

## Internalist Semantics: Comments on Paul Pietroski, *Conjoining Meanings*

Zoltán Gendler Szabó  
Yale University

### 1. How to be a meaning internalist?

Suppose you say to yourself “If I cut the pie the way it should be then meanings must be in the head!”<sup>1</sup> In other, less catchy words: expressions of natural languages understood by intrinsically identical people must have the same meaning. What could meanings be if this is true? The standard answer is that they could be *concepts* – composable mental symbols – assembled in the minds of those who interpret linguistic expressions. For example, the meaning of ‘mothers’ could be a complex concept built from the meanings of its constituents (presumably, the noun ‘mother’ and the plural morpheme ‘-s’) in a way corresponding to its morpho-syntactic structure.

There are at least three objections to this type of meaning internalism. The first is that it is vacuous. We are told that for one thing (e.g. the word ‘mother’) to have a certain meaning is for it to be associated with another (e.g. the concept MOTHER) that has the same meaning. This seems about as revealing as saying that for a dish to be spicy is for it to be associated with a spicy taste. Interpreting a word may involve translating it to *Mentalese* but then we still have to say what the translation means.<sup>2</sup>

The second objection is that the proposal overcounts meanings. ‘Mother’ is associated with multiple concepts: one that applies to someone whose egg unites with a sperm producing a child, another that applies to someone who gives birth to a child, a third that applies to someone who raises a child. And yet, ‘mother’ does not seem to have all these different but closely related meanings. If you introduce me to someone saying ‘I’d like you to meet my mother’ and I mistakenly conclude that she is the person who gave birth to you then I misunderstood *you*, but I did not misinterpret *your words*.<sup>3</sup>

The third objection is that the proposal isn’t general enough. Even if intrinsically identical people must associate the same concepts with ‘mother,’ they could associate different ones with ‘water’ or with ‘arthritis.’ No concept associated with ‘water’ is employed in thinking about XYZ, and yet, someone intrinsically identical to me living on a planet where river and lakes are filled with XYZ could be thinking of XYZ using such a concept. Similarly, no concept associated with ‘arthritis’ is employed in thinking about inflammations in thighs, and yet, someone intrinsically identical to me living in a linguistic community whose members used ‘arthritis’ to talk about inflammations in thighs could be thinking about that using such a concept.<sup>4</sup>

---

<sup>1</sup> “Cut the pie any way you like, ‘meanings’ just ain’t in the head!”, Putnam wrote. Cf. Putnam, Hilary (1973) ‘Meaning and Reference.’ *Journal of Philosophy* **70**: 704.

<sup>2</sup> This is how I read the paragraph that contains the famous quip “Semantics with no treatment of truth-conditions is no semantics” in Lewis, David (1970) ‘General Semantics.’ *Synthese*, **22**: 18.

<sup>3</sup> This is a central theme in Chomsky, Noam (2000) *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

<sup>4</sup> See Putnam *op. cit.* and Burge, Tyler (1979) ‘Individualism and the Mental.’ *Midwest Studies in Philosophy*, **4**: 73 – 121.

There are answers to these objections. One can argue that the view is informative because it reduces linguistic meaning to mental content, that polysemous words are just special homonyms, and that concepts have narrow contents. Alternatively, one can articulate a different version of meaning internalism that bypasses these concerns. This is what Paul Pietroski has done in *Conjoining Meanings*.

The meanings of natural language expressions, we are told, are not the concepts assembled in interpreting them but *instructions* for the assemblage (19). The meaning of ‘mothers’, for example, is the instruction to fetch a monadic concept associated with the noun ‘mother’, fetch another associated with the plural morpheme ‘-s’, and then conjoin the two. Instructions get more intricate as we consider more complicated expressions, but all linguistic interpretation is a mechanical process of fetch, join, and repeat (106) – building complex concepts from simple ones. What, if anything, these concepts apply to in the external world matters to cognition but makes no difference to semantics.

A meaning, on this view, is not a representation. It does not *have* meaning or content, and hence, is not subject to interpretation. This immediately disarms the first objection. In telling what sort of instruction a given meaning is, we are saying everything that needs to be said about what it is – no passing of the buck to a future theory of intentionality. The second objection is also moot. Lexical meanings are instructions to fetch a concept meeting some constraint from a mental address. Nothing rules out that there are multiple concepts meeting the constraint there. If so, linguistic interpretation will not have a determinate outcome and we must rely on pragmatic mechanisms to figure out what was intended. Finally, there is nothing in this view of linguistic meaning that is incompatible with externalism about mental content. The meaning of ‘water’ (or ‘arthritis’) is an instruction to fetch an appropriate concept from a mental address – the same instruction in the minds of all who understand it. What concepts are stored at the address may depend on facts about the physical and social environment.

Externalists think being meaningful is being related in a certain way to something external. Internalists who identify meanings with concepts hold on to this relational view – they simply replace the external thing with an internal representation. On Pietroski’s view, being meaningful is an *intrinsic* feature. But the view goes even further:

If meanings are concept assembly instructions, Slang expressions [Pietroski’s term for expressions of natural languages] can *be* these meanings/instructions. [...] This does not imply that every feature of a Slang expression is relevant to an encoded meaning. On the contrary, a generable expression can also be a phonological instruction [...] generable expressions can be partially described in various ways: as pronounceable meanings; as meaningful pronunciations; or as syntactic structures that are meaningful and pronounceable.” (292-3)

I think this idea is a problematic, even by Pietroski’s own light. Consider one of the initial examples he uses to illustrate the kind of data semantic theories should explain (41). (1) is paraphraseable as (1a), but not (1b) or (1c) – why is this so?

- (1) This is a cow that we saw yesterday.
- (1a) This is both a cow and we saw it yesterday.
- (1b) This is either a cow or we saw it yesterday.
- (1c) This is a cow if we saw it yesterday.

We might conjecture that the significance of combining a word with a relative clause is fixed by the innate endowment children use to acquire language. Then it is a law of natural language that given what ‘that we saw yesterday’ and ‘a cow’ mean ‘a cow that we saw yesterday’ must mean what it does. So far so good. But the modality here is restricted – there is no reason to deny that we could *invent* a language where ‘that we saw yesterday’ and ‘a cow’ mean just what they do in English, and yet, ‘a cow that we saw yesterday’ means something entirely different.<sup>5</sup> If sentences *are* their meanings then they could not mean anything other than what they actually mean, and hence, the perfectly sensible question “Why does (1) have the meaning of (1a), rather than that of (1b) or (1c)?” would be much like the absurd question “Why is water H<sub>2</sub>O, rather than H<sub>3</sub>O or HO<sub>2</sub>?”

Pietroski can hold on to all that is attractive in his meaning internalism without opting for the Chomsky-inspired view that natural languages generate pronounceable meanings. They generate words, phrases, and clauses – things that *have* meanings and pronunciations intrinsically. But they do so contingently.

## 2. Should we be meaning internalists?

How can we go about deciding whether meanings are mental instructions or what is represented by the outputs of such instructions? The colloquial meaning of ‘meaning’ would be a poor guide. The claim that meanings are instructions in the mind sounds like a category mistake, but so does the claim that they are things in the external world. To paraphrase Bishop Butler, meaning is what it is, not another thing. Or, anyway, so it seems when we rely too heavily on what sounds felicitous in everyday conversation. The question is what we *should* call meaning when we try to develop a *theory* of meaning. This depends, in part, on where we think the most fruitful questions in this area lie.

Those who think we are very much in the dark when it comes to investigating the psychological mechanism of linguistic interpretation will be drawn to externalist, referential semantics. Such theories explain how we could, in principle, understand natural languages without making a commitment that the compositional rules they identify are the ones we actually follow. Those who decry making ontological claims based on linguistic evidence will find internalist, instruction-based semantics appealing. These theories purport to give the exact mechanism at work in the interpretation of natural language but set aside the question how to understand people who use language to say various things that can be evaluated as true or false.

Partisans in this debate often justify their choices in stark terms. David Lewis, for example, expresses doubts that we can make objective sense of the claim that a given language has a particular grammar. He argues that people speak a language *L* in virtue of belonging to a community where a convention of truthfulness and trust in *L* holds, but such a convention cannot pin down how the sentences of *L* are generated from their parts. He does not discard the notion of meaning for subsentential expressions of *L* (“that would be absurd” he says), but he thinks this notion is “no clearer and no more objective than our

---

<sup>5</sup> Cf. Szabó, Zoltán Gendler (2000) ‘Compositionality as Supervenience.’ *Linguistics and Philosophy* 23: 475 – 505. There I argue that the principle of compositionality should be construed in a way that rules out the existence of a *natural* language like this.

notion of a *best* grammar” for *L*.<sup>6</sup> Lewis’s reasons for skepticism about grammar are highly speculative. Like Quine and Davidson, he thinks all evidence in favor of interpreting a language must be manifestable in behavior, and that the only relevant form of behavior is acceptance of a sentence as true. Neither of these assumptions is forced upon us.

Noam Chomsky, on the other side of the debate, is even more forceful. He thinks the technical terms ‘reference’, ‘extension’, ‘denotation’, etc. have been introduced to provide a semblance of respectability for our inherently messy ideas about intentionality. He thinks “a word, even of the simplest kind, does not pick out an entity of the world” but also that this fact should not be seen as a denial that “there are banks, or that we are talking about something (or even some thing) if we discuss the fate of the Earth (or the Earth’s fate) and conclude that *it* is grim.”<sup>7</sup> Chomsky’s reasons for skepticism about reference are also highly speculative. Like Putnam and Burge, he thinks what we use our words to talk about depends on norms – in particular, norms of who we should defer to – and that there is no credible way to fit the study of these norms into a naturalistic inquiry. Depending on how one understands what counts as naturalistic inquiry, one can either insist that theories of reference can be naturalistic, or deny that their non-naturalism is a blemish.

Pietroski is clearly sympathetic to the Chomskian hard line: he suspects typical concepts (and *a fortiori*, typical expressions of natural languages) don’t have extensions (9). But what he tries to show in *Conjoining Meaning* is something different: that even if expressions of natural languages have extensions, those extensions play no significant role in a theory of meaning. He has two main arguments for this, one based on the semantic paradoxes developed in Chapter 4, and another based on event-semantics presented in Chapter 5. Here I’ll focus on the second.

Davidson has argued that action sentences existentially quantify over events. For example, ‘Jones buttered the toast in the bathroom at midnight’ says that for some event *e*, *e* is a buttering of the toast by Jones, and *e* occurred in the bathroom, and *e* occurred at midnight, which is why the sentence entails both ‘Jones buttered the toast in the bathroom’ and ‘Jones buttered the toast at midnight’ and fails to be entailed by their conjunction.<sup>8</sup> Subsequent work has deepened the empirical case for event semantics.<sup>9</sup> The first premise of Pietroski’s argument is that Davidsonian semantics is on the right track.

It is well-known that the entities event semantics posits are very fine-grained – there are a lot more of them than of the entities we would ordinarily call ‘events’. Not only do we have to distinguish Jack’s selling of a book to Jill from Jill’s buying of a book from Jack (say, because the former could be illegal even if the latter is not) we may even be forced to split up Simon’s playing the song on his tuba into his

---

<sup>6</sup> Lewis, David (1975) ‘Languages and Language.’ In: Gunderson, Keith (ed.), *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science*, Vol. 7. Minneapolis, MN: University of Minnesota Press, p. 177. Note the indefinite article – Lewis thinks for different purposes we may rank candidate grammars differently.

<sup>7</sup> Cf. Chomsky *op.cit.* p.181.

<sup>8</sup> Cf. Davidson, Donald (1967) ‘The Logical Form of Action Sentences.’ In N. Rescher (ed.), *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press, pp. 81–95

<sup>9</sup> Cf. Higginbotham, James (1983) ‘The Logical Form of Perceptual Reports.’ *Journal of Philosophy*, **80**: 100–127, Parsons, Terence (1990) *Events in the Semantics of English*. Cambridge, MA: MIT Press, and Schein Barry (1993) *Events and Plurals*. Cambridge, MA: MIT Press, among others.

playing the song and his playing his tuba. After all, given eventish interpretations, (2) and (3) together with the assumption that they are made true by the same thing entails the ungrammatical (4):

- (2) Simon played the song in three minutes.  
(2')  $\exists x(\text{Played}(x, \text{Simon}, \text{TheSong}) \wedge \text{InThreeMinutes}(x))$
- (3) Simon played his tuba.  
(3')  $\exists x(\text{Played}(x, \text{Simon}, \text{HisTuba}))$
- (4) # Simon played his tuba in three minutes.  
(4')  $\exists x(\text{Played}(x, \text{Simon}, \text{HisTuba}) \wedge \text{InThreeMinutes}(x))$

And yet, it would be ludicrous to claim that when Simon is playing the song on his tuba there he is doing two different things. This is the second premise.

Pietroski thinks the best way to reconcile the two premises is to give up the assumption that (2), (3), or (4) have truth-conditions. They can, of course, be used to make true claims, but that is not because any of thoughts one can build following the instructions that constitute their meanings are true. Those thoughts are all false – the things they existentially quantify over are simply not there. Talk of musical performances is “like talk of lovely sunrises, blue skies, and other ‘things’ we talk about without using words that are true of them.” (199)

Well, that’s one way to go. For my part, I would be more inclined to reject the second premise than to accept this conclusion. On reflection, we can find a clear modal difference between Simon’s playing the song and his playing his tuba. Imagine a counterfactual possibility where the song has never been written. Simon is playing his tuba in exactly the same way but he is not following a score – he is just improvising. (2) would then be false but (3) would still be true.

This probably won’t move Pietroski. He agrees that there are apparent differences between Simon’s playing the song and his playing the tube: we might want to say, for example, that the former reaches its culmination after three minutes, while the latter simply ends then. But this sort of difference, he insists, is not in the events themselves but in the way we *represent* them (219). My question is: so what? What is wrong with saying that Simon’s playing the song is not a bare event but an event represented in a certain way – *as a playing of the song by Simon*?

This is the sort of view Barry Schein holds.<sup>10</sup> Pietroski briefly discusses the proposal but rejects it because he thinks it gives up the rationale for referential semantics: we are no longer in the business of relating linguistic expressions to items in the objective, mind-independent world (207). But this job-description is tendentious: referential semantics seeks to relate expressions to things they are about (in a rather broad *de re* sense of aboutness, modeled with the technical notion of reference) whether or not these things are representation-dependent. Natural language sentences often quantify over representation-dependent things: when Lucy imagines a unicorn she imagines *something*, as does Franz when he fears

---

<sup>10</sup> Cf. Schein, Barry (2017), *‘And’: Conjunction Reduction Redux*. Cambridge, MA: MIT Press, especially Chapter 9. He calls the events represented in a particular way resolutions.

*everything* Lucy imagines. The novelty in Schein's proposal is the idea that we might be doing this all the time, not just when we employ overtly intentional vocabulary.

All of us (except for the idealists) agree that there are objective, real word events, independent of our thinking about them that can be described using natural language. How does this happen? One hypothesis is that sentences of natural languages encode instructions to build thoughts which describe nothing in the real world, but there are other thoughts in the neighborhood which do describe objective events. We pick out one of these thoughts relying on context and background information. Another hypothesis is that sentences of natural languages encode instructions we use to build thoughts which describe events as being represented in a certain way. There are also other thoughts in their neighborhood which describe the same events more objectively, in ways less tied to the character of natural languages. We design artificial languages whose sentences can do this job. Pietroski opts for the first one – I think the second is at least as plausible.

### 3. Eliminating semantic types

I think the most original part of *Conjoining Meanings* is the semantics sketched in Chapters 6 and 7. It is an internalist theory, but its main ideas could be adopted by externalists as well.

Many semanticists nowadays are working within framework whose basic tenets go back to Frege. They include (i) that every expression belongs to a semantic type which determines what kind of meaning it has, (ii) that semantic types are generated recursively from basic types  $e$  (for expressions denoting entities) and  $t$  (for expressions that have truth-values), and (iii) that the interpretation of syntactic composition is functional application. Each of these initial assumptions can be relaxed – semantic theories often allow type shifting as well as additional basic types or composition rules beside functional application – but departures from the bare-bones account must always be justified and kept at a minimum.

This sort of theory raises some uncomfortable questions. Why do expressions of natural languages belong to only a handful of semantic types? It is glaring that even some of the simplest types – like  $\langle e, e \rangle$  or  $\langle t, e \rangle$  – are apparently uninhabited. Why do semantic types look so gerrymandered from the point of view of syntax? Despite their vast differences, most nouns, verbs, and adjectives are standardly ascribed the same semantic type –  $\langle e, t \rangle$ ; despite their massive similarities, proper and common nouns are standardly ascribed different semantic types –  $e$  and  $\langle e, t \rangle$ , respectively. Why are pro-forms not available more freely across semantic types? They tend to be of type  $e$  ('she', 'this'), or type  $\langle e, t \rangle$  ('so', 'how'), but never in otherwise well-inhabited types like  $\langle e, \langle e, t \rangle \rangle$  or  $\langle \langle e, t \rangle, t \rangle$ .

Pietroski suggests that we should radically flatten the semantic type hierarchy. In his view, all natural language expressions are predicates. All complex expressions as well as most of the simple ones are one-place. The exceptions (which are all functional items) are two-place. He argues that proper nouns are predicates, that sentences are predicates applying to everything or nothing, that transitive verbs are two-verb constructions (and so verbs can all be one-place), and that nouns without their count/mass, singular/plural features are number neutral (i.e. can apply to can apply to one, many, or much). The

outline covers roughly the same empirical ground as standard semantics textbooks do, showing the viability of the proposal.

I am very much on board with this project and I find a lot of the details congenial. But I'd like to mention one area where I prefer to go in a different direction. On Pietroski's account, quantificational determiners fetch concepts that apply (non-distributively) to ordered pairs – e.g. ALL/SOME/MOST applies to some ordered pairs iff all/some/most of their second components are one of their first components (330). The interpretation works but it feels farfetched and parochial. Farfetched, because it seems unlikely that everyone who understands 'all', 'some' or 'most' must have the capacity to think about ordered pairs. Parochial, because it is hard to see how it can be generalized to cover the interpretation of the intuitively quite similar (but focus-sensitive) 'always', 'sometimes', and 'often.'

Other concepts Pietroski had to posit to get his semantics rolling can all be plausibly attributed to competent speakers. If you understand 'horse,' you presumably have a concept that applies to horses; if you understand 'see,' you presumably have a concept of that applies to episodes of seeing; if you understand 'see a horse,' you presumably have a concept that applies to those who see a horse; and if you understand 'most horses,' you presumably have a concept that applies to most of the horses. But what sort of concept do you have to have if you understand 'most'? There is nothing natural to say, so Pietroski ends up saying something artificial.

I think it's better to reject the question. 'Most' is a functional item and functional items are not associated with concepts, at least not if concepts are supposed to be mental symbols used to "think about things in certain ways." (78) They are not associated with representations, but rather with ways we can build complex representations from simpler ones. Within Pietroski's framework the natural thing to say would be that their meanings are not instructions to fetch a concept, but instructions to join concepts.

There are already syncategorematic items in Pietroski's semantics – the polarity markers (317) – so treating functional items syncategorematically would not change the basic architecture. It would be a loss to have to accept a large (but finite) number of operations instead of just the handful the current version posits. But eliminating appeal to artificial concepts seems worth the price.