjoining Meanings: Semantics Without Truth Values*.
Priest, G. (1979). Logic of Paradox. *Journal of Philosophical Logic* 8: 219–241.
Priest, G. (2006). In Contradiction: A Study of the Transconsistent, second edition. Oxford: Oxford University Press.
Quine, W. (195x)
Rescher, N. (1962). Plurality Quantification. *Journal of Symbolic Logic* 27:373-4.
Schlenker, P. (2009). Local contexts. *Semantics and Pragmatics*, 2: 1–78.
Segal, G. (1989). A Preference for Sense and Reference. *Journal of Philosophy* 86:73-89.
Segal, G. (1991). In the mood for a semantic theory. *Proceedings of the Aristotelian Society* 91: 103–118.
Soames, S. (1984). Linguistics and Psychology. *Linguistics and Philosophy* 7: 155-179.
Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.
Strawson, P. (1950). On referring. *Mind*, 59(235):320–344.
Tarski, A. (1933). 'The Concept of Truth in Formalized Languages', reprinted in *Logic, Semantics,* Metamathematics, tr. J. H. Woodger (Indianapolis: Hackett).
Wiggins, D. (1980). Most' and 'all': some comments on a familiar programme, and on the logical form of quantified sentences. In M. Platts (ed.) *Reference, truth and reality, Essays on the philosophy of language* (London: Routledge & Kegan Paul) 318–346.