Language and scientific explanation: Where does semantics fit in?

Eran Asoulin

Conceptual Foundations of Language Science 99

language science press

Series information: cfls-info.tex not found!

ISSN: 2363-877X

Language and scientific explanation: Where does semantics fit in?

Eran Asoulin

language science press

Eran Asoulin. 2020. *Language and scientific explanation: Where does semantics fit in?* (Conceptual Foundations of Language Science 99). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog

ISBN: no digital ISBN
no hardcover ISBN
no softcover ISBN
ISSN: 2363-877X
no DOI

Cover and concept of design: Ulrike Harbort
Fonts: Linux Libertine, Arimo, DejaVu Sans Mono
Typesetting software: XeLaTeX

Language Science Press
Unter den Linden 6
10099 Berlin, Germany
langsci-press.org

Storage and cataloguing done by FU Berlin

no logo

Contents

Acknowledgments

This book is the culmination of ten years of thinking and writing about internalist semantics and the nature of meaning. During those years I have benefitted from discussion with many colleagues and friends in various universities, conferences, and personal correspondences. I would like to especially thank Peter Slezak, Nick Riemer, Mengistu Amberber, Debra Aarons, Michael Levot, Justin Colley, and all the regular participants at the cognitive science discussion group at the University of New South Wales who read drafts of some of the chapters in this book.

I am also grateful to Language Science Press and to the Conceptual Foundations of Language Science series editors and reviewers.

Preface

In their 1963 paper "The structure of a semantic theory", Jerrold Katz and Jerry Fodor argued that a characterisation of the abstract form of a semantic theory is given by a meta-theory that answers questions such as *What is the domain of a semantic theory? What are the descriptive and explanatory goals of a semantic theory? What mechanisms are employed in pursuit of these goals? What are the empirical and methodological constraints upon a semantic theory?* Even though the Katz and Fodor paper was an early attempt to develop a semantic theory that would be compatible with a Chomskyan syntax, their introductory comments are applicable to semantic theories in general. That is, in order to be taken seriously, any semantic theory must be able to answer such meta-theoretical questions. Moreover, the extent to which competing semantic theories give similar answers to these questions is the extent to which such theories can be compared, for different answers will result in different explanatory aims and perhaps in incommensurable domains of inquiry. In regard to linguistic science and the way in which linguists think and work, sorting out what the domain of a semantic theory is and what explanatory goals it has are paramount in assessing the success or otherwise of the theory.

This book discusses the two main construals of the explanatory goals of semantic theories. These two construals, I argue below, are not so much in opposition as they are orthogonal. The first understands semantic theories in terms of an interpretive (or hermeneutic) explanatory project, this is often referred to in philosophy of language as externalism. As I detail in the second half of the book, this construal sees the task of a semantic theory as specifying how expressions are to be interpreted. For example, in their two volume study of truth-theoretic semantics, Lepore and Ludwig remark that "there is no question of a standpoint for understanding meaning that is outside of language altogether." That is, they argue that "the most fundamental and powerful devices for representation can obviously not be explicated without the use of just those devices. We can then at best show how they work by showing how they systematically contribute to how we understand sentences in which they appear" (Lepore & Ludwig 2007: 9). This construal, often implicit, is the standard one in philosophy and in formal

semantics, but it is far from being the only one.

The second construal understands semantic theories in terms of the internalist study of the psychological mechanisms in virtue of which meaning production and comprehension are made possible. There is a sense in which there is no competition between the internalist and externalist understanding of semantics, for each approach asks different questions and has different explanatory aims. Unfortunately, this is not the way in which the debate has often been couched, for it is often assumed that both sides are engaged in the same research project. This has led to much misunderstanding and ill-founded criticism from both sides. The internalist side is often criticised for not doing semantics in the way in which the externalist and hermeneutic side assumes semantics should be done. In other words, psychological theories of semantics are often criticised for eschewing the interpretive aspects of semantics that form the basis of the hermeneutic approach to meaning. But these critics fail to see the force and difference in the internalist approach. Regardless of what one thinks of the internalist approach to semantics, its explanatory project both in theory and practice is not hermeneutic but rather scientific in the sense to be spelled out below.

This book argues that a fruitful scientific explanation is one that aims to uncover the underlying mechanisms in virtue of which the observable phenomena are made possible, and that a scientific semantics should be doing just that. I should note at the outset that nothing follows about approaches that are not scientific in this particular sense. There is clearly a great deal to learn from the hermeneutic approach and much good work has been done that takes this approach, but we should not confuse ourselves by claiming that this approach is scientific. Another way to put the matter is as follows. Until recently (perhaps until the mid twentieth century) it was not possible to do semantics qua science, and so it was done in a hermeneutic fashion with much success and offering many insights into the nature of language and mind. However, if (as I detail in chapter 4) we understand scientific explanations to be unearthing the underlying mechanisms in virtue of which the observable phenomena are made possible, then the hermeneutic approach does not offer scientific explanations (and most of its practitioners do not claim to be doing so). The externalist project is one that often aims to provide meta-linguistic semantic descriptions that are essentially interpretive and hermeneutic. Nothing follows about the validity or fecundity of this hermeneutic approach by showing that it is not scientific, except clarifying that it does not aim to unearth the psychological mechanisms in virtue of which meaning comprehension and production are made possible. Showing that this is the case is important in the context of any field that studies meaning, whether it

be linguistics, philosophy, psychology, or cognitive science. To see why this is the case, let me offer a few remarks about the current status of semantics, both within linguistics and in other fields that study the phenomenon of meaningfulness.

The introduction to the recent *Routledge Handbook of Semantics* is titled "Semantics – a theory in search of an object". The editor of the handbook argues that current linguistic semantics is "a subfield whose object – meaning and reference – could hardly be more ambiguous or protean, and which is studied by a highly various scatter of often incompatible theoretical approaches, each of which makes truth-claims, at least implicitly, in favour of its own kind of analysis" (Riemer 2015: 1). The editor of the handbook is also the author of a semantics textbook in which he notes that there is a "lack of disciplinary agreement over the basic theoretical questions" at the core of semantics (Riemer 2010: xiii). Such a diagnosis (and indeed philosophical self reflection of this kind) is rare in linguistics, yet it is accurate, and Riemer remarks that due to this theoretical heterogeneity "it is no surprise that consensus is almost wholly absent about any of the key questions semantics sets out to answer". These questions include, "what meaning as an object of study might, in detail, amount to; how it – whatever 'it' is – should be theoretically approached; how – even pretheoretically – it should be characterized on the level of individual expressions, constructions, and utterances; and what relation semantics should entertain with other fields of enquiry within and outside linguistics" (Riemer 2015: 1-2). Indeed, as Riemer remarks, "it's striking how little explicit theory-evaluation is undertaken by semantics researchers, and how rarely theoretical bridges between different research programmes are even sought, let alone found" (Riemer 2015: 2). This lack of explicit theory evaluation is a primary reason for the lack of consensus on fundamental linguistic phenomena. This is a major hurdle faced by linguistic semantics, and the fact that it is rarely noticed or acknowledged calls for a remedy.

This book aims to provide the beginning of such a remedy by discussing the two major construals of the nature of meaning. By investigating the debate between internalist semanticists and those who advocate for a hermeneutic and interpretive semantics, I hope to clarify the theoretical landscape and provide a rigorous characterisation of what meaning is according to these two schools of thought. I should note that, historically and to this day, linguistics has attracted the interest of many philosophers of language that seek to understand the nature of meaning. However, as Riemer and others have noted, this interest has not been reciprocated. Too few linguists have investigated the philosophical approach to meaning or compared their own semantic theories to those offered by philoso-

phers. This is unfortunate, for the divide between philosophy and linguistics is artificial. In the same way as people working on the nature of time or the interpretation of quantum physics are often in philosophy (and not physics) departments, there are people working on the nature of meaning that are in philosophy (and not in linguistics) departments (or they may be in psychology, literature, anthropology, or sociology departments). Linguists, then, and semanticists in particular, have as much to learn from philosophers of language as philosophers of language have to learn from linguists. But again, this divide is artificial.

This book compares the internalist and externalist approach to semantics, describing their different motivations and theoretical assumptions. I do this from the point of view of explanatory scientific theories. This is an important issue to sort out, for the way in which we construe the nature of meaning is essential for a fecund explanatory language science. I argue that a science of semantics is unlikely to be an externalist one, for reasons having to do with the subject matter and form of externalist and hermeneutic theories. Unlike the internalist approach to semantics, the externalist approach is not usually discussed in terms of scientific explanations, and so my argument might be open to the charge that externalists do not see their enterprise as scientific and thus it is a moot point to compare them to other scientific pursuits. However, as will become evident, there are leading externalists and formal semanticists who explicitly state that their theory is a scientific one. Thus, it is both possible and illuminating to look at the externalist research program from the perspective of scientific explanatory strategies and to ask whether it is a promising avenue in regard to constructing an explanatory scientific theory.

I argue that externalist explanations of meaning are concerned with *ascription and description* of meaning rather than the *mechanisms* of meaning. That is, externalism is not concerned with the mental mechanisms in virtue of which humans produce and comprehend meaning. Therefore, it is not part of the psychological explanation of the mechanisms in virtue of which meaning is made possible. Rather, externalist explanations are a hermeneutic explanatory project in that they are an inherently interpretive project. Works in favour of the internalist approach are currently in the minority, and thus this book also meets the need of describing and helping in advancing a particular understanding of meaning that has been used in the philosophical and linguistics literature for a long time. I provide a critical examination of externalism and present the internalist alternative that, I argue, is better placed to provide the foundation upon which to build a fruitful explanatory science of semantics.

Lastly, I should note that in addition to discussing recent debates, I will also be

discussing many of the classic references in the field because the latter still represent mainstream positions in the discipline. Much can be learned by considering the classic references in light of current debates.

1 Clarifications and methodological preliminaries

The problem of intentionality is the problem of how some entities can be "about" something. That is, words and sentences, among others, display intentionality in that they are about something else; they are said to be a representation of something. The notion of intentionality can be traced back at least as far as Aristotle, though the German philosopher Franz Brentano is generally credited with introducing the notion to contemporary philosophy in the late nineteenth century. Brentano's oft-quoted remark is that "[e]very mental phenomenon is characterized by [...] the intentional (or mental) inexistence of an object" and "reference to a content, direction toward an object." In other words, "[e]very mental phenomenon includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on" (Brentano 1995 [1874]: 68). The usual way to frame the problem of intentionality is in terms of meaning or content. What is the status of the meaning of a sentence over and above its syntactic aspects? What makes it the case that a particular proposition has the content that it does? Is meaning only dependent upon mind-internal properties? Or must we make use of mind-external factors such as the context of the utterance or the speaker's social history in order to determine the meaning?

Those who argue that the relevant and scientifically interesting properties that are involved in meaning are overwhelmingly, though not entirely, within the mind are referred to as internalists. On the other hand, externalists argue that there is something more to meaning than purely mind-internal events and their happenstance connection to the world: externalists insist that the meanings of our words (or sentences, or the contents of our thoughts, etc.) depend on some deep metaphysical (perhaps causal) connection between the mind and objects in the world that are independent of the mind. Externalists argue that a semantic theory needs to provide an account of the relation between linguistic expressions and things in the world (Cann 1993). In other words, the claim is that in order to explain meaning we must provide an account of the purported for-

mal/causal/metaphysical relation between linguistic expressions and the things that they can be used to talk about.

The externalist position has become a widely held position in the philosophy of language. The classic arguments for externalism are found in Putnam (1975), Burge (1979; 1986), and Kripke (1980). Broadly speaking, externalist theories take a model theoretic approach to semantics. They model the interpretation of natural language sentences by making use of set theoretic structures and truth conditions. According to such theories, understanding a sentence involves at least in part the grasping of its truth conditions. This approach is not limited to philosophy of language, for there is a great deal of work in linguistics, for example in formal semantics (Heim & Kratzer 1998; Portner 2005) and formal pragmatics (Kadmon 2001), that takes the truth conditional approach. For example, in his formal semantics textbook, Portner (2005: 11, 13) argues that "meanings are not internal to language, are not in the mind, and are not merely social practices. Rather, they are based in language- and mind-external reality." Moreover, "knowledge of meaning involves (at least) the knowledge of the conditions under which a sentence is true". In other words, "all there is to the meaning of a sentence is its truth-conditions". It should thus be clear at the outset that philosophers *and* linguists take the externalist understanding of meaning seriously both in theory and in practice. Moreover, just like the externalist philosophers discussed in this book, some linguists also aim "to approach meaning as scientists" (Portner 2005: 4) and not, presumably, as members of the project that construes meaning in a hermeneutic or interpretive fashion.

Putnam argues that "a better philosophy and a better science of language" must encompass the "social dimension of cognition" and the "contribution of the environment, other people, and the world" to semantics (Putnam 1975: 193). His Twin Earth thought experiment is the most famous argument in favour of externalism; it claims to show that two subjects can have identical internal psychological mental states but that the content of these states can be different due to particular variations in the environment. Putnam asks us to imagine a world (Twin-Earth) in which water is not composed of H_2O like it is on our world but is rather composed of XYZ. When a person (call him Oscar) says *water* on Earth the word refers to H_2O, but when a different person (call him Twin-Oscar) says *water* in a different place (on Twin-Earth) the word refers to XYZ. This seems intuitively clear; the word *water* refers to what the word is about in that particular environment (so when Oscar utters *water* that word is about H_2O in his environment). Putnam asks what would happen if Oscar is transported to Twin-Earth. Would the word *water* uttered by Oscar on Twin-Earth now refer to H_2O

or XYZ? Notice that the thought experiment legislates that the only change that takes place when Oscar is transported from Earth to Twin-Earth is the change in his environment (i.e., all of his psychological states remain unchanged). Now, Putnam reasons that if knowing the meaning of a term is just a matter of being in a certain psychological state, then *water* on Twin-Earth when uttered by Oscar should refer to H_2O and not to XYZ as we might expect. This is because Oscar's psychological state was fixed on Earth, and if the psychological state fixes the reference then *water* refers to H_2O regardless of the environment the subject is in.

Another way to put the matter is as follows: when Twin-Oscar on Twin-Earth says *water* whilst pointing to a lake that is entirely composed of XYZ, as all watery things are composed of on Twin-Earth, *water* refers to XYZ and not to H_2O. But, Putnam's argument claims, if knowing the meaning of a term is just a matter of being in a certain psychological state then *water* uttered on Twin-Earth by Oscar transported from Earth cannot mean XYZ and must mean H_2O. Something seems to be wrong here. If two people utter the same word in the same environment we expect that word to refer to the same thing. Thus, if we want to hold on to the claim that the meaning of a term determines its reference or extension then, the argument claims, we must concede that, as Putnam famously put it, "[c]ut the pie any way you like, 'meanings' just ain't in the head!" (Putnam 1975: 144). That is, the claim here is that mind-internal properties on their own cannot fix the meanings of words or what their reference is. Note again that for Putnam and others such claims are made within the realm of a science of language. They are claims about the nature of meaning in the mind that they see as having direct bearing on the psychology of meaning.

One might conclude that externalism has to be right, for how could meaning not depend on the outside world? Surely the meaning of the word *elephant* cannot be due to only mind-internal properties. The word is about elephants, it could be argued, which are in the mind-external world, not inside the mind. As we will see, however, internalists argue that there are good reasons to question the externalist claim that meanings are connected to the world in the way in which externalists claim they are. In other words, internalism does not deny the link to the outside world but rather has a different explanation of how our mind generates and interprets semantic content. Internalism argues that, for the purposes of scientific inquiry into language and mind, the internal properties of the human mind are the most relevant and fruitful subject matter. Thus construed, internalism is not so much a solution to the issues that externalists grapple with. Rather, as I detail in the forthcoming chapters, internalism is a different research

program. There is a difference in the sorts of questions that externalists and internalists attempt to answer. This is important to stress at the outset, for there has been a great deal of misunderstanding due to terminological choices. This is because the label *internalism* has been used to refer to several different and opposing research programs.

In the remainder of this chapter, then, I show that the research project that surrounds the debate between individualism and anti-individualism or externalism is separate to the research project of internalism. Too often individualism and internalism are used interchangeably, but the way in which internalists in linguistics and the philosophy of language (such as Noam Chomsky and Paul Pietroski) understand and practice their research project is very different from that of individualism. I don't want to engage in a terminological dispute about what *internalism* means or should mean, rather I want to describe and thus help in preserving a particular understanding of internalism (and of meaning) that has been used in the literature for a long time and has provided interesting and valuable insights into the nature of language and mind. Individualism, anti-individualism and externalism are each concerned with the criteria for the ascription of meaning, whereas internalism is concerned with the underlying generative mechanisms of meaning.

Burge (1986: 3-4) defines individualism as "a view about how kinds are correctly individuated, how their natures are fixed." According to individualism about the mind, then, "the mental natures of all a person's or animal's mental states (and events) are such that there is no necessary or deep individuative relationship between the individual's being in states of those kinds and the nature of the individual's physical or social environments." Individualism is concerned with attribution, with the proper labels that should be assigned to particular mental states. In the case of language, the individualist (and the anti-individualist or externalist) project is part of the approach that aims to provide meta-linguistic semantic descriptions of the linguistic usage of speakers in particular contexts. Internalism, on the other hand, is concerned with the underlying mechanisms in virtue of which language use is made possible. Anti-individualism or externalism argues that there is a deep individuative relationship between mental states and the environment, but this is not a claim about the underlying psychological mechanisms of language but rather about how to interpret particular utterances given their context of use.

Wikforss (2008) argues that externalism is "the thesis that meaning (and content) fails to supervene on internal facts." In other words, "[f]or all natural kind terms T, and all meanings M, the totality of facts that determine that T expresses

M include external facts" (Wikforss 2008: 161). Wikforss then remarks that "internalism, by contrast, is the thesis that the determination basis includes only internal facts." However, the main force and substance of the internalist position discussed in this book is not a mirror image or a negation of the externalist or anti-individualist position. That is, internalism is not concerned with how we determine (qua speakers) what the meaning of a particular utterance is. A fortiori, it does not claim that the determination basis of utterances includes only internal facts. Internalists are skeptical that there is a deep metaphysical relation between the "things in the world" and linguistic expressions, and they dispute the externalist contention that the relations between linguistic expressions and the "things in the world" are desirable or even tractable in a scientific theory of language. The literature in the philosophy of language that discusses individualism is immersed in debates about the correct attribution of semantic content to utterances given a particular context of language use; indeed, this question is the focal point of the individualism/anti-individualism debate. But whatever the merit and explanatory force of the search for the correct attribution conditions, it is clearly separate to internalism. Note again the parallel with linguistics: semanticists also aim to uncover the correct criteria for the attribution of semantic content. But this is not the only way in which to do semantics. Indeed, as detailed in this book, the internalist semantics of biolinguistics has a different understanding of the aims of a semantic theory.

Burge (2003) remarks that even though he at times uses the term *internalism*, he prefers to use the term *individualism*. He admits that some of Chomsky's "arguments for 'internalism' do not directly connect with my objections to the view I designate with the term 'individualism.' So some apparent disagreement may not be real" (Burge 2003: 453). Indeed, for Chomsky's internalism is not the same as individualism. Burge is explicit about this: "Internalism, in *my* sense, concerns not the locus of the psychological states, or the best ways to study them, but whether being in them presupposes individual-environmental relations. It concerns whether the existence and nature of certain psychological kinds depends necessarily on the existence and nature of certain relations to specific kinds or situations in the environment" (Burge 2003: 454, emphasis in original). As we will see below, however, despite Burge's explicit distinction between his sense of internalism (individualism) and Chomsky's internalism, the conflation of the two continues.

Let us now briefly rehearse a classic argument about the individuation of meaning in order to detail and clarify the thesis of individualism and to separate it from internalism as understood in this book. This is important to sort out because too

often internalism is dismissed as being another species of individualism; but it is far from that and in fact offers genuine insights into the nature of meaning in natural language. Moreover, this discussion shows that the concerns of philosophers of language often overlap with the concerns of linguists. The two camps may use different terminology and they may have little contact with each other, but it is clear that they are often involved in the same research program with similar theoretical working assumptions.

Burge (1979) argues against individualism by stressing the necessity of the inclusion of mind-external factors in the descriptions of an individual's mental states. Burge's anti-individualist argument revolves around a Twin-Earth thought experiment in which a person is said to have a large number of propositional attitudes with the content of *arthritis*. So, for example, this person correctly thinks that he has had arthritis for years, or that stiffening joints is a symptom of arthritis. In addition to these attitudes, Burge's thought experiment continues, the person falsely believes that he has developed arthritis in his thigh: this is impossible by definition, as the person is informed by his doctor, since the speech community does not use the term *arthritis* to apply to ailments outside of the joints. Next is the counterfactual supposition of the thought experiment in which we are asked to imagine a second person whose life has proceeded from birth through an identical course of physical events, right to and including the time at which the first person initially reports his fear that he has arthritis in his thigh to his doctor. The only difference that is postulated to exist between the two people is that the latter's community of physicians and informed laymen apply the term *arthritis* not only to arthritis but to various other rheumatoid ailments.

So in the first case, the person falsely believes that he has arthritis in his thigh. Whereas in the second case, the person correctly believes that he has arthritis in his thigh. Burge concludes that the "upshot of these reflections is that the patient's mental contents differ while his entire physical and non-intentional mental histories, considered in isolation from their social context, remain the same", and the "differences seem to stem from differences 'outside' the patient considered as an isolated physical organism." That is, the "difference in his mental contents is attributable to difference in his social environment" (Burge 1979: 79). So the contents of one's thoughts, according to Burge, are individuated by and depend on the meaning of the terms as used in one's linguistic community. That is, "social factors may enter in complex ways into individual psychology and the semantics of idiolects" (Burge 1989: 275). Burge argues that the correct attribution of meaning is impossible without reference to the social context in

which the individual uses that meaning. Burge is here concerned with the criteria for semantic attribution, which is related to but independent of the study of the underlying mechanisms that make the production and comprehension of meaning possible. Michael Devitt makes a similar point when he argues that "thoughts are one thing, their ascription another." He believes that it is a mistake for philosophers to "start with the theory of thought ascription, leaving the theory of thought pretty much to look after itself" (Devitt 1984: 385). The upshot of the difference between individualism and internalism is that one can search for the correct criteria for the individuation of mental states (or for the correct way in which to produce meta-linguistic semantic descriptions) without committing to the nature of the mechanisms that underlie these mental states.

Incidentally, the amalgamation of internalism with a form of individualism (thus excluding the semantic internalism discussed in this book) is not limited to the externalist literature. Consider the internalist (individualist) accounts of Segal (2000), Farkas (2008), Mendola (2008), and Georgalis (2015). These books are perhaps the most notable of the so-called internalist accounts of meaning of the last two decades, but revealingly none deal with internalism in the sense discussed here. The work of Chomsky and others is barely mentioned let alone discussed in sufficient depth (or at all). This is not meant as a criticism. They do not do so because they deal with individualism, with matters of ascription, description, and truth-conditional semantics, and argue against anti-individualism (externalism). Indeed, as Yli-Vakkuri & Hawthorne (2018: 63) remark in a recent critical book on narrow content, "the most natural of our structural conditions" is "nearly universally accepted by internalists." This condition is "that narrow content should be truth-conditional." Chalmers (2003) also argues for this sort of internalist content. And Farkas (2008: 184) concludes her book by remarking that "[t]here is no need for the internalist to give up the idea that contents are truth conditional."

One might wonder whether internalists qua individualists also claim their account to be scientific like some leading externalists do. The answer is that some do so, and so as far as their account is a variation of truth-theoretic semantics, what I will have to say in regard to externalist theories of meaning will apply to individualists too. Mendola (2008) argues that science can settle the debate between internalism and externalism. He understands internalism to be a claim about the content of, say beliefs and desires, and argues that the neuroscience of vision and other sciences support his internalist position. But there is a problem with this strategy that Mendola (2008: 10) himself notices, but he draws the wrong conclusion from it. He says that this strategy is even more popular on

the other side, with externalists, and that for "every internalist who claims to be deferring to cognitive science, there are two externalists who do the same." The problem is that, "even if we take our current cognitive science and psychology as gospel, the deference-to-science strategy doesn't work right now, for internalists or externalists" because "it doesn't now clearly cut one way or the other, or at the very least there is no consensus on how it cuts." But the reason for the lack of consensus is not that science has not yet shown which side is the clear-cut winner. Rather, the reason there is no consensus about whether science supports internalism (individualism) or externalism (anti-individualism) is that science is a different project altogether. The two projects can and should inform one another, but they are distinct. To repeat, the notion of internalism understood as the mirror image of externalism (that is, understood as rejecting anti-individualism but still clinging to reference and content understood truth-conditionally) is very different to the internalist position described in this book.

As noted above, externalism has become a widely held position that is especially popular within the philosophies of mind and language. Indeed, some feel that "externalism has been so successful that the primary focus of today's debate is not so much on whether externalism is right or wrong, but rather on what its implications are" (Wikforss 2008: 158), and that "[o]ver the past 30 years much of the philosophical community has become persuaded of the truth of content externalism" (Majors & Sawyer 2005: 257). Externalism has thus become "almost an orthodoxy in the philosophy of mind" (Farkas 2003: 187). Since the internalist position is very much in the minority, it is necessary to begin by outlining its conception of a semantic theory before it can be compared with the received view of externalism. A clear understanding of this strand of internalism is essential, for not only is it a minority view but it is also widely misunderstood. After comparing the two approaches to semantics, I will argue that internalism is significantly more promising in regard to constructing an explanatory scientific theory of meaning. Note again that what follows is that the externalist (hermeneutic and interpretive) approach is a different research project to that of a semantics construed scientifically, and so nothing follows about the validity or fecundity of externalism construed hermeneutically.

2 Internalism

In this chapter I detail the internalist approach, which is taken by generative linguistics as well as the broader, generative-oriented, biolinguistics program. Language is here regarded as an internal computational system that produces a set of hierarchically structured expressions that are employed by the systems of thought and the sensorimotor systems to yield language production and comprehension. I discuss the work in internalist semantics of Paul Pietroski and others according to which linguistic meanings are computational instructions to build monadic concepts.

Internalism, as the name suggests, studies internal states, including those that in philosophy are regarded as mental states. Chomsky (2003) makes clear that internalism is not the doctrine that denies that mental states are individuated by reference to the subject's environment, nor is it the doctrine that holds that subjects in the same internal states are therefore in the same mental states. That is, as mentioned above, internalism is not the same as individualism. Rather, internalism is "an explanatory strategy that makes the internal structure and constitution of the organism a basis for the investigation of its external function and the ways in which it is embedded in an environment" (Hinzen 2006: 139). In other words, internalism "is primarily a *conjecture about a proper object of the scientific study of language* (which internalists claim to be *I-language*)" (Lohndal & Narita 2009: 324, emphasis in original). This chapter will outline what this amounts to in the case of semantics.

2.1 E-language and I-language

Biolinguistics treats language as an internal computational system, a recursive mechanism that produces a potentially infinite set of hierarchically structured expressions that are employed by the conceptual-intentional systems (systems of thought) and the sensorimotor systems to yield language production and comprehension. As I detail below, this particular functional design of the language faculty is strongly shaped by its interface with the systems of thought, rather than by the peripheral process of externalisation inherent in the link with the sensori-

motor systems (Chomsky 2013; Hinzen 2013; Asoulin 2016; Berwick & Chomsky 2016). Biolinguistics takes its object of study to be the underlying mechanisms of language, which are a subsystem of our cognitive system and are composed of a computational system (called an I-language) that is encoded in individual brains. The subject matter of biolinguistics (and internalism) is thus competence, as opposed to performance. As Chomsky put it in an oft-quoted phrase, generative linguistics is primarily concerned with an ideal speaker/hearer who resides in "a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance" (Chomsky 1965: 3). Competence, then, refers to the speaker/hearer's knowledge of his/her language, whereas performance refers to the actual use of this knowledge by a particular person. The actual use of one's linguistic knowledge in language production and comprehension involves many other factors, only one of which is one's competence, and it is only under strict idealisation conditions that performance might be seen as reflecting competence. Chomsky (1986) developed a different characterisation of the competence/performance distinction, a clearer and more useful distinction that is still used today: I-Language versus E-language.

Externalised (E-) language refers to the actual or potential speech events. From the E-language point of view, a grammar is a collection of descriptive statements concerning performance; the grammar describes or taxonomises the corpus of linguistic performance data. This is the way language is studied in structural and descriptive linguistics, behavioural psychology, and some branches of cognitive science, where language is viewed as a collection of linguistic forms (words or sentences) that are paired with meanings. Even though this description glosses over the subtleties of and the differences between specific E-language approaches, the main thread of them all is the view of language as "the totality of utterances that can be made in a speech community" (Bloomfield); or language as "a pairing of sentences and meanings over an infinite range", where the language is used by a population when certain regularities hold among the population with respect to the language and are sustained by an interest in communication (Lewis).[1] What the E-language approaches share is the view that language can be understood (indeed often it is claimed that it exists) independently of the properties of the mind/brain. That is, language is understood as a collection of actions or behaviours, and "a grammar is a collection of descriptive statements concerning

[1] See Chomsky (1986: 19) for discussion and more references.

the E-language, the actual or potential speech events (perhaps along with some account of their context of use or semantic content)" Chomsky (1986: 20).

In other words, this approach sees a grammar as a function that enumerates the elements of the E-language. But this function need not be unique. From the E-language perspective, there need not be one *real* or *correct* grammar that corresponds to the corpus data: as long as it yields a correct description of the corpus data, any number of grammars could in principle apply. Lewis, for example, says that he can find no way to "make objective sense of the assertion that a grammar Γ is used by population P whereas another grammar Γ′ which generates the same language as Γ, is not" (Lewis 1975: 20). Lewis believes that a language is an abstract, formal system that a population selects by convention (Lewis 1969). Another manifestation of E-language can be seen in Devitt & Sterelny (1989), who argue that rather than being about competence, linguistics is about the properties and relations of *observable* linguistic symbols (see also Devitt 2006). According to the E-language conception, then, language is, as it were, out there, it is not intimately related to the mind. Deacon, for example, argues that in contrast to the claim of generative linguistics that support for language acquisition originates inside the brain (in the language faculty), "the extra support for language learning is vested neither in the brain of the child nor in the brains of parents or teachers, but *outside brains, in language itself*" (Deacon 1997: 105, emphasis mine). The E-language is the real object of study here, not the grammar which generated it, which is a derivative notion because it is assumed that any grammar is suitable so long as it correctly generates the observable corpus.

On the internalised (I-) language perspective, however, there is a particular grammar that generates and is responsible for the observable corpus of utterances. More precisely, it generates a set of structural descriptions that provide the basis for interpretation. It is the generative grammar that is the object of study (as opposed to the set generated by the grammar), and this grammar qua generative computational device is instantiated in the brain. Language is thus conceived as some real structure in the brain of the speaker/hearer that is responsible for (indeed it is) the language that that speaker/hearer knows. So, unlike the E-language conception of language, a generative grammar qua I-language is a theory of a real mental structure to which "questions of truth and falsity arise [...] as they do for any scientific theory" Chomsky (1986: 22). As we'll see in detail below, the I-language approach, which biolinguistics takes, sees the proper subject matter of a scientific linguistics to be the knowledge a speaker/hearer has of his or her language, the knowledge that underlies and makes possible, along with other factors, that speaker/hearer's language production and comprehen-

sion. This is also the research program of internalism.

Let us be clear about the relation between the internalism/externalism distinction and the I-language/E-language distinction. Internalism in the sense understood by the authors discussed in this book is clearly and explicitly rooted in the I-language approach to semantics. The E-language approach, on the other hand, is exemplified by several externalists that are cited in Chomsky (1986) where the I-language/E-language distinction was first articulated. The labels I-language and E-language, then, denote approaches to the study of language and meaning. There is more than one way to flesh out an I-language or E-language approach to semantics, and I discuss some of these variations below. I should also note that there are other criticisms of the externalist position apart from those offered here from the point of view of internalist semantics, so it of course does not follow that one must agree with biolinguistics in order to see the problems with externalist semantics of the Putnam or Davidson sort. For example, as discussed below, Paul Horwich offers both a critique of externalist semantics and an alternative semantic theory. But his theory still clings to an externalist (in the E-language sense) understanding of meaning.

With that in mind, let us now explore the nature of I-language before moving on to internalist semantics. An I-language is a computational system that is in the mind of individual language users. It is a generative procedure that outputs structural descriptions that provide the basis for interpretation. There is a stress here on the intensional nature of particular I-languages, meaning that there is a specific procedure encoded in the mind that generates the structural descriptions; this is in contrast to the extensional nature of E-language grammars. Another way to put the matter is in terms of formal mathematics, in which a sequence can be defined extensionally by listing its members, say $0, 1, 1, 2, 3, 5, 8, 13\ldots$, or intensionally by providing a formula that generates the members of the sequence, say the formula $F_n = F_{(n-1)} + F_{(n-2)}$ that generates all and only the numbers of the Fibonacci sequence. An intensional definition is much more useful for large sets and is essential for potentially infinite sets like the ones associated with natural languages. This analogy should not be taken too literally, for as we'll see in chapter 3, there are crucial differences between formal languages and natural languages.

The internal computational processes of the language faculty generate linguistic objects that are employed by the conceptual-intentional systems (systems of thought) and the sensorimotor systems. Lexical items, then, and all expressions generated from them, must have properties that are interpretable at both these interfaces. Notice that on this view the language faculty is embedded within, but

separate from, the performance systems. So an I-language is a device that generates structured expressions of the form $Exp =< Phon, Sem >$, where $Phon$ provides the sound instructions of which the sensorimotor systems make use, and Sem provides the meaning instructions of which the systems of thought make use. $Phon$ contains information relating to linear precedence, stress, temporal order, prosodic and syllable structure, and other articulatory features. Sem contains information relating to event and quantification structure, and certain arrays of semantic features. The term *instructions* is here used in a technical sense, so that to say that

> [...] phonetic features are "instructions" to sensorimotor systems at the interface is not to say that they have the form "Move the tongue in such-and-such a way" or "Perform such-and-such analysis of signals." Rather, it expresses the hypothesis that the features provide information in the form required for the sensorimotor systems to function in language-independent ways. (Chomsky 2000a: 91)[2]

The same is true for the semantic features at the Sem interface, which are not "instructions" to the conceptual-intentional systems of the form "this pronounced word means such-and-such" or "link this pronounced phrase with this concept". Rather, as detailed in the next section, the Sem interface is part of the procedure that generates instructions to build new mental representations.

The expression Exp is generated by the operation Merge, which takes objects already constructed and constructs from them a new object. So, for example, $Merge(X, Y)$ will yield the unordered set $\{X, Y\}$. The structure-building operation Merge follows the principle of Minimal Computation (compute and articulate as little as possible), for it is the simplest possible computational operation for the task at hand (Berwick & Chomsky 2016; Chomsky 2016a). There are two cases of Merge: External Merge refers to the operation where two syntactic objects are merged but where neither one is part of the other. Internal Merge, on the other hand, refers to the operation where one of the syntactic objects is part of the other. For example, Internal Merge takes place when a syntactic object is combined with the set that contains it: so if $Merge(X, Y)$ yields $Z = \{X, Y\}$, then $Merge(Z, X)$ yields $\{X, Y\{X\}\}$. For concreteness, take the following simplified example of External Merge. *The silver saucer broke yesterday* is produced by Merge as follows: lexical items are merged to (separately) create *The*, *silver*,

[2] For more on the phonetic implementation of phonological features, see Halle (1983; 1995). See Kenstowicz (1994); Hale & Reiss (2008); Volenec & Reiss (2020) for an overview of generative phonology.

and *saucer*. Then *silver* and *saucer* are merged to created the Noun Phrase (NP) *silver saucer*. Then *silver saucer* is merged with *the* to create the NP *the silver saucer*. Then that NP is combined with the Verb Phrase (VP) *broke yesterday* (which was produced by Merge when lexical items were merged to create *broke* and lexical items were merged to create *yesterday*, and then *broke* and *yesterday* were merged together to create the VP *broke yesterday*).[3]

As for Internal Merge, suppose we merged *which saucer* with *John broke which saucer* to produce *which saucer John broke which saucer*, which via further computations is then externalised as *which saucer did John break*. Before externalisation, there are two copies of the same linguistic object (X): the original one and the displaced one. They are both essential for interpretation. As Chomsky has remarked in various places when he discusses Merge, this is an example of the ubiquitous phenomenon of displacement in language, where phrases are heard in only one place but are interpreted both there and in another place. So we interpret the above sentence to mean "for which X, John broke the saucer X". Merge, then, defined as recursive set-formation, produces hierarchical structures and allows for the unbounded embedding of these structures (namely, it allows for discrete infinity).

There are independent reasons to believe that cognitive processes satisfy the principle of Minimal Computation (see Cherniak 1994; Cherniak, Mokhtarzada & Nodelman 2002), and since Merge satisfies this principle and is able to account for the underlying mechanisms of language, we have strong grounds for its existence as a core computational principle of human language. Furthermore, as Chomsky (2013) shows considering examples such as the above where two copies of *which saucer* are required for the interpretation of the sentence, Merge yields structures suited for interpretation at the conceptual-intentional interface but "these are clearly the wrong structures for the SM [sensorimotor] system: universally in language, only the structurally prominent copy is pronounced" (Chomsky 2013: 41). That is, the second copy must be deleted when it is transferred to the sensorimotor interface resulting in articulated sentences having gaps that create problems for language comprehension and communication but that are necessary for interpretation at the conceptual-intentional interface. These are the so-called filler-gap problems,[4] where the hearer has to figure out where the unarticulated element is in order to parse and interpret the sentence correctly. There is thus an asymmetry between the interfaces in favour of the semantic side,

[3] For recent discussion of Merge, see C. Collins (2017); Chomsky, Gallego & Ott (2019)

[4] See Sprouse & Hornstein (2013) for a recent collection of work on long-distance filler-gap dependencies.

pushing externalisation (via $Phon$) to the periphery. If the language faculty is structured in this way then it follows that the underlying computational mechanisms of language "will provide structures appropriate for semantic-pragmatic interpretation but that yield difficulties for perception (hence communication)" (Chomsky 2013: 41). In other words, the design of language favours minimal computation, often at the expense of ease of communication (Sigurðsson 2004; Burton-Roberts 2011; Asoulin 2016; 2020).

2.2 Internalist semantics

The expression Exp is of course not the same as a linguistic utterance but rather provides the information required for the sensorimotor systems and the systems of thought to function, largely in language-independent ways. Since these two systems operate independently of (but at times in close interaction with) the faculty of language, a mapping to each interface is necessary, for these two systems have different and often conflicting requirements. The systems of thought require a particular sort of hierarchical structure in order to, for example, calculate relations such as scope; the sensorimotor systems, on the other hand, often require the elimination of this hierarchy because, for example, pronunciation must take place serially. The instructions at the Sem interface that are interpreted by the performance systems are used in acts of talking and thinking about the world – in, say, reasoning or organising action. Linguistic expressions, then, provide a perspective (in the form of a conceptual structure) on the world, for it is only via language that certain perspectives are available to us and to our thought processes. This is in line with a long rationalist tradition in the philosophy of language and linguistics (Chomsky 1966), most famously articulated by Humboldt in the nineteenth century, according to which language provides humans with a *Weltansicht* or worldview that allows us to form the concepts with which we think certain kinds of thought (but, crucially, not *all* kinds of thought: for we share many kinds of thought processes with animals that do not have language).

In his recent study of Humboldt, Underhill (2009) remarks that Humboldt's "rich and dynamic model of language" is one "in which the individual both shapes and is shaped by the *organ of speech*" (Underhill 2009: xi, emphasis in original). The worldview concept of *Weltansicht*, which forms the cornerstone of Humboldt's linguistic philosophy, is understood as "the configuration of concepts which allow conceptual thought" of a certain kind (Underhill 2009: 56). Language is an instrument of thought in this sense (Asoulin 2016), but note that this is not a Whorfian claim of linguistic determinism, for thought is certainly inde-

pendent of particular natural languages, and what can be expressed or thought by a speaker of one language can certainly be expressed or thought by a speaker of a very different language. As Underhill (2009: 57) remarks, Whorfian claims are merely "weak echoes of Humboldt's voice". Language provides us with a unique way of thinking and talking about the world that is unavailable to non-linguistic animals. Though of course animals have thoughts of many kinds (many of which are shared with humans), but since they lack the language faculty there is a specific kind of thought that they lack (Hinzen 2013; Asoulin 2019). Let us see how this rationalist understanding of the role of language in cognition is manifested in current biolinguistics and internalist semantics.

As mentioned, an I-language is a device that generates structured expressions of the form $Exp =< Phon, Sem >$ with a double interface property: they have phonological and semantic features through which the linguistic computations can interact with other cognitive systems. But the link to and influence of each interface is not symmetrical, for there is mounting evidence that there is an asymmetry between the interfaces in favour of the semantic side, pushing externalisation via $Phon$ to the periphery (Chomsky 2013; Berwick & Chomsky 2016; Chomsky 2016a; Asoulin 2016; 2020). Merge implements the basic properties of I-language (Chomsky, Gallego & Ott 2019). Collins & Stabler (2016) show that all the essential syntactic operations, such as c-command, can be formally defined in terms of Merge. For reasons of computational efficiency, the computations of Merge should apply freely so that the only constraints imposed on them are those derived from the interfaces with the external systems. There are independent reasons to believe that cognitive processes satisfy this principle of Minimal Computation (Cherniak 1994; Cherniak, Mokhtarzada & Nodelman 2002; Chomsky 2016a). Investigation of the structures generated by Merge shows that they are well suited to the Sem interface (hence, for internal thought), but cause predictable problems at the $Phon$ interface. In other words, the normal course of the derivation generated by Merge simply proceeds towards Sem, then at some point in the derivation some parts of the expression are sent to $Phon$ for externalisation. The 'point' of the derivation is the generation of interpretable structures: its externalisation via sound or sign is secondary at best.

The Sem interface is the way in which biolinguistics and internalism explain meaning in natural language. A theory of Sem must satisfy three basic conditions of adequacy, so that in order to capture what the language faculty determines about the meaning of an expression, Sem must "be universal, in that any thought expressible in a human language is representable in it; an interface, in that these representations have an interpretation in terms of other systems of

the mind/brain involved in thought, referring, planning, and so on; and uniform" (Chomsky 1995: 21). Sem must be uniform "for all languages, so as to capture all and only the properties of the system of language as such" (Chomsky 1995: 21). In other words, the way in which the meanings at Sem are generated (and then sent to the conceptual-intentional interface) is uniform in the sense that any meaning generated via the language faculty is expressible in any natural language. Note the stress on all and only the properties of the system of language: the language faculty allows humans to use available concepts (some of which are shared with other animals) to generate formally new concepts. The claim is not that Sem is the interface of all conceptual content or of all of thought.

Pietroski (2008; 2010; 2018) has developed one of the most interesting and detailed accounts of an internalist semantics, the leading idea of which is that "in the course of language acquisition, humans use available concepts to *introduce* formally new concepts that can be fetched via lexical items and combined via certain operations that are invoked by the human faculty of language" (Pietroski 2010: 247, emphasis in original). That is, meanings are (internal, and unconscious) instructions for how to access and assemble concepts of a special sort. Meaning is here understood not in an extensional sense but rather in terms of the cognitive resources (the computational procedures) that humans deploy in *generating* the meanings. So, for example, the Sem of *white sheep* is an instruction to fetch a concept from each lexical address and then conjoin them. There are a number of steps and notions here that require unpacking: (i) what is a concept, (ii) how is a concept lexicalised, and (iii) how are these lexicalised concepts conjoined. I discuss each in turn below.

Concepts are, roughly, constituents of mental states. To give Fodor's favourite example, believing that cats are animals is a paradigmatic mental state, and the concept *animal* is a constituent of the belief that *cats are animals*. The latter is a proposition, and propositions are generally understood to be structured objects of which concepts are the constituents. As Fodor has discussed in his various works on concepts (for example, Fodor 1998; 2003; Fodor & Pylyshyn 2015; see Murphy 2002 for an overview), some concepts are structured and some are primitive. So the concept *white cat* is a structured concept that might include the two primitive concepts *white* and *cat*. The meaning of a structured concept depends on its primitive elements and on the way in which they are combined. But not any combination is possible: there is a syntax that determines how concepts can (and cannot) be combined. Frege's famous metaphor of saturating concepts is helpful here. Statements can be thought of in the same way as mathematical equations, argued Frege (1980 [1892]), in that they are split into two parts (a function and an

argument). Consider the sentence *Caesar conquered Gaul*. The first part (*Caesar*) is the subject expression, which can stand on its own, but the second part (*conquered Gaul*), which is the predicate expression, is in need of supplementation or saturation for it contains an empty place that needs to be filled in. So a proper name like *Caesar* is said to saturate the function *conquered Gaul* by filling in the empty place, giving a complete sense.

As Pietroski (2010: 249) discusses, singular concepts can saturate concepts like ARRIVED(X) and SAW(X, Y), which are used to classify and relate represented individuals, allowing humans to form sentential representations like ARRIVED(BRUTUS) and SAW(BRUTUS, CAESAR). Abstracting away from completed concepts leaves what Pietroski calls a sentence frame "that can be described as an unsaturated concept whose adicity is the number of saturable positions: ARRIVED(X) is monadic, SAW(X, Y) is dyadic, GIVE(X, Y, Z) is triadic, etc." (Pietroski 2010: 249). There is of course a limit to the saturable positions that natural language concepts can possess, and Pietroski argues that tetradic concepts may be common (compare the difference between selling and giving). As he puts it, "we seem to have higher-order numeric/set-theoretic/quantificational concepts that can be saturated by monadic concepts, as in THREE/INCLUDE/MOST[BROWN(X), COW(X)]. In short, concepts compose and exhibit a limited hierarchy of types" (Pietroski 2010: 249).

Now, as Fodor (1975) famously argued, there are parallels between propositions and sentences and between words and concepts (and thus between thought and language). That is, "propositions are what (declarative) sentences *express*, and (excepting idioms, metaphors, and the like), *which* proposition a sentence expresses is determined by its syntax and its inventory of constituents" (Fodor & Pylyshyn 2015: 8, emphasis in original). But how far do these parallels go? How much of conceptual thought is influenced, constructed, or determined by the computational procedures of I-language? If our concepts are parallel to linguistic expressions in their systematicity and productivity, how did these concepts emerge? There is perhaps a spectrum of answers to these questions, but in general there are two answers: either the concepts were there prior to lexicalisation or else the process of lexicalisation introduced new sets (or new types) of concepts. The biolinguistic and internalist semantics claim is of the latter sort. To put the matter in Pietroski's terms, already existing concepts (many of which we share with other animals) are lexicalised and in the process distinctively new concepts are produced that we are then able to combine to form linguistic expressions. This process of lexicalisation and concatenation is part of the explanation of the creative aspect of language use (Chomsky 1966; McGilvray 2001; 2005; Asoulin 2013).

So how is a concept lexicalised? There are several ways in which to flesh out the idea that lexicalisation is the process by which pre-existing concepts are used to introduce formally new concepts. The externalist answer, which will be discussed in the next chapter, is a compositional theory of meaning modelled on the work of Davidson (1967; 1973) and Montague (1974); but let us continue with an internalist answer. Pietroski's answer moves away from the Fregean idea that combining expressions is an instruction to saturate a concept and towards a Conjunctivist account of linguistic composition (Hornstein & Pietroski 2009; Pietroski 2018). According to the latter account, lexicalisation is not a process in which a previously available concept is merely labelled using a lexical item that inherits its content from the concept itself. Rather lexicalisation is a device for accessing previously available concepts which become lexical items that are used as input to I-language operations that combine the lexical items in specific ways to introduce new formally distinct concepts. Accordingly, the Sem of any expression $Exp =< Phon, Sem >$ is not a concept that is paired with a pronunciation. Indeed, as Pietroski puts it, "evaluating SEMs as if they were concepts may be a category mistake, like evaluating *an instruction to fetch* a rabbit as male or female" (Pietroski 2010: 252, emphasis in original). So a Sem is an instruction to fetch (i.e., lexicalise) a previously available concept that is then used to build a formally new concept(s). This formally new concept will be stored in the mind somehow (and perhaps be recombined with other concepts to create yet more formally new concepts), but the Sem itself is not a concept.

Another way to put the matter is as follows. Humans possess a great variety of pre-lexical mental representations (many of which we share with other animals). On the Conjunctivist account, these pre-lexical mental representations are linked to formally distinct but analytically similar concepts. The latter are sometimes referred to as I-concepts (Jackendoff 1989; 1990) to signal that the way in which these concepts are to be studied is on the model of the study of language signalled by the use of I-language as opposed to E-language. Thus, "the repertoire of I-concepts expressed by sentences cannot be mentally encoded as a list, but must be characterized in terms of a finite set of mental primitives and a finite set of principles of mental combination that collectively describe the set of possible I-concepts expressed by sentences" (Jackendoff 1990: 9). I-concepts, then, are a uniquely human *subset of concepts* that humans can use to think about the world. It thus follows that "there may be many human concepts that cannot be fetched or assembled via SEMs: acquirable I-languages may not interface with all the concepts that humans enjoy", for "there may be ways of assembling concepts that SEMs cannot invoke" (Lohndal & Pietroski 2018: 325).

If lexical meanings are understood to be instructions to fetch concepts, then phrases are understood to be instructions to combine these fetched concepts in specific ways. So how are these lexicalised concepts conjoined? Let us look at two simple examples taken from Pietroski (2010: 249-250) (see also Pietroski 2018). Consider the phrase *kick a brown cow*. The acquisition of *kick* might involve the process in which a dyadic concept like KICK(X, Y) is paired with a *Phon*, stored in the mind, and then used to introduce a concept of events, KICK(E). The latter is then fetched and conjoined with the other concepts of the phrase *kick a brown cow*, which were fetched in a similar way. *Kick a brown cow* can then be analysed in terms of the instructions to build concepts like •[KICK(E), ∃•[PATIENT(E, X), •[BROWN(X), COW(X)]]]. "•" indicates a conjunction operator, PATIENT(E, X) is a concept of a thematic relation exhibited by certain events and participants affected in those events, and "∃" existentially closes the participant variable (X). In the same way, *kick a carrot to a cow* can be analysed in terms of the instructions to build a monadic concept like •[•[KICK(E), ∃•[PATIENT(E, X), CARROT(X)]], ∃•[RECIPIENT(E, X), COW(X)]], which applies to kicks that have carrots as patients and cows as recipients.

The internalist semantics claim, then, is that understanding an expression of I-language (or perceiving its meaning) is a matter of (unconsciously) recognising that that expression is an instruction to construct concepts of a special kind. This explanation of meaning also offers an explanation for the creative aspect of language use, for it suggests the procedure by which we combine concepts in recursively productive ways to yield formally new concepts and phrases.

It may seem as if this internalist explanation of meaning merely passes the buck to concepts, thus avoiding the crucial explanatory question as to what makes a proposition mean what it does. That is, claiming that lexical meanings have as primitives non-lexical concepts that are then combined to produce formally new concepts might be criticised for assuming (and thus leaving unexplained) the meanings of the primitives. This criticism is unwarranted for several reasons. First, we cannot expect a theory of semantics to explain every primitive, for at some point we move beyond semantics and into cognitive or perceptual psychology (and ultimately into neuroscience). We are concerned here with lexical meaning, not with pre-lexical or non-linguistic conceptual structure. If biolinguistics is on the right track, then it follows that lexical concepts are a distinct subset of concepts available to humans, a subset that is uniquely human in comparison to what is shared with other animals. But what about the other concepts? That depends on what one takes non-linguistic (or pre-lexical) concepts to be, if one takes them to be concepts at all. Since concepts are constituents of thoughts,

the debate about the nature of non-linguistic concepts often overlaps with the debate about whether animals can think or with the debate about whether language is the medium of thought. There is a tradition in philosophy that argues that all thought requires language (Malcolm 1972; Davidson 1975; 1982; Dummett 1989; McDowell 1994), whereas others have agreed with Fodor (1975: 56) that the "obvious (and, I should have thought, sufficient) refutation of the claim that natural languages are the medium of thought is that there are nonverbal organisms that think" (see also Ryle 1968; Slezak 2002; de Waal 2016). I do not want to weigh in on the debate of whether cognitive processes as understood by cognitive scientists are the same as what philosophers such as Malcolm and Davidson understand to be thought processes. Whether or not we should conceive of the cognitive processes that humans share with animals as thoughts with concepts is orthogonal to my concerns here. I've argued elsewhere that humans share with animals a great deal of thought processes but that humans also possess a unique type of thought that is not available to animals without a language faculty (Asoulin 2016; 2019) (see also Gallistel 1991; 2011).

An interesting corollary of the internalist understanding of meanings as effectively tools for reformatting pre-lexical concepts to yield formally new concepts is that these newly created concepts lie at a greater remove from the environment than the concepts that get lexicalised. As Pietroski (2010) discusses, pre-lexical concepts are different to the newly created concepts because the latter require an additional kind of abstraction. That is, the creative aspect of lexicalisation "promotes cognitive integration, by giving us new concepts that fit together in recursively productive ways, [but] at the cost of giving us concepts that fit the world less well than the concepts initially lexicalized" (Pietroski 2010: 250). The process of lexicalisation results in the new concepts (and the phrases of which they form the constituents) not fitting the world in the same way that pre-lexical concepts do. The pre-lexical concepts that we share with other animals, on the other hand, do exhibit "a functioning isomorphism between processes within the brain or mind [...] and an aspect of the environment to which those processes adapt the animal's behavior" (Gallistel 1990: 1-2). There is "a mapping from external entities or events (temporal intervals, numerosities of sets, rates of food occurrence, shapes of patterns, chemical characteristics of foods, members of a matrilineal family within a monkey group, and so on) to mental or neural variables that serve as representatives of those entities" (Gallistel 1990: 2). In other words, in the case of pre-lexical concepts, "the concept of a mental or neural representation depend[s] on the demonstration of a mapping from world variables to mental or neural variables and on a formal correspondence between opera-

tions in the two domains" (Gallistel 1990: 4). But lexical concepts and natural language sentences do not work that way. As we'll see later on, however, the kind of referential semantics that externalism proposes assumes that natural language is referential in this problematic way. But if "our 'I-fetchable' concepts are systematically combinable because they were *introduced* to ensure such combinability, then these concepts lie at greater remove from the environment than the concepts we lexicalize" (Pietroski 2010: 266, emphasis in original).

There is thus a lack of one-to-one (or even one-to-many) relations between mental representations and things in the world when it comes to natural language. As Chomsky (2000; 2003a; 2003b; 2016) has discussed at length, this is clear for even the simplest words. Take Chomsky's famous example of London (Chomsky 2000: 37). He remarks that London is of course not a fiction created by our minds but "considering it as London – that is, through the perspective of a city name, a particular type of linguistic expression – we accord it curious properties" such as the following:

> […] we allow that under some circumstances, it could be completely destroyed and rebuilt somewhere else, years or even millennia later, still being London, that same city. [….] We can regard London with or without regard to its population: from one point of view, it is the same city if its people desert it; from another, we can say that London came to have a harsher feel to it through the Thatcher years, a comment on how people act and live. Referring to London, we can be talking about a location or area, people who sometimes live there, the air above it (but not too high), buildings, institutions, etc., in various combinations (as in *London is so unhappy, ugly, and polluted that it should be destroyed and rebuilt 100 miles away*, still being the same city).

We use terms such as *London* to talk and think about the mind-external world but "there neither are nor are believed to be things-in-the-world with the properties of the intricate modes of reference that a city name encapsulates" (Chomsky 2000: 37). Another way to put the matter is as follows. The properties that we attribute to the world via our everyday language are for science the products of our minds (see McGilvray 2002 for discussion). That does not mean that the mind creates all objects and thus there are no mind-independent objects out there. Rather, it means that the scientific study of the way in which humans conceive the world will be internalist.

Language-world relations are assumed in externalist semantics, and it is almost as often assumed that generative linguistics can provide an analysis of the

language side of this relation. D'Ambrosio (2019: 214) understands semantics to be a "theory of the contents of natural-language expressions, where such contents are ultimately found in the world, or constructed mathematically out of pieces of reality." On this view, "semantics makes use of lexical postulates that express genuine relations between words and objects or collections of objects, and from these premisses, semanticists derive theorems about what the world must look like for natural-language sentences to be true." Semantics is thus "partly a metaphysical theory—it is a version of the theory of truthmaking". D'Ambrosio claims that his account of the verbs we use in our semantic theorising (such as *refers (to)*, *applies (to)*, and *is true (of)*) shows that internalist and externalist semantics are compatible. King (2007; 2018) also argues that one can accept the central features of generative linguistics and still endorse an externalist semantics for I-languages. But this is far from clear. J. Collins (2007: 805) discusses the "presumption that—at some level and in some way—the structures specified by [generative] syntactic theory mesh with or support our conception of content/linguistic meaning as grounded in our first-person understanding of our communicative speech acts." This presumption is currently shared by many top philosophers of language, but Collins shows that generative syntactic structure both provides too much and too little to serve as the structural basis for the notion of content as understood in philosophy. He argues that the philosopher's "content, as it were, is the result of a massive cognitive interaction effect as opposed to an isomorphic map onto syntactic structure" (J. Collins 2007: 806).

2.3 What about mind-world relations?

A common misconception of internalist semantics is that it sees as irrelevant or ignores the environment of the speaker or the context of the speech act. A corollary of this misconception is the argument favouring externalism that argues that, if meaning is construed as "an internal phenomenon" that rejects relations to the extra-mental world, then there will be "a wholesale and incomprehensible relativism concerning truth" and "a total collapse in our belief in the existence of the external world" (Ferguson 2009: 299, quoting Millikan 1984: 7). That is, "if meaning is equated with intensions of the individual language users" then the extension of words is derivative of these mental representations and thus "this derived extension is never actually put into direct correspondence with external objects but only with user's *concepts* of such objects" (Ferguson 2009: 299-300, emphasis in original). In other words, the externalist worry is that if meaning is construed as an internal phenomenon then words are not connected to the ex-

ternal world but only to the user's internal mental representations of them and thus there is no way to distinguish between a true representation of something in the world and a misrepresentation.

Moreover, it is claimed that realism itself is at stake. Millikan is explicit about this point: she argues that the "assumption that whatever meaning is, it must determine reference or extension is the very essence of realism" (Millikan 1984: 329). She takes issue with what she calls meaning rationalism, with the claim that we can know a priori "that in seeming to think or talk about something we *are* thinking or talking about – *anything at all*." That is, meaning rationalism claims that we can "know a priori *that we mean*" and "know a priori or with Cartesian certainty *what* it is that we are thinking or talking about" (Millikan 1984: 10, emphasis in original). This is "a view of meaning that is completely internal", a "theory of meaning that sees the extension of words as a function of the intensions of individual speakers, with no way to ensure that these intensions actually correspond to anything in the external world" (Ferguson 2009: 299). Millikan claims that meaning rationalism "permeates nearly every nook and cranny of our philosophical tradition" and that "[i]n order even to come to *comprehend* what meaning rationalism is, what various forms it can take, it is necessary forcefully to fling down on the table something with which to contrast it" (Millikan 1984: 92, emphasis in original). Millikan's own externalist theory of meaning (teleosemantics) is then offered as a contrast. I should note that meaning rationalism, understood as an internalist (individualist) view of meaning, certainly no longer "permeates nearly every nook and cranny of our philosophical tradition": the tables have completely turned. As noted above, it is now claimed that "externalism has been so successful that the primary focus of today's debate is not so much on whether externalism is right or wrong, but rather on what its implications are" (Wikforss 2008: 158).

The externalist worry, then, is that an internalist view of meaning and concepts either leads to an attack on realism or to an idealism of some sort. Millikan remarks that

> The important thing is that meaning rationalism led to the conclusion that all our genuine concepts are of things that have a most peculiar ontological status. They are things that *are* and that can be *known* to be, yet that have no necessary relation to the actual world. They are things that do not need the world about which we make ordinary judgements in order to be. They must be Platonic forms, or reified "concepts" or reified "meanings" or things having "intentional inexistence" or reified "possibilities" – or else they must be *nothing at all!* (Millikan 1984: 328, emphasis in original)

This worry about internalist theories of meaning of any sort is not limited to Millikan. Fodor (2000; 2007), for example, worries that Chomsky's internalist semantics is a sort of idealism about meaning. Fodor understands internalist semantics to eschew relations between concepts and the world in favour of relations among the concepts themselves, and he believes that Chomsky is motivated by epistemological concerns. So, as Fodor paints the scene, since knowledge involves representation, the question is how we can know the world independently of the ways in which we represent it. That is, "if representation is itself a kind of mind-world relation, we can't know whether we ever do succeed in thinking about the world. (/about what our words mean, etc.)" (Fodor 2007: 6). Fodor thinks that one of the motivations for an internalist semantics is this epistemological question, which would then be answered by internalists by holding that representation is constituted by relations among our thoughts only. And since "we can know about such relations [among our thoughts] (by introspection for example) we likewise can know for sure such putatively analytic truths as that bachelors are unmarried, that cats are animals, and so forth" (Fodor 2007: 6). In other words, Fodor argues that internalist semantics is in effect a proposal "to avoid skepticism about knowledge by adopting a sort of Idealism about meaning: all our ideas are ideas about ideas" (Fodor 2007: 6).

Fodor's misconstrual of Chomsky's internalism is revealing in several respects, and so it is worth exploring in detail. It shows the intuitive pull that externalist theories of meaning have as well as the theoretical motivations driving externalism (see also the discussion in the next chapter and Slezak 2002; 2004; 2018). It is especially revealing in Fodor's case because he was one of the first to present (along with Jerrold Katz in 1963) a version of semantics that was compatible with generative syntax. Even though Fodor admits that he's "not at all sure that this is Chomsky's view", he gives a long list of reasons for why "succumbing to representational Idealism strikes me as a strategy that is to be avoided at all costs" (Fodor 2007: 6) (Fodor at times refers to Chomsky's internalist semantics as semantic Idealism or representational Idealism). Let us look at some of those reasons, addressing each in turn below. Fodor (2007: 7-8) claims the following:

> (i) That it "is wildly implausible that we don't, at least some of the time, think about the world. Semantic Idealism seems to deny this and hence to be false on the face of it." That is, the claim is that internalist semantics "rejects the notion of mind-world correspondence".
>
> (ii) The view of meaning Fodor supposes Chomsky endorses apparently requires that there be a great deal of analytic propositions and thus "avoids skepticism about whether bachelors are unmarried; we really can know that

they are; in fact, anyone who has the concept BACHELOR must know that they are". But Fodor counters that "it's very unclear how this is supposed to work for knowledge of 'contingent' propositions (for example the case of one's perceptually grounded true belief that the cat is on the mat.) In such cases, our knowledges simply can't come from our grasp of relations among ideas: It's not part of the idea CAT that this one (the one I'm, just now looking at) is on a mat; and it's not part of the idea MAT that this one now has a cat that's on it." If we suppose that empirical knowledge is a mind-world relation, then "[s]emantic Idealism avoids skepticism about 'conceptual truths' only at the cost of making a total mystery of empirical truth". That is, the claim is that there is something wrong with a semantics that implies that "our concepts are constrained by their relations to one another but not by their relations to the world".

(iii) Fodor also claims that "semantic Idealism can't account for the fact that, at least some times, we are able to make rational choices among conflicting beliefs; in particular, among conflicting scientific theories". That is, it seems to follow that "theories can't be rationally compared because what their terms in a theory mean is determine[d] internal to the theory. If I think dogs have tails and you think they don't, then we must 'mean something different' by 'dog' so there's no way of settling what appears to be the disagreement between us."

In summary, Fodor's belief is that if we take internalist semantics to claim that semantic relations hold only among ideas, then it follows that we can only think about mind-dependent things. But, he says, "it's simply untrue that whatever we can think about is mind dependent", for "we can think about The Grand Canyon, which surely was around before there were any minds and presumably will continue to be when all the minds are gone. The world (consider[ed] as the potential object of indefinitely many thoughts) is prior to the mind. A fortiori, the objects of thought can't themselves all be mental." Fodor concludes that it is "an infallible sign of bad semantics that it leads to bad metaphysics" (Fodor 2007: 8).

Note that Fodor (like Millikan and others) is here importing epistemological and metaphysical worries into the debate about the fecundity of a semantic theory. It is not at all clear that epistemological worries about the nature of knowledge and skepticism thereof are relevant to semantic theories. Fodor (2007: 6) takes such relevance to be "more or less truistic" but herein lies the problem. That the externalist conception of semantics is taken to be a truism is perhaps too strong, but the underlying intuition is clear and goes back to the early days

of externalist critiques of internalist semantics. In the oft-quoted words of Lewis (1970: 18), "[s]emantics with no treatment of truth conditions is not semantics." Lewis was very critical of the internalist semantics of the 1960s (Katz & Fodor 1963; Katz & Postal 1964), a version of which is still held by biolinguists today, because he felt that it dealt with "nothing but symbols". That is, assigning to sentences particular symbols or semantic markers that are supposed to explain their semantic interpretation (what Lewis called Semantic Markerese) is supposed to be invalid because it "amounts merely to a translation algorithm from the object language to the auxiliary language Markerese" (Lewis 1970: 18). The translation of sentences into symbolic (but, crucially, non-linguistic) expressions is seen by Lewis to be "at best a substitute for real semantics, relying either on our tacit competence (at some future date) as speakers of Markerese or on our ability to do real semantics at least for the one language Markerese" (Lewis 1970: 18). Lewis erroneously takes Markerese to be akin to a natural language and jokes that, given certain qualifications, we might as well translate the sentences into Latin. In other words, Markerese semantics cannot deal "with the relations between symbols and the world of non-symbols – that is, with genuinely semantic relations" (Lewis 1970: 19).

But we do not (and cannot) speak Markerese as a natural language (any more than we can see the visual mental representations our brain uses to process edges in a visual scene, or speak physics) and the fact that Lewis thought this was implied by internalist semantics is again revealing. Internalist semantics proposes a mental structure (an I-language) in virtue of which language production and comprehension are made possible. This structure is not available to introspection, nor is it anything like a natural language that we can understand qua speakers, and so the claim that internalist semantic theories rely on the tacit competence of us as speakers of Markerese is a misunderstanding of the internalist project. The claim that internalist semantics is "nothing but" symbol manipulation has remained a consistent misconstrual of internalism since Lewis, the main reason being that semantics is taken to essentially make epistemological and metaphysical claims. Fodor and Lewis are far from alone in this. Ludlow (2003), for example, argues for a certain kind of language-world isomorphism. As he puts it, "[t]he idea is that the linguistic representations will indeed underwrite our metaphysical intuitions, but that because of this we can expect our metaphysical intuitions to shed some light on the nature of I-language" (Ludlow 2003: 154). In other words, a "plausible hypothesis [...] is that any grasp we have on metaphysics is by virtue of our having the linguistic representations that we do" (Ludlow 2003: 155). Ludlow tries to have it both ways. He assumes with Fodor that there ex-

ists a language-world relation such that studying the structure of language can shed light on the structure of the world, but he wants to do this within an internalist semantic picture. That is, Ludlow argues that our metaphysical intuitions can be underwritten by the structure of I-language. As he puts it, we "do have substantial independent knowledge of the language faculty, and we can use that knowledge to gain insight into the nature of reality" (Ludlow 2003: 154).

It appears that Ludlow has the same motivation as Fodor in trying to preserve a certain conception of mind-world relations so that it could underwrite truth and realism. It is thus no coincidence that both Fodor and Ludlow (albeit in different ways) propose an externalist referential semantics that could be compatible with I-languages. However, the language-world relation that is posited by externalists is problematic at best (see chapter 3 for discussion). We can understand language in the language-world relation as either the surface structure of sentences and their pronunciations or as the underlying structure. Both are unhelpful in regard to metaphysics or epistemology. If we take the relation to be between the surface forms of language (say the words *sheep* or *river*) and the world then we can of course agree that those words are used to refer to sheep or rivers and thus agree that there exist sheep and rivers. That is, Ludlow's and the externalists' claim is that the word *sheep* refers to sheep in the world, and that we can use this fact to underwrite our metaphysical theory about what can exist in the world. That is, the claim is that we can successfully refer to sheep in the world because sheep exist in the world. If we can refer to them then they exist, and so we can infer the latter from the former. However, as Chomsky quips, we "can accept all this at the level at which we abandon curiosity about language and mind, about human action and its roots and properties" (Chomsky 2003a: 290). If we take the relation to be between the surface forms of language and the world, are we making a metaphysical claim about the constituents of the world? It is difficult to see how this position can be substantiated, at least if we construe the metaphysical claim to be part of natural science. Alternatively, if we take the relation to be between the underlying structures of language (say the logical forms of linguistic expressions) and the world then we also run into difficulties. If the underlying structures in the relation are couched in terms of logical form or in terms of syntax, then as J. Collins (2007) shows, there is both too much and too little structure in syntax to serve the purposes that philosophers want them to serve.

The language-world relation, however, is problematic in a deeper sense. As Chomsky (1995a; 2000; 2003a; 2003b) argues, the analogous argument in regard to sound is absurd, so why is it taken seriously in regard to meaning. That is, sup-

pose (following the thought experiment in Chomsky 2003b: 270 ff.) that we talk of a denotational or referential phonology as parallel to a denotational or referential semantics. So that instead of $Phon$ in the expression $Exp =< Phon, Sem >$, which explains the sound properties of the expression in an internalist fashion, the expression lacks $Phon$ and is instead said to *P-denote* some object that is external to the speaker, call it a phonetic value (PV). One can even suppose that some computation on the phonetic value yields the sound component of expressions. The phonetic value could be said to be a construction from physical sound waves and the proposal could be elaborated by taking into account the social context of the speakers. Given this denotational account of phonology, we could then offer an account of communication, translation, and perhaps even language acquisition. So, for example, we could say that "Peter is able to communicate with Tom because the same PV is denoted by their expressions in the language they share (but only partially know)" (Chomsky 2003b: 271). However, as should be clear, a denotational phonology gets us nowhere and in fact "leaves all problems where they were, adding a host of new ones", for we "understand nothing more than before about the relation of [the expression] E to its external manifestations" and thus such an "account of communication and other processes is worthless" (Chomsky 2003b: 271). The introduction of phonetic values is completely unhelpful in regard to explaining how humans interpret the sounds of linguistic expressions and there are of course no serious theories of denotational phonology.

The language-world relation in regard to sound is never postulated, whereas the language-world relation in regard to meaning is commonplace. These two postulated relations are of course not identical, but they are similar enough to present a puzzle as to why denotational phonology is rightly dismissed as absurd whereas denotational semantics is taken seriously. One could perhaps make the counter argument that sound doesn't "mean" anything, but this begs the question. Indeed, semantics doesn't "mean" anything in the everyday sense either. Semantics explains interpretation but does not provide interpretation per se. More on this in the next chapter.

Fodor's misconstrual of semantic internalism claims that Chomsky's language faculty hypothesis is an epistemological proposal. As J. Collins (2004) has argued at length, Fodor reads into Chomsky's work (and into internalist semantics in general) epistemological proposals about what speakers know about their language. This type of knowledge is said by Fodor to be propositional knowledge as understood in philosophy, thus bringing in issues of truth, belief, and justification. It follows, then, contrary to explicit pronouncements by internalists and

biolinguists, that the hypothesis of the language faculty is not a proposal about a specific mental structure in the mind. But as we saw above, the biolinguistic claim is just that. Indeed, whether it is true or not, internalist semantics does not make epistemological claims at all. Jackendoff's remarks in regard to the tension that exists between fundamental questions in a theory of mind are relevant here. There is what he calls a philosophical approach that "grows out of questions of epistemology", and there is the psychological approach that "grows out of issues in perception" and questions of how the brain functions. Jackendoff argues that this tension has "the flavor of a paradigm split in the sense of Kuhn" (Jackendoff 1991: 411-412).

Moreover, Jackendoff points to the fact that Fodor "embodies in a single person *both* sides of the paradigm split" (Jackendoff 1991: 412, emphasis in original). On the one hand, Jackendoff notes, Fodor wants to be a psychologist by insisting that an organism's behaviour is determined by its computational mental representations. On the other hand, Fodor is "holding on very hard to his roots as a philosopher in his use of the terms *true* and *false* with respect to mental representations" (Jackendoff 1991: 412). Jackendoff argues that since the philosophical approach "leads to uncomfortable metaphysical problems", one should abandon it in favour of the psychological approach that "permits – at least in principle – a revealing account of the phenomena" (Jackendoff 1991: 416). The conclusions Jackendoff draws are supposed to apply to the philosophical and psychological approaches to perception, but as we'll see in the next chapter, the same "paradigm split" exists in the debate between semantic internalism and externalism where externalists insist upon the relevance of truth and reference to semantic theories whereas internalists reject the relevance of such notions to a scientific semantic theory.

***

The next chapter will discuss the externalist approach to semantics, but before moving on let me briefly outline where the argument is going. After detailing the externalist approach I will argue in the final chapter that internalism is a promising solution to the problem of constructing a scientific theory of semantics, a solution that does not include the various additives that externalists demand be included in a semantic theory. Externalism will thus be discussed from the point of view of explanatory scientific theories. I will ask what externalism is supposed to explain, and whether such an explanation is or can be a scientific one. In contrast to leading externalists, I argue that whatever merits externalism may possess, it is unable to provide a fruitful explanatory framework for a scientific

theory of meaning. I argue that an externalist explanation of meaning is concerned with *ascription* and *description* of meaning rather than the *mechanisms* of meaning. That is, externalism is not concerned with the mental mechanisms in virtue of which humans produce and comprehend meaning. In chapter 4, I argue that a fruitful scientific explanation is one that aims to uncover the underlying mechanisms in virtue of which the observable phenomena are made possible, and that a scientific semantics should be doing just that. Therefore, externalist explanations are not part of the psychological explanation of the mechanisms in virtue of which meaning is made possible. Rather, externalist explanations are an interpretive and hermeneutic explanatory project.

In his discussion of externalism, Chomsky remarks that it faces a choice: if it is conceived as part of ethnoscience, "it is making the factual claim that people (in our culture, or universally) attribute thoughts, beliefs, etc., which they individuate by reference to environment or social context, and then faces the task of clarifying and defending that empirical thesis". If, on the other hand, externalism is conceived as part of psychology, then "it is making the claim that among the entities of the world, alongside of complex molecules and (maybe) I-languages, are mental states individuated by environment and social context, and it will again have to explain what these entities are, show how they function, and provide empirical confirmation for its conclusions about these matters" (Chomsky 2003: 269-270). Chomsky is skeptical about the prospects for *both* of these construals of externalism and stresses that whichever way we understand externalism, the normal criteria for scientific theory evaluation should be satisfied. Chomsky's skepticism is warranted in regard to externalism understood as part of psychology, but there is still value to the way in which externalism approaches semantics as hermeneutics in that (if sharpened and explicitly understood in this way) it can help in shedding light on the way in which language users ascribe meanings to words or phrases in particular contexts. But this should not be confused with the scientific task of unearthing the mechanisms in virtue of which language production and comprehension are made possible.

This is where the argument is headed. But first let us explore the externalist approach to semantics.

3 Externalism

As noted above, the externalist approach is currently somewhat of an orthodoxy in the philosophy of mind and in the philosophy of language. Moreover, this approach is not limited to philosophy, for much work in linguistics (for example, in formal semantics and formal pragmatics) takes the truth conditional approach to meaning. There are a number of positions that go by the name of externalism, but I will focus here on Twin Earth externalism. That is, the accounts that rest on a Twin-Earth-style thought experiment, the most famous version of which is Putnam (1975). This is the standard understanding of externalism. Indeed, as Farkas (2003) notes, some philosophers use Twin Earth thought experiments as part of the very definition of what externalism is (see also McLaughlin & Tye 1998). Moreover, standard linguistics textbooks in formal semantics also approvingly discuss these thought experiments (see, for example, Portner 2005: 7ff.). The Twin Earth thought experiment claims to show that two subjects can have identical internal mental states but that the content of these states can be different due to particular variations in the environment. In other words, the claim is that the content of mental states can vary with variations in the environmental or socio-environmental conditions of two subjects while their respective internal mental states remain identical. The conclusion of externalist semantics, then, is that meanings are individuated by reference to environmental features or social contexts, and that therefore in order for a person's utterance to have a particular meaning it must be related to the environment in the right way.

Note that even though much of the discussion in the externalist literature is couched in terms of content, it is clear that the conclusions in regard to content are meant to apply to linguistic meaning as well. Burge, for example, is explicit about this: "The arguments for anti-individualistic individuation of mental kinds can be extended in relatively obvious ways to show that much of semantics is not purely individualistic [i.e., it is externalist]" (Burge 1989: 279). This can be generalised, so that externalist arguments couched in terms of mental content can be applied to linguistic meaning, as indeed they often are. Content is perhaps a broader term that can apply to non-linguistic mental states such as visual perceptions, or perhaps content is entirely distinct from linguistic meaning. But

whatever content turns out to be, externalists argue that its nature will have direct bearing on the nature of meaning. In addition, the fact that Burge says that his philosophy of language has a direct bearing on much of semantics shows that he construes his externalist theoretical aims to overlap with those of linguistics and semantics in particular.

There are two traditional assumptions that lurk in the background of the debate in the externalist literature. These are (i) the claim that the meanings of words are fixed by the psychological states of those who use them, and (ii) the claim that the meanings of words determine their extension or reference. Externalism is supposed to entail that (i) and (ii) are incompatible. Thus, the Twin Earth thought experiments purport to show that since meaning determines reference – and so terms with the same meaning will have the same reference – the psychological states of the twins cannot determine their meaning because the reference of their utterance of *water* is different. In other words, if one agrees that a difference in meaning implies a difference in reference, then one cannot hold that meaning is determined by psychological states; this is because, *ex hypothesi*, the psychological states of the twins are identical but the reference is different. Putnam argued that it is possible for two speakers to be in exactly the same psychological state, even though the extension of a term in one speaker is different to the extension of the *same term* in the second speaker. If this is correct, he argued, one must give up one of the traditional assumptions (i.e., give up either (i) or (ii)).

I want to argue, however, that it is possible to hold both (i) and (ii), but not for the reasons one might expect. That is, (i) is a psychological explanatory project in regard to meaning and mind (as illustrated by internalist semantics in the previous chapter), and (ii) is a hermeneutic explanatory project. As a result, there is no tension between holding that a person's psychological states fix their meanings and holding that meaning determines reference. In other words, (i) is part of the explanation of the psychological mechanisms in virtue of which meaning is made possible, whereas (ii) plays an interpretive function for the theorist by linking internal psychological states and the world.

As noted above, Jackendoff (1991) makes a similar point when he compares the tension that exists in regard to fundamental questions for a theory of mind. Barry Smith (1992) draws a parallel distinction between what he calls the interpretive, descriptive, and explanatory stances to a theory of meaning. At one end of the spectrum "we have the language-dissolving view of interpretation, which at the limit slides into hermeneutics and literary theory where there are no standards of correctness at all. ('Anyone can do the philosophy of language')." At the other

end of the spectrum there is "an extreme philosophy of language" which "gives way to empirical research. ('Best left to the scientists')." Smith advocates a theory of meaning that "must come somewhere in between", and he takes Chomsky's framework as setting "a constraint on any satisfactory solution [to the problem of meaning]" (Smith 1992: 138-139). Smith's position is more conciliatory and closer to the position I argue for here, for it leaves some room for the philosophical and hermeneutic approach to semantics. In contrast, Jackendoff's claim that the philosophical and psychological approaches to a theory of mind are disparate in the sense of Kuhn is too strong, for it neglects the interest and value that lies in the hermeneutic explanatory project. Nevertheless, there *is* a systematic difference between the questions that externalists attempt to answer and the questions that internalists attempt to answer. Thus, if the externalist mode of explanation is a hermeneutic one, then despite the claims of many externalists it will be unable to provide an explanatory framework for a science of semantics.

If we assume that meaning is externalist as defined by Lewis and others and that we cannot individuate the meanings of utterances without reference to mind-external factors, what follows in regard to the science of meaning? I agree that one can only discern what a person's utterance refers to by consulting the external environment, and that the referents or extensions or denotations of thoughts cannot be exclusively determined by mind-internal matters – the question, say, of whether a referent is a sheep or a bush that looks like a sheep cannot be determined without consulting the external environment. But these are questions of meaning ascription or individuation. I argue in what follows that scientific explanations do not attempt to answer such questions, at least not in the way that is claimed in the externalist literature. That is, I argue that the externalist claim that there exists a deep-rooted link between inner psychological states and their extension is not problematic – indeed, it is essential and potentially fruitful – for the hermeneutic explanatory project of thought contents. However, in regard to the psychological explanatory project of the latter, questions of attribution or individuation do not play a key role.

3.1 The subject matter of externalism

Leading externalists explicitly and repeatedly state that their theories are part of the scientific project. What are we to make of such claims? Do criteria for ascription and description of meaning belong in a scientific theory? We should question the assumption that externalism is directly relevant to a scientific theory of meaning that attempts to unearth the mechanisms in virtue of which meaning

works in the mind. Putnam argues that a better philosophy and a better science of language is externalist. But what does Putnam understand to be the details of his science of language? One of his other thought experiments supposes that "[i]f Twin Earth organisms have a silicon chemistry, for example, then their 'tigers' aren't really tigers, even if they look like tigers, although the linguistic habits of the lay Twin Earth speaker exactly correspond to those of Earth speakers" (Putnam 1975: 167). This stems from Putnam's (and the externalists') belief that "*extension is tied to the notion of truth*" and that the "extension of a term is just what the term is true of" (Putnam 1975: 154, emphasis in original). Thus, according to this reasoning, we can only determine the meaning of a person's utterance – whether the meaning of the person's utterance *really is tiger* – by consulting the external environment and checking whether the utterance is true of that environment. A person on Earth and his doppelgänger on Twin Earth can have the same internal psychological states or concepts (tigers are striped, they're quadrupeds, they have paws, they have whiskers, etc.) and yet mean different things when they utter *tiger* because the tigers in their environment are different – one is carbon-based and the other is silicon-based. Putnam concludes that the same utterance spoken by him and by his doppelgänger can have different meanings, "but this will not be an assertion about our psychological states" (Putnam 1975: 165).

This is a curious statement. If we take a science of language to encompass externalist relations in the way that Putnam urges, and if we agree that language processing is somehow instantiated in the brain, what are we to make of the claim that externalist theories of meaning do not make assertions about psychological states? Fodor similarly argues that "[i]t is, to put the point starkly, the heart of externalism that *semantics isn't part of psychology*. The content of your thoughts (/utterances), unlike for example, the syntax of your thoughts (/utterances), does not supervene on your mental processes" (Fodor 1994: 38, emphasis in original). It is far from unusual to find such statements claiming that semantics (or philosophy of language, for that matter) is not about underlying psychological states. Soames (1984), for example, argues that "linguistic theories are conceptually distinct and empirically divergent from psychological theories of language acquisition and linguistic competence" (Soames 1984: 155). Soames denies that linguistic theories are theories of "[c]omplex, unconscious, computational states and processes [that] underlie language acquisition and mastery" (Soames 1984: 155). There is much more to say on the matter, but for the moment let us note that such claims are in direct opposition to the way in which internalist semantics and biolinguistics practice their research programs. Internalist semantics, as we saw above, explicitly takes its theories to be about the

underlying computational mechanisms in virtue of which language production and comprehension are made possible. Soames, however, claims that generative linguistics and psychology "are concerned with different domains, make different claims, and are established by different means" and thus "linguistics does not yield computational and representational theories in cognitive psychology" (Soames 1984: 157). Soames insists on separating the conceptual and empirical foundations of generative linguistics from those of psychology. But contrary to his claims, his construal of linguistics does not apply to generative linguistics and in fact he has misunderstood its aims and methodology.

At the risk of gratuitously discussing yet another misconstrual of internalism, I would like to briefly discuss Soames's misconstrual. It is important to understand how widespread this confusion is. People like Fodor, Soames, and others discussed here, are leading figures who know the literature of generative linguistics quite well. It is thus of great interest to unearth the source of their misconstrual of internalism, for it both sheds light on the underlying (and I would argue mostly implicit) theoretical assumptions of hermeneutic projects such as externalism and helps to clarify the research program of internalism. The claim of generative linguistics (and, later, biolinguistics) to be part of psychology (and ultimately biology) has been debated for decades with little consensus. As has the internalist semantics claim that meanings are internal psychological phenomena. Unpacking the reason for this seemingly recalcitrant debate shows the Kuhnian split between the two camps is due to the often unnoticed fact that one is a hermeneutic project and the other is a scientific project.

Soames proposes three "Leading Questions" that linguistics should attempt to answer; they are:

> (i) In what ways are English and Italian/the Romance languages/all Indo-European languages/all natural languages alike and in what ways do they differ?
>
> (ii) What (if anything) distinguishes natural languages from some set of artificial languages (such as finite state languages) or from animal communication systems?
>
> (iii) In what ways has a particular language (a description (or list) of one or more natural languages) changed and in what ways has it remained the same? (Soames 1984: 158)

Later in the article he remarks that "by *linguistics* I mean the discipline defined by the Leading Questions discussed earlier and practiced today by generative

grammarians" (Soames 1984: 178, fn. 24, emphasis in original). However, whatever virtue and interest the answers to the Leading Questions may yield, it is not at the core of what generative linguists do. The separation that Soames wants to draw between linguistics and psychology is that between a linguistics that "aims at providing theories of natural languages" and that of a "cognitive psychology [that] aims at providing theories of natural language users" (Soames 1984: 157). But this distinction is not applicable to generative linguistics, for generative linguistics explicitly deals with providing a theory of what is in the head of language users in virtue of which their language production and comprehension are made possible.

Soames wishes, then, to distinguish between "non-mentalistic" linguistics and "non-linguistic" psychology; he remarks that "[s]ince the job of a linguistic theory is to specify the similarities and differences among (possible) languages, such a theory must be sensitive to truth conditions (or elements that determine them)" (Soames 1984: 163). The appeal to truth conditions (a hallmark of externalism) is the real reason for Soames's insistence that linguistics is not psychology: "languages may differ not only with respect to syntactic and phonological properties, but also with respect to semantic properties involving truth conditions", and thus a linguistic theory "that failed to account for truth conditions would miss these differences" (Soames 1984: 162). In other words, if "linguists' grammars were simply psychological theories, then claims about truth conditions would themselves be psychological", but since "these claims are not (purely) psychological in nature, it follows that grammars are not wholly psychological in nature and that linguistics is not merely a branch of psychology" (Soames 1984: 163). The argument is that since semantics (as construed by Soames and other externalists) makes use of "extra-psychological notions" (that is, "to give the truth conditions of sentences is to specify the non-linguistic conditions that would make them true"), "it follows that linguistic semantics is *conceptually distinct* from psychological models of semantic competence" (Soames 1984: 163, emphasis in original). More recently, Soames (2009) has been somewhat more charitable to psychological models of semantic competence, but he still argues in favour of a "nonpsychologistic perspective" according to which "[s]entences, and other expressions, have grammatical structures and representational contents that can be studied in abstraction from questions about how they initially came to have those structures and contents, what psychological states and processes are responsible for their retaining them, or how speakers come to know whatever they do know about them" (Soames 2009: 1-2).

One is of course free to choose the perspective within which to study meaning

that they feel will yield the most fecund explanation. One can agree with Soames (2009: 183) that the "job of semantics is to specify the principles by which sentences represent the world". That is, since it "is impossible to represent the world as being a certain way without implicitly imposing conditions that must be satisfied if the world is to conform to the representation", "whatever else a semantic theory must do, it must at least characterize truth conditions." Soames sees this construal of the job of semantics to be part of the "basics" of the field, implying that if one were not to do semantics in this way then one would not be doing semantics. But this at best reduces to the a priori assertion that "real semantics" (Lewis 1970) is concerned with truth conditions. It of course follows from this assertion that linguistics isn't part of psychology, for in order to specify truth conditions one must make use of "extra-psychological notions". But this assertion begs the question: the issue in generative linguistics (which Soames claims is characterised by the Leading Questions) is what kind of semantic theory fits within a scientific framework of the study of the mind. One cannot legislate what "real semantics" is and then proceed to argue that therefore semantics is not part of psychology. Moreover, by asserting that a linguistic theory must be sensitive to truth conditions, Soames has indeed distinguished linguistics from psychology, but this characterisation of linguistics, *pace* Soames (1984: 178, fn. 24), is not that of generative linguistics.

It could be objected here that truth-theoretic semanticists do not see their semantic theory as contributing to a scientific theory of the mind that seeks to unearth the underlying mechanisms in virtue of which meaning is made possible, and thus it would be moot to compare the two. But there are many leading externalists and formal semanticists who explicitly claim that their project is a scientific one with similar aims to those of generative linguistics. It is thus both warranted and illuminating to look at the externalist research program from the perspective of scientific explanatory strategies and to ask whether it is a promising avenue in regard to constructing an explanatory scientific theory.

That said, however, there are plenty of philosophers of language who wish to separate an externalist semantics from a semantics construed psychologically. Dummett, for example, argues that philosophy "is not concerned with what *enables* us to speak as we do, but what it is for utterances to have the meanings that they have" (Dummett 1994: 187-188, emphasis in original). That is,

> [...] a theory of meaning is required to make the workings of language open to our view. To know a language is to be able to employ a language; hence, once we have an explicit account of that in which the knowledge of a language consists, we thereby have an account of the workings of that lan-

> guage; and nothing short of that can give us what we are after. (Dummett 1993: 4)

Dummett insists that "once we can say what it is for someone to know a language, in the sense of knowing the meanings of all expressions of the language, then we have essentially solved every problem that can arise concerning meaning" (Dummett 1993: 4). In other words, "[a]ny theory of meaning which was not, or did not immediately yield, a theory of understanding, would not satisfy the purpose for which, philosophically, we require a theory of meaning" (Dummett 1993: 4). The claim here is that semantics is not part of psychology, nor is it part of a linguistics construed as part of psychology (as biolinguistics and internalist semantics take themselves to be). Indeed, Dummett (1994: 187) complains that Chomsky's theory of meaning "is really a theory of something very complicated that goes on in the brain" and "that is a completely unphilosophical way of looking at the matter."

So what is the subject matter of an externalist semantics if not psychology or the mental states of language users? Externalist semanticists see their task as providing an account of the relation between linguistic expressions and things in the world. This relation can be fleshed out in several ways, but what underlies them all is the concern with the conditions or rules of ascription: when is one justified in ascribing a particular content to a particular utterance, and what is the correct content that should be ascribed. Given such aims, then, what does the research project take itself to be explaining? Once this question is answered one can address the further question of whether externalist explanations are scientific as some leading externalists claim to be the case. In other words, does externalism employ a form of understanding that is appropriate to a cognitive psychological explanation? Or, in contrast, is the way in which externalism understands the mind orthogonal to that of cognitive psychology? This highlights an important distinction, one between different research projects that are often conflated. That is, the distinction is that between, on the one hand, the notion of understanding that refers to interpreting a sentence as a speaker of the language (*verstehen*) and, on the other hand, the notion of understanding that refers to explaining as in science (*erklären*). As Slezak (2004) shows, the question of semantics or the content of mental representations is often confused between whether representations are *intelligible* to the *theorist* and whether they are explainable by the *theory* (see also Slezak 1990; 2018).

Peacocke (1993), which details how explanation by externalist states is supposed to work, is a case in point of such a conflation. It is, he says, "partially constitutive" of the identity of any externalist state that "it can explain, or be ex-

plained by, relational properties of external objects or events" (Peacocke 1993: 206). Relational properties, such as the "highly relational property of saying something which has a certain meaning", are argued to be explained psychologically (Peacocke 1993: 204). Psychology can be done in an externalist fashion on Peacocke's conception (and thus aid in the study of semantics), for a psychological explanation of an event explains a particular set of its relational properties. Peackocke thus explicitly merges the externalist approach with the psychological approach, taking the externalist framework as a means to achieve a psychological explanation of the mind.

But what is the nature of these relational properties? These are properties that exist only in virtue of relations to other objects; the properties depend on the relations and there is no sense in which they could exist independently of them. Nuccetelli (2003: 3) gives the example of the property of being west of Central Park, where "whether one has it depends on how one is geographically situated with respect to Central Park." She then argues for the externalist claim that such relations are analogous to thoughts having a certain content. That is, "[g]iven externalism, having either the belief that water is wet or other propositional attitudes with certain contents would be in some sense analogous to being west of Central Park, simply because the content *type* of some such attitudes would supervene on the relations of those who entertain them with their physical and/or social environments" (Nuccetelli 2003: 3, emphasis in original). This analogy is of course only illustrative, but it outlines the foundations of externalism in that it is claimed that even though thoughts are inside a person's head, their content supervenes on external factors in the environment of the person who has those thoughts. Thus, as Ben-Menahem (2005) notes in regard to one of Putnam's examples, "to speak of coffee tables it does not suffice for us merely to have the concept of a coffee table, but we must be in contact with *actual* coffee tables" (Ben-Menahem 2005: 10, emphasis in original).

Another analogy is that of sunburn: Davidson (1987) remarks (in the context of a discussion of meanings and their relations to objects outside of the head) that "[m]y sunburned skin may be indistinguishable from someone else's skin that achieved its burn by other means (our skins may be identical [...]); yet one of us is really sunburned and the other not" (Davidson 1987: 451-452). That is, in order for one to have the property of *being sunburned* (to *really* be sunburned), one's skin must have had the proper relationship to the outside environment, namely, to the sun. Such analogies illustrate the kind of reasoning that lies behind the externalists' claim that a particular thought or utterance cannot have a particular meaning unless it has had the proper relation to the outside envi-

ronment. In other words, according to externalism, content or meaning (or their individuation), in this case the meaning of *sunburn*, is in some way *essentially* tied to the environment (see Egan 1999 for discussion). Thus, as McGinn (1989: 9-10) remarks, "mind and world are not, according to externalism, metaphysically independent categories, sliding smoothly past each other. To regard them as so is to commit oneself to an 'untenable dualism', to marking a metaphysical boundary that does not exist."

Talk of such relations is systematic and their justification and validity is rarely explicitly defended. Burge (2003: 466), for example, remarks that "[t]aking account of language–world relations is part of the way semantics is actually practiced" and thus there is "no reason to think that there is anything scientifically wrong or fruitless in studying language–world relations, or with taking them to be part of the formal structures elaborated in semantical theory." This has led to the situation in which relational properties are read into research programs such as internalism that explicitly deny them. Higginbotham (1991: 556), for example, argues that Chomsky's notion of competence involves "an epistemic relation between a person and the principles that determine her language." He argues against a view he calls representationalism that generative linguists supposedly hold according to which "having a language just amounts to having a system of mental representations" and "that one stands in epistemic relations to the principles of one's own language" (Higginbotham 1991: 557). Higginbotham complains that according to this doctrine that he pings on generative linguistics "there is only a pedantic distinction between representation and represented" (Higginbotham 1991: 557). That is, "it conflates questions about what is apprehended, language, with questions about the means of apprehension." Analogously, Higginbotham remarks, "[n]o one confuses the mental representation of a tree with a tree" and so "[w]hy should it be so common among linguists to write as if intending to confuse the mental representation of a sentence with a sentence?" (Higginbotham 1991: 555). In contrast to the representationalism that Higginbotham claims conflates the two, he wishes to retain the "philosophical distinction between language and its representation" (Higginbotham 1991: 558). The latter distinction encapsulates the E-language conception of language.

There are two questions that, according to Higginbotham, generative linguistics investigates: (i) what is the nature of language? and (ii) what is the relation between speakers of languages and the languages that they speak? These questions arise, he says, "if we are interested in a systematic metaphysical view of the conception of linguistics as a chapter of the cognitive sciences": "in short, they are philosophical problems." Referring to generative linguistics, Higginbotham

claims that a "popular but confused answer" to the aforementioned questions "is that languages are systems of mental representation, and that the growth of competence is the growth of such systems under appropriate environmental contingencies" (Higginbotham 1991: 555). This answer is confused, according to Higginbotham, because of the conflation between the mental representation of language and language itself that rejects the philosophical distinction between language and its representation.

Despite persistent efforts by Chomsky (1975; 1995; 2000; 2016) and others to clarify their position, misunderstandings of this sort persist. Devitt (2003; 2006; 2006a) has for years maintained a version of the *linguistics is not psychology* argument, arguing forcefully for an epistemological reading of generative linguistics. He claims that "there is a natural interpretation which takes Chomsky pretty much at his word" in which the answer to the question of what constitutes knowledge of language "urges that competent speakers of a language have propositional knowledge of its rules" and that this "knowledge underlies the speakers' intuitive judgements about the syntax of expressions" (Devitt 2003: 107-108). This reading of Chomsky, which Devitt curiously takes to be the natural interpretation, leads to the conclusion that "there is something theoretically interesting for a grammar to be true about other than the internal reality of speakers" and thus the "grammar might be true about a symbolic system, a *linguistic* reality." In other words, Devitt claims that "we can take the grammar realistically without taking it to be true of *psychological* reality" and that "given the weight of evidence adduced for a grammar, it is plausible that it *is* (more or less) true of linguistic reality" (Devitt 2003: 131, emphasis in original).

The claim that a grammar can be "true of" something other than the internal structure of the mind originates with Stich (1972), but since then there has been no clear elucidation of the implication that, as Devitt puts it, evidence from psychology, psycholinguistics, and the like, "bears on the grammar even without the assumption [that the grammar is psychologically real]" (Devitt 2003: 128, fn. 28). What does it mean to say that the grammar is *linguistically* real but is not in the head? Devitt claims that the grammar is true of a symbolic system, "a linguistic reality made up of the spoken, written, etc., symbols that speakers produce" (Devitt 2006a: 483, fn. 5); but what is the nature of this system? Devitt takes a grammar to be about a non-psychological realm of expressions, a realm where physical entities form representational systems that are somehow distinct from the creatures that use them. Moreover, he claims that "the truth of a grammar for a language leaves the question of the psychological reality of the language open" (Devitt 2003: 136). He argues that we should begin with studying

the grammar's linguistic reality and that only "in the end we will need to study the psychological [reality] in order to explain the linguistic [reality]. But in the beginning we do not" (Devitt 2003: 135). Such an approach is perhaps correct in regard to well-formed formulae (wffs) in formal logic, where we can study, to use Devitt's terminology, the structure rules (the rules governing outputs of a person's competence) without consideration, at least initially, of the processing rules (the rules governing the psychological production of the aforementioned outputs). Devitt correctly argues that these two rules are very different in regard to formal logic (the structure rules of formal logic are very different to the processing rules of the mind or a computer that govern these outputs), but *natural language* does not fit this mould.

Formal languages are invented, natural languages are not. We stipulate the properties of formal languages to suit particular purposes in logic and mathematics, but in natural languages we do not stipulate but rather *discover* their properties. Chomsky (2002) remarks that there is no right answer to questions such as *What are the true rules of formation for well-formed formulas of arithmetic?* or *What are the axioms of arithmetic?* because, at least in principle, any set of axioms can generate the theorems in question. Particular axioms are a particular way of describing the theorems, but not the only way. The same goes for computer languages: the rules that are chosen to characterise their expressions can be almost anything because they can be implemented on a wide range of distinct platforms. That is, the expressions themselves are the language, not the specific computational system that characterises them. In natural language, however, the reverse is the case. That is:

> In natural language there is something in the head, which is the computational system. The generative system is something real, as real as the liver; the utterances generated are like an epiphenomenon. This is the opposite point of view [to that of a computer language or to formal logic]. (Chomsky 2002: 110)

There is thus no analogy between formal languages and natural languages because in the former one can choose any set of axioms to generate the same theorems and it is thus these theorems that are the language. In *natural* languages, however, one cannot choose any set of axioms because there is a computational system in the head with a specific set of axioms or principles, and it is these principles that are the language. This is because expressions are generated by a computational system that is the same for all language users (the output of this generative system varies depending on various factors such as what natural language is spoken by the community of individual language users, but the

underlying system that makes this possible is universal). Thus, an explanatory scientific theory of human language cannot, to use Devitt's terminology, separate psychological reality from linguistic reality, nor can it postulate grammars that are merely "true of" the speaker but are not internally represented in the speaker. This is of course an empirical question that could turn out to be wrong, but we should not misunderstand the claim of internalist semantics, which explicitly postulates a computational procedure that is instantiated in the mind. Devitt claims that "the grammar is describing the syntactic properties of (idealized) linguistic expressions, certain sounds in the air, inscriptions on paper, and the like" but that even though these items "are produced by minds" and "presumably get many of their properties somehow from minds", "they are not themselves mental" (Devitt 2006: v). If one assumes an E-language perspective, then one can agree with Devitt that generative linguists conflate a theory of language with a theory of linguistic competence and that "a person could be competent in a language without representing it or knowing anything about it: she could be totally *ignorant* of it" (Devitt 2006: 5, emphasis in original). But internalist semantics takes the I-language perspective and so there is no conflation. Notice that this is not a matter of terminology: the biolinguistic proposal is an empirical proposal that could be right or wrong and that will stand or fall on the merits of its explanatory fecundity. We should be wary of reading into it epistemological relations that are explicitly denied.

Why do Soames, Higginbotham, Devitt, and others insist on retaining the distinction between the mental representation of the language and the language itself and thus misinterpret internalist semantics as also holding the same distinction? The answer has to do with their conception of knowledge. From the I-language perspective the distinction between language itself and its mental representation is superfluous: speakers of a particular language do not represent their language like they would represent some aspects of the external world; they just have their language (qua I-language). The mental representations proposed by generative linguistics to account for, say, a particular language, are not *about* that language; *they are that language*. This is language understood intensionally in terms of the generative procedure that produces the set of structural descriptions as opposed to the extensional understanding that sees language as the set itself that is the output of the generative procedure. Higginbotham is not satisfied with this because he wants a linguistic theory to explain not only the properties of language but also how one *knows* (qua justified true belief) one's language. He remarks that the representationalist doctrine, which he reads generative linguistics as holding, fails when it comes to semantics because '"[w]e

seem to use words with their meaning, when we have only a partial or even a mistaken conception of what that meaning is" (Higginbotham 1991: 563).

Knowledge of meaning (and language itself), says Higginbotham, "becomes social when we acknowledge others as knowing more than we about meaning, or as correcting us about meaning, and where our grounds for doing so are cognitively based." But if language is social, Higginbotham continues, then the distinctions "between mental representations and what they represent" and "between what we think the properties of our language are and what they are in fact" are significant (Higginbotham 1991: 563). In other words, if you want a linguistic theory to explain not only the mechanisms in virtue of which language production and comprehension are possible but also how it is that speakers can misrepresent or be mistaken about a particular meaning or grammatical feature, then the "language with its properties must be distinguished from what the speaker knows about it" (Higginbotham 1991: 563). This is a clear expression of externalism, which is "principally a view about the conditions for truth and reference, and invokes the same considerations whether it is the condition for the truth of a sentence, or for the truth of a belief is in question" (Farkas 2006: 328).

Notice again that the externalist position conflates two notions of understanding: the notion of understanding that refers to interpreting a sentence as a speaker of the language (*verstehen*) and, on the other hand, the notion of understanding that refers to explaining as in science (*erklären*) (see Slezak 2004; 2018). This is clear in Peacocke's discussion of the "how-questions" in science. He says that when scientists know that something has a given property, they then look to find out how is it able to have that property. Peacocke's how-questions include: "How is the human body able to avoid waste products building up in the blood?" and "How is a person able to understand a sentence he has never previously encountered?" (Peacocke 1994: 315). However, the former question is of a different type to that of the latter – at least on the usual externalist reading of the latter question. But herein lies the problem, for there is an ambiguity in the question "How is a person able to understand a sentence he has never previously encountered?": an ambiguity between understanding qua speaker and understanding qua scientist. It is all too easy to slip in and out of these two very different projects. In other words, the question "How is a person able to understand a sentence he has never previously encountered?" can be answered hermeneutically in terms of interpreting the sentence as a speaker of the language; this is the externalist approach. But the question can also be answered by taking the internalist approach and trying to unearth the psychological mechanisms in virtue of which interpretation is made possible.

Externalist relational properties and their corresponding how-questions are of a different type to the *scientific* how-questions of, say, how certain functions of the blood operate or, crucially, how a person can comprehend a novel sentence understood scientifically. Peacocke's discussion is indicative of externalism's conflation of the two types of question. He states that to have "one of the properties identified in Mendelian genetics is to have a highly relational property" and that for a person "to have a recessive gene for red hair" is "something that involves *his relations* to hair colour, to other genes (or factors), and to parents and empirically possible descendants" (Peacocke 1994: 315, emphasis mine). Peacocke assumes that these relational properties are in the same category as properties in biology in virtue of which genetic explanations are possible. He claims that Mendelian theory spells out exactly what these relational properties are. But this assumption is far from obvious, and it's unclear how a scientific explanation is at all improved by postulating that a *person* has certain relations to *their own* hair colour or to their genes. One can of course postulate such relations, but not without explaining them fully and unpacking the underlying assumption that such relations are relevant to fruitful scientific explanations – for it is this assumption that is doing all the work (as I argue in the next chapter, this assumption is problematic, for fruitful scientific explanations are mechanistic). This assumption is exemplary of the underlying conflation in the externalist literature in which externalists erroneously assume that their questions and theoretical aims are identical with the questions and theoretical aims of scientists studying similar phenomena.

3.2 Externalism as a hermeneutic explanatory project

As we have seen, in the founding article of externalism Putnam claimed that "a better philosophy and *a better science of language*" must encompass the "social dimension of cognition" and the "contribution of the environment, other people, and the world" to semantics (Putnam 1975: 193, emphasis mine). Other externalists make similar claims: Paul Horwich (1998; 2005) argues that his externalist use-based semantics is compatible with a linguistics construed as an empirical science. Moreover, he says of Davidson's externalist truth-theoretic program that it "became widely accepted, instigating several decades of 'normal science' in semantics" (Horwich 2001: 371). Davidson himself is somewhat ambivalent, but still holds that "my own approach to the description, analysis (in a rough sense), and explanation of thought, language, and action has [...] what I take to be some of the characteristics of a science" (Davidson 1995: 123). And Burge says

that he sees no reason why formal semantics, which postulates "reference, or a technical analogue, as a relation between linguistic representations and real aspects of the world, should not be an area of fruitful systematic scientific investigation" (Burge 2003: 465). Moreover, Lassiter (2008: 607) claims to have responded to "Chomsky's challenge to articulate an externalist theory of meaning that can be used in the scientific investigation of language."

In order to assess such claims, let us look at a particular externalist theory of meaning in detail: Davidson's truth-conditional semantics. Davidson is one of the most influential philosophers of the second half of the twentieth century, and his work has had a significant impact not only on philosophy, but also on linguistics and cognitive science. Lepore & Ludwig (2005: viii) remark that "Davidson's proposal to use a Tarski-style truth theory as the core of a theory of meaning for natural languages [...] sparked a revolution in philosophical semantics." Davidson argues that the best way to construct a compositional meaning theory for natural language is to construct a truth theory (based on the work of the logician Alfred Tarski) that assigns, from a finite set of axioms, truth conditions to each sentence of the language. The assignment of truth conditions to the sentences of a natural language is supposed to allow a person to be able to interpret those sentences. The notion of truth, which for Davidson is "the most obvious semantic property" and "one of the clearest and most basic concepts we have" (Davidson 2005: 2, 55), thus acquires a central place in a theory of semantics. Davidson argues that the meaning of a sentence *is* its truth conditions. This conception of semantics is so widespread that, as Lepore & Ludwig (2004: 310) point out, "any approach to the semantics of natural languages is now likely to begin by stating whether it is based on adopting or rejecting a truth-conditional approach inspired by Davidson's work."

Is a theory of meaning construed in this way compatible with a scientific semantics? Do they have the same aims and explanatory goals? After discussing Davidsonian truth-conditional semantics I will argue that the answer to these two questions is negative. My analysis is applicable not only to the Davidsonian program, for what I highlight is a symptom of all externalist theories of meaning. Such attempts fail, I argue, not so much because there is no connection between the outside world and what is in the head; rather, an externalist theory of meaning fails because the sort of connection claimed by externalists is either nonexistent or so amorphous that its attempted systematisation puts into question a coherent, fruitful and scientific externalist theory of semantics.

In "Truth and meaning", the classic paper that convinced many philosophers and linguists of the indispensability of truth conditions in semantics, Davidson

attempts to construct a compositional theory of meaning that can "give the meaning of all expressions in a certain infinite set on the basis of the meaning of the parts" (Davidson 1967: 305). He stresses the primacy that sentences have in his theory of meaning and argues (with Frege) that only in the context of a sentence do words have a meaning. Davidson considers the fruitfulness of using the locutions *means that* or *meanings* and concludes that "the one thing that meanings do not seem to do is oil the wheels of a theory of meaning." That is, his "objection to meanings in the theory of meaning is not that they are abstract or that their identity conditions are obscure, but that they have no demonstrated use" (Davidson 1967: 307). We saw above that internalist semantics posits that syntax (broadly conceived) plus a lexicon are sufficient to explain semantics in natural language. Davidson rejects such a proposal by appealing to "the fact" that "recursive syntax with dictionary added is not necessarily recursive semantics" (Davidson 1967: 308). The locutions *s means m* or *meanings* cannot form the basis of a semantic theory because they cannot be used to provide "for every sentence *s* in the language under study, a matching sentence [… that] 'gives the meaning' of *s*" (Davidson 1967: 309). Such locutions, says Davidson, lead us nowhere and present problems that are as hard as or identical to the problems that a theory of meaning attempts to solve. Davidson then concludes that the only way to construct a theory of meaning is to "sweep away the obscure 'means that'" and replace it with truth conditions. That is, for a semantic theory to "have done its work" it must provide for each sentence in the language under study, a matching sentence that gives the meaning of the former sentence.

Davidson argues that instead of constructing a theory of meaning using such locutions as *s means that p*, one must replace them with the following T-sentence: "*s* is *T* if and only if *p*". This schema allows one to translate the sentence *s* by replacing the sentence *p* with a sentence in a metalanguage. Davidson then argues, augmenting Tarski's Convention T (Tarski 1956), that "it is clear that the sentences to which the predicate 'is *T*' applies will be just the true sentences of [a language] *L*" (Davidson 1967: 309). In other words, this proposal amounts to replacing the locution *s means that p* with *s is true if and only if p*. That is, "a theory of meaning for a language L shows 'how the meanings of sentences depend upon the meanings of words' if it contains a (recursive) definition of truth-in-L", and thus "to give truth conditions is a way of giving the meaning of a sentence" (Davidson 1967: 310).

The classic example of what truth definitions look like in a Davidsonian theory of meaning is the following T-sentence: "*Der Schnee ist weiss*" *is true iff* "*snow is white*". Thus, an object language sentence replaces *s*, and a metalanguage that

provides conditions under which *s* is true replaces *p*. As Lepore & Ludwig (2007: 4) put it, "the key idea of truth-theoretic semantics is that placing certain constraints on an axiomatic truth theory will [...] put us in a position, knowing that the theory meets the constraints, to use it to interpret object language sentences and to see how understanding of them depends on an understanding of their parts and mode of combination." Truth-theoretic semanticists at times say that truth is not supposed to be understood as meaning and that stating conditions under which the object language is true is not to state what that sentence means. However, Davidson early on argued that "'[s]ince a truth definition determines the truth value of every sentence in the object language (relative to a sentence in the metalanguage), it determines the meaning of every word and sentence" (Davidson 1967: 322, fn. 8). But this is just a terminological matter, for the theory is not conceived as a meaning theory per se; but rather as an interpretive theory that provides "all the knowledge that a compositional meaning theory is intended to" (Lepore & Ludwig 2004: 317). Davidson thus aims to "sweep away the obscure 'means that'" while still retaining the explanatory insights a compositional meaning theory offers (Davidson 1967: 309).

To summarise, Davidson's theory includes the following three interrelated claims: (i) that a theory of meaning for *L* is a truth-conditional semantics for *L*, (ii) that to know the meaning of an expression in *L* is to know a satisfaction condition for that expression, and (iii) that meanings are satisfaction conditions. Horwich (2008: 309) remarks that a version of a truth-theoretic approach to semantics is "widely endorsed amongst both linguists and philosophers". This conception of semantics is so ingrained that, as Cummins puts it, "it is something of a challenge to get philosophers of language to realize that The Conjecture [of Davidson's] is not obviously *true*. Generations of philosophers have been trained to regard The Conjecture as a truism. What else could semantics be? Surely, to understand an expression, one must know the conditions under which it is satisfied!" (Cummins 2002: 153, emphasis in original). Like Horwich but for different reasons, Cummins is skeptical of truth-theoretic semantics and remarks that "we are now in a position to see that it is probably false, but I do not expect many to agree with me about this" (Cummins 2002: 153). He argues that the effect of taking truth-conditional semantics as a truism meant that much of philosophy of mind took it upon itself to explain how mental representations could have the satisfaction conditions that truth-conditional semantics required. In other words, once you assume "a Davidsonian story about the semantics of natural language, it is nearly irresistible to conclude that intentional states or mental representations (or both) must have a truth-conditional semantics as well" (Cummins 2002:

153).

But this inference from semantics to the nature of mental representation is problematic, says Cummins, for we have good independent reasons to think that mental representations do not have a truth-conditional semantics. Cummins distinguishes two notions of meaning: the first is the *communicative meaning* of a term, which is whatever must be in the mind to allow the understanding of the term. The second is the *truth-conditional meaning* of a term, which is the term's satisfaction condition or its role in generating one in context. Cummins argues that if one accepts that a constraint on a theory of meaning is that it needs to explain whatever it is that has to be grasped or possessed for linguistic communication to be successful, then the mental representations required for linguistic understanding do not have a truth-conditional semantics. Thus, "a theory of language understanding will make no use of truth-conditional semantics", for "there is no good reason to think that a truth-conditional semantics for natural language will have any place in a mature psycholinguistics" (Cummins 2002: 155).

The independent reason Cummins refers to that shows that mental representations do not have truth conditions is as follows. On the externalist understanding of meaning, the concept of, say, *horse* is the mental representation the reference of which is either horses or the property of being a horse. Cummins rejects this conception of meaning in favour of the claim that the concept of a horse is "a body of knowledge loosely identified by its topic. Just as a book about horses has horses as its topic, but not its referent, so a concept of horses has horses or the property of being a horse as its topic rather than its referent" (Cummins 2002: 158). A loose analogy to this conception of meaning would be to say that "a concept (of horses, say) is a *theory* (of horses), the idea being that theories are organized bodies of knowledge that we identify in much the way we identify concepts—viz., by specifying a topic" (Cummins 2002: 158, emphasis in original). Theories, of course, are identified by their topics not by their referents or satisfaction conditions, and the same is true in the case of concepts. If we conceive of concepts as tacit theories, and agree that they are what you need to have in the mind in order to understand particular terms, then concepts do not semantically combine in the way that is required by truth-conditional semantics. In other words, the sort of semantics invoked by "Tarskian combinatorics [is] hopeless in connection with the sorts of psychological structures concepts must be to do their jobs" (Cummins 2002: 159).

Cummins argues that truth-conditional semantics is psychologically implausible, for we have independent reasons to think that the content of concepts is not truth-conditional. But there is a deeper problem with the sort of externalist

semantics exemplified by Davidson. As Slezak (2014; 2018) discusses, the semantic intuitions of the theorist are in this case relied upon to identify the objects to be explained by the theory. But this reliance on the intuitions of meaningfulness is illegitimate for it relies upon the very mental ability to be explained. The T-sentence is explicitly invoked as a sentence that is expressed in a theoretical metalanguage understood by the theorist. Davidson is clear about this when he remarks that the "inevitable goal of semantic theory is a theory of a natural language couched in a natural language (the same or another)" (Davidson 1973a: 71), and that "it is one condition on the correctness of a theory of meaning that it be such that if an interpreter knew it to be true of a speaker, the interpreter could understand what the speaker said" (Davidson 1995: 131).[1] But understanding the language as speaker is implicitly relied upon here in order to make the explanation work.

The two volume study of truth-theoretic semantics by Lepore & Ludwig (2005; 2007) is very clear about the conception of semantics being an interpretive enterprise, and thus as being in direct opposition to the internalist conception of semantics. As they put it:

> An interpretive truth theory shows how we understand complex expressions on the basis of understanding their significant components. But, as we have said, it does not state how we do it. For the illumination for a particular language presupposes grasp of another language, the metalanguage, in which the theory is given. It is through our already grasping a language which is at least equal in expressive power to the object language, and in some respects greater (the object language need not have the resources to give its own truth theory), that we are able to see in detail how the semantic combinatorics of the object language work. (Lepore & Ludwig 2007: 9)

This should come as no surprise, they say, for "there is no question of a standpoint for understanding meaning that is outside of language altogether." That is, "the most fundamental and powerful devices for representation can obviously not be explicated without the use of just those devices. We can then at best show how they work by showing how they systematically contribute to how we understand sentences in which they appear. And there will be no way to do this that does not mirror the structure of the sentences whose structure we seek to illuminate" (Lepore & Ludwig 2007: 9).

Riemer (2005; 2010; 2015) also argues that semantics is interpretive. He says that semantics is a project that is "essentially hermeneutic" and that it is "a hu-

[1] See Slezak (2018) for discussion of these quotes and others.

manistic discipline closely linked with literary studies" (Riemer 2019: 42-43). It follows from this that the "subjective character of semantic analysis is irreducible, and that real empirical progress in all varieties of linguistics is dependent on an acceptance of this fact" (Riemer 2005: 4). On this view, the irreducibly interpretive character of the study of meaning is due to "the fact that central theoretical features of the explanation of semantic phenomena have no other justification than the subjective judgement of the investigators" (Riemer 2005: 3). One might worry that the inherent subjective character of semantic analysis clashes with the goals of empiricism, psychological realism, and scientificity in linguistics, but Riemer argues that analyses of meaning such as those of cognitive semantics (Lakoff & Johnson 1980; Lakoff 1987; Langacker 1987) "remain genuinely useful and explanatory – despite, or rather because of, the acknowledged subjectivity at their core" (Riemer 2005: 4). In other words, if their "explanatory power turns out not to be a 'scientific' one, then so be it", for such semantic models preserve "their explanatory value regardless of their ultimate status as 'science'" (Riemer 2005: 4).

The description of meaning is, according to Riemer, "infinitely less constrained and more open to varying characterizations than is the description of morphology or syntax", and as a result "semantics has much more to lose by a tolerance towards alternative descriptions, and runs the risk that any analytical specificity about the nature of a single meaning/conceptualization will be lost in a scatter of divergent but equally endorsed analyses" (Riemer 2005: 8). The comparison here with syntax is revealing, for of course we cannot study syntax hermeneutically in the same way that we study semantics hermeneutically (there is no meaning to understand qua speaker in syntax in the same way that there is in semantics). However, we *can* study semantics *non*-hermeneutically. That is, we can study semantics on the model of syntax, phonology, and morphology. One can agree with Riemer (2005: 8) that "any meaning is open to a variety of different, often incompatible, descriptions" and that "the choice of the optimal description is a prerequisite if the analysis is to attain a minimal degree of empirical specificity", but such an argument applies only to semantics construed as a project aimed at producing meta-linguistic *descriptions* of meaning. The internalist project discussed in this book has a different aim, one that is not descriptive but rather generative in the sense that it explicitly aims to unearth the mechanisms that are responsible for the generation of meaning. That is, the aim of internalist semantics is not to provide an optimal description of a particular meaning but rather to discover the underlying mechanisms that generated that meaning in the first place. Riemer concludes that semantics is not science because of its irre-

ducibly interpretive character, and he claims that real empirical progress in *all varieties of linguistics* is dependent on an acceptance of this fact. But this is only the case if semantics is understood hermeneutically, and it fails to acknowledge that semantics in the internalist and biolinguistic tradition is not understood nor practiced in this way.

Unlike leading externalists, Riemer recognises that semantics qua hermeneutics is not scientific. Cognitive semantics is inherently interpretive in the same sense that truth-theoretic semantics is. However, as Slezak (2018) shows, if conceived as a *scientific* project, truth-theoretic semantics has a fatal explanatory error, one that is apparent in other domains and that renders such theories explanatorily vacuous from a scientific perspective. The recent study of Quine and Davidson by Kemp (2012) comes to the same conclusion. He remarks that the "bottom line is that the intuition or semantical judgement of the interpreter cannot be removed from the loop, and thus the theory fails to measure up to the standards of impersonal science" (Kemp 2012: 12). Davidson's is a "non-naturalist standpoint" that is "an unscientific if intuitive standpoint" (Kemp 2012: 12); it is a pragmatic account that "relies ineliminably on an inarticulate human skill or art" (Kemp 2012: 143).

Given the above, one would perhaps be content to conceive of truth-theoretic semantics as a hermeneutic project, but what is one to make of Davidson's proclamations to the contrary? As noted above, he remarks that "my own approach to the description, analysis (in a rough sense), and explanation of thought, language, and action has [...] what I take to be some of the characteristics of a science" (Davidson 1995: 123). Moreover, his unified theory of speech and action (of which his theory of meaning is part) "presents a clear and precise formal structure with demonstrable merits", and it does so "like any scientific theory" (Davidson 1995: 126). Davidson takes his theory to be a psychological theory, but he hedges his bets by saying that whether "the features of a psychological theory I have been rehearsing [...] show that a psychological theory is so different from a theory in the natural sciences as not to deserve to be called a science I do not know, nor much care." What he is "sure of is that such a theory, though it may be as genuine a theory as any, is not in competition with any natural science" (Davidson 1995: 134). Here and elsewhere Davidson is clearly ambivalent about the scientific aims of his theory, but as we'll see in the next chapter there are plenty of externalists (both Davidsonians and not) who unambiguously claim their project to have the same explanatory aims as scientific pursuits like cognitive psychology.

What is going on here? On the one hand, we have claims that "[e]xternalism sets limits to how complete psychological explanation can be, since it introduces

into the heart of the subject elements that no psychological theory can pretend to explain." But on the other hand, "this feature in itself makes psychological theory no less scientific than volcanology, biology, meteorology, or the theory of evolution" (Davidson 1995: 129-130). Davidson at times claims that there is a fundamental difference in aims and interests between his theory and Chomsky's in that he is not trying to understand and explain the same phenomena. Yet at other times he explicitly conceives of his externalist theory as psychological and scientific. If Slezak (2018) is right, then Davidson's truth-theoretic semantics is a specific and revealing example of an explanatory theory that relies on the inner abilities it purports to explain. But conceiving of Davidsonian semantics as a hermeneutic project shields it from such criticism, for we can understand each other in a hermeneutic manner as speakers of a language and not as theorists of language in the scientific sense. As noted above, there are two senses of understanding here that are often conflated. One sense of understanding language is the project that takes semantics to be essentially hermeneutic, as a largely humanistic discipline. The other sense is the scientific or psychological sense qua internalist semantics.[2]

It is noteworthy that late in their careers, both Quine and Davidson acknowledged the affinities between their Analytical philosophical methods, attitudes and traditions and that of the hermeneutic and Continental project (see Slezak 2018 for quotes and discussion). This affinity is explicit in Malpas (2011), which is a collection of leading scholars reflecting on Davidson's work. In the book's foreword, Dagfinn Føllesdal writes that it is "easy to connect him [Davidson] with the hermeneutic tradition, particularly with the new hermeneutics, Heidegger and Gadamer and their followers" (xii).

[2] See Riemer (2019) for evidence that cognitive linguistics makes the same conflation.

4 The science of semantics: Aims, methods, and aspirations

In this concluding chapter I link the discussion of internalist and externalist semantics with a discussion of what scientific explanations look like in general. I argue that a fruitful scientific explanation is one that aims to uncover the underlying mechanisms in virtue of which the observable phenomena are made possible, and that a scientific semantics should be doing just that. If this is so, then a science of semantics is unlikely to be an externalist one, for reasons having to do with the subject matter and form of externalist theories.

I should make clear at the outset that even though I argue that externalism is a hermeneutic project and thus not a scientific one, my criticism should not be taken to be dismissive of the hermeneutic approach to semantics. The hermeneutic approach has provided and continues to provide great insight into the use of language in human social interaction, but this book is concerned with what a science of semantics should look like. My criticism is thus not aimed at externalists per se but rather at those of them who claim to be part of the scientific project. Considering the latter type of externalists, one should of course judge their theories by the same standards as internalist theories (and scientific theories in general). In other words, taking for granted the claim by both sides to be doing science, the real interest in the externalism/internalism debate comes when one considers which questions, aims, and theoretical interests are more likely to produce a fruitful explanatory scientific semantics.

4.1 The nature of scientific explanations

What follow are some remarks on the nature of scientific explanation. I argue that the aims and practices of externalism are orthogonal to those of cognitive psychology. The main reason for this is that, unlike the externalist approach, a fruitful scientific explanation is one that aims to uncover the underlying mechanisms in virtue of which the observable phenomena are made possible. I first unpack this view of scientific explanation, and then offer some remarks on the

implication of this view for externalism.

One can make the strong claim that many scientific explanatory theories – perhaps all but physics – follow what Thagard (2012) refers to as the *mechanista* view of scientific method, according to which to explain a phenomenon is to unearth the mechanism that produces it. Fodor (1968) and Cummins (1975; 1983) are early versions of this sort of approach to explanation. It has been developed more recently by, among others, Bechtel & Richardson (1993), Glennan (1996), Machamer, Darden & Craver (2000), Craver (2007), and Bechtel (2009), though this conception goes back to Descartes and Boyle (see Bechtel 2011 for discussion). On this view, the aim of science is the discovery of mechanisms rather than laws. Machamer, Darden & Craver (2000) argue that much of the practice of science can be understood in this way. Mechanisms, according to them, are "identified and individuated by the activities and entities that constitute them, by their start and finish conditions, and by their functional roles" (Machamer, Darden & Craver 2000: 6). A mechanism is defined by them as a regular series of activities of entities that bring about a particular phenomenon. The emphasis here is on what the activities of mechanisms produce, rather than merely on the changes in the properties of the mechanisms. The construal of scientific explanation in terms of the unearthing of mechanisms is a different project to that of the discovery of laws. In fact, subsumption under law is a misunderstanding of how fruitful scientific explanation works (see Cummins 2000). Machamer, Darden & Craver (2000: 8) give an example from biology according to which if a single base were changed in DNA and the mechanism of protein synthesis operated as usual, then a counterfactual would be supported. "No philosophical work is done," they say, "by positing some further thing, a law, that underwrites the productivity of activities." Activities are constitutive of mechanisms, and it is they that make phenomena intelligible. In other words, the intelligibility consists "in mechanisms being portrayed in terms of a field's bottom out entities and activities" (Machamer, Darden & Craver 2000: 21). So it is not regularities or laws that explain. Rather, what does the explaining are the mechanisms in virtue of which the observed regularities are made possible.

It should be noted that mechanistic explanations are not reductive explanations – one cannot use them to deductively predict from a lower level what will occur at a higher level. The decomposition into mechanisms (and into mechanisms of mechanisms) preserves the higher levels, and indeed a mechanistic explanation would be incomplete without a hierarchy of levels. In other words, multiple levels are required in order to properly explain a particular phenomenon, and it is the integration of different levels that makes phenomena intelligible.

Moreover, it is striking that despite its apparent prevalence, mechanistic explanation was largely ignored in the philosophy of science in the twentieth century. Bechtel (2009) shows how biologists and psychologists rarely make use of laws in giving explanations, and in the relatively few cases in which they do the laws tend to be those of physics or chemistry (see also Bechtel 2008; Bechtel & Abrahamsen 2005). In the case of biology, say, there is an "ubiquity of references to mechanism" and a "sparseness of references to laws" (Bechtel & Abrahamsen 2005: 423). Cummins (2000: 140) speaks of the scandal in regard to the widespread belief that scientific explanation is subsumption under law: "Laws tell us what the mind does, not how it does it. We want to know how the mind works, not just what it does." He gives the example of the McGurk Effect in psychology: "no one thinks that the McGurk effect explains the data it subsumes", no one "would suppose that one could explain why someone hears a consonant like the speaking mouth appears to make by appeal to the McGurk effect"; this is because that "just *is* the McGurk effect" (Cummins 2000: 119, emphasis in original). In other words, laws describe the data, they do not explain the data.

Bechtel & Abrahamsen (2005) discuss how the identification of phenomena in biology precedes their explanation. There is a sense in which there is no other way to go about scientific explanation – we cannot know in advance what it is that needs explaining. Asking the right questions in science is an important part of what makes a particular explanatory theory successful. In cognitive psychology, identifying phenomena of, say, behavioural dispositions or of language use precedes their theoretical explanation. We of course need to know what it is that humans are doing when they use language, but that is not an explanation – it is a description. What we have in externalist theories is a description of regularities and "laws" of language use or of behaviour. There is much debate about ascription, about what a particular behaviour or linguistic output should be labelled as. But regardless of the value and interest of such descriptions – and it is far from nil – such theories are not fruitful *explanatory* theories: they are the explananda, not the explanantia. To conceive of the externalist research project as scientific – specifically, as having the same aims as cognitive psychology and internalist semantics – is a misunderstanding that fails to see the force and value of hermeneutic research. Investigating the way in which words are individuated and the way in which social norms come into play when people use language, amongst other topics, is a valuable and interesting project. But this tells us little about the mechanisms in virtue of which such language use is made possible. Another way to put the matter is as follows. Hermeneutic research does not explain the underlying mechanisms of language but rather *uses* them to investigate the

world via language. Internalism, on the other hand, takes its subject matter to be the underlying mechanisms themselves, not their use.

Externalism is a hermeneutic project in which linguistic abilities are used in the investigation of the world. Its methods are ill-suited to the investigation of those underlying mechanisms of language use, for that leads to explanatory vacuousness in which the abilities that are purportedly being explained are relied upon implicitly. The next section is devoted to showing in detail that if externalism is understood as science then it indeed leads to such illegitimate reliance on the very phenomenon to be explained.

4.2 Externalism and scientific explanations

As noted above, there are a number of leading externalists who explicitly conceive their projects to be scientific ones. Horwich, for example, says of Davidson's externalist truth-theoretic program that it "became widely accepted, instigating several decades of 'normal science' in semantics" (Horwich 2001: 371). Despite the immense popularity of truth-theoretic semantics, Horwich is one of a handful of critics within the externalist camp that have called into question the concept of truth as a basis for a semantic theory. Their deflationary theory of truth argues that to assert that a particular sentence is true is equivalent to merely asserting the sentence on its own. That is, the claim is that asserting the sentence "'snow is white' is true" is equivalent to merely asserting that snow is white. In other words, "[o]ur use and grasp of the concept of truth is adequately explained by our tendency [...] to accept $<A>$ *is true* when, and only when, we are prepared to accept A" (Armour-Garb & Beall 2005: 12, emphasis in original). Deflationists argue that in order to understand a concept one must consider its function, and therefore in order "to understand what truth *is*, we must consider what truth *does*" and "consideration of what it does reveals that it has no underlying nature or structure at all – there is nothing to truth" (Armour-Garb & Beall 2005: 17, emphasis in original). The deflationary account thus precludes the analysis of the nature of meaning in terms of truth conditions.

The deflationary alternative to truth-theoretic semantics, however, is still an externalist account of semantics. It is worth looking at this alternative, for it shows that the problems with truth-theoretical theories of meaning are due to their externalist conception of meaning and not due to their truth-theoretical formal apparatus. Horwich's use-based semantics, whilst not truth-theoretical, is still externalist and suffers from similar problems if construed as a cognitive psychological scientific project. Horwich argues that his use-based semantics,

essentially a defence of Wittgenstein's idea of a use theory of meaning, is compatible with a linguistics construed as an empirical science, but the reasons for rejecting this claim are the same as the reasons for rejecting any externalist theory of meaning if one wishes to construct a fruitful scientific semantics.

Horwich is critical of mainstream formal semantics and argues that "as far as explaining our linguistic activity is concerned, there is no reason at all to think that understanding has a truth theoretic basis" (Horwich 2008: 318). He claims that while the problems truth-theoretic semantics presents "are highly challenging, requiring considerable skill and ingenuity, and that enormous progress has been made in these endeavours over the last forty years or so", citing "such progress is not enough to vindicate truth-theoretic semantics as an *empirical* subject, as an integral part of the global scientific enterprise" (Horwich 2008: 318, fn. 12, emphasis in original). He argues that in order to be a part of science, truth-theoretic semantics must show how their derivations have contributed to the explanation of observable events. However, "that has not, and cannot, be done" (Horwich 2008: 318). His main objection has to do with compositionality and the assumption of formal semanticists that "the project of semantics needs to start [...] with theoretical assumptions about *the meanings of sentences*" (Horwich 2008: 314, emphasis in original). As we saw above, truth-theoretic semantics is an analysis that focuses on sentences. As Lepore & Ludwig (2007: 4) put it, the "goal is not to provide an analysis of the concept of meaning, or an analysis piecemeal of particular words or what it is for someone to understand them, but to illuminate as a whole the set of concepts deployed in understanding other speakers by considering how one could confirm such a theory on the basis of evidence described without appeal to those concepts." Horwich takes the opposite approach, for he believes that compositionality is relatively easy to accommodate and thus one needs to first "somehow identify [...] the theoretical-meanings of *words*, and then, presupposing compositionality, to trivially deduce the theoretical-meanings of sentences" (Horwich 2008: 314, emphasis in original).

Inverting the focus of semantics from sentences to words, says Horwich, has the effect of nullifying truth-theoretic semantics, for truth conditions apply to sentences and cannot apply to words. Given this focus on words, Horwich argues that the theoretical characterisation of the meanings of words will be deduced from "certain facts concerning sentence *usage*, rather than sentence *meaning*" (Horwich 2008: 314, emphasis in original). That is, once the meanings of words are deduced from observations of sentence usage, "we will – in light of compositionality – be able to arrive [...] at the meanings of sentences." Thus, according

to Horwich's use-based semantics and *pace* Davidson, "it will be trivially easy to deduce what any sentence means from the structure of that sentence and what its words mean" (Horwich 2008: 314). A simple example of Horwich's use-based semantics is the following:

> Presumably our understanding of the sentence "dogs bark" arises somehow from our understanding of its components and our appreciation of how they are combined. That is to say, "dogs bark" somehow gets its meaning (or, at least, one of its meanings) from the meanings of the two words "dog" and "bark", from the meaning of the generalization schema "ns v", and from the fact that the sentence results from placing those words in that schema in a certain order. (Horwich 1998: 154)

So on this account the meaning of the sentence *dogs bark* is deduced by combining a word meaning *dog* with a word meaning *barks*. Therefore, according to Horwich, understanding complex expressions is nothing over and above understanding their parts and knowing how they are combined. This is what he means when he claims that compositionality is a relatively trivial matter. Knowing the meanings of words and being aware of their mode of combination is all that is required, says Horwich, in order to understand the meanings of sentences: "No further work is required; no further process needs to be involved, leading from those initial conditions to the state of understanding the sentence" (Horwich 1998: 155).

It might seem that Horwich's account is compatible with an internalist semantics, for the latter also claims that all that is required for the explanation of sentence meaning is the primitive lexical elements and the syntax defining the ways in which they can be combined. But the way in which the primitives are explained in internalist semantics is very different to Horwich's use-based account. Horwich is critical of truth-theoretic semantics because he thinks it lacks the necessary explanatory power. That is, he says, the observable events that are of interest to semantics are "items of verbal activity – both mental and behavioural", and semantics "is obliged to explain [...] facts concerning the circumstances in which sentences are *accepted*" (Horwich 2008: 315, emphasis in original). But truth-theoretic semantics cannot sufficiently explain such facts and is thus a failed enterprise: it cannot (but it must, says Horwich) "tell us what it is about, e.g., 'The sky is blue' that explains why it tends to be *recognized* as true if and only if it *is* true" (Horwich 2008: 317, emphasis in original).

This is clearly still an externalist semantic theory, for even though it rejects truth conditions it claims that "the underlying basis of each word's meaning is

the (idealized) law governing its usage— a law that dictates the 'acceptance conditions' of certain specified sentences containing it" (Horwich 2005: 26). This law of acceptance conditions, argues Horwich, solves the puzzle of why it is that *The sky is blue* tends to be recognised as true. The law would stipulate, for example, that the meaning of *red* "stems from the fact that its law of use is a propensity to accept 'That is red' in response to the sort of visual experience normally provoked by observing a clearly red surface" or that "'and' means what it does because the fundamental regularity in its use is our acceptance of the two-way argument schema, 'p, q // p and q'" (Horwich 2005: 26). The law of acceptance conditions, which is supposed to underwrite Horwich's semantic theory, is explicitly understood to be on par with a linguistics construed as an empirical science. But as we'll now see, laws of this kind are problematic at best.

Horwich argues that the phenomena that semantics needs to explain are those of sentence acceptance. He elaborates: "I don't mean 'accepted as *grammatical*', but 'accepted as *true*', i.e., 'in the belief-box'. Acceptance sometimes leads to *utterance* (depending on the speaker's desires); therefore explaining the acceptance of a sentence may contribute to explain its being uttered" (Horwich 2008: 315, fn. 9, emphasis in original). Sentence acceptance is explained by the following:

> The meaning of a word, w, is engendered by the non-semantic feature of w that explains w's overall deployment. And this will be an acceptance-property of the following form:— "that such-and-such w-sentences are regularly accepted in such-and-such circumstances" is the idealized law governing w's use is [*sic*] (by the relevant "experts", given certain meanings attached to various other words). (Horwich 2005: 28)

According to the use theory of meaning, then, a word means what it does "*in virtue* of its basic use; a word's use is *responsible* for its meaning what it does. Thus, not only does a meaning-property *supervene* on a basic acceptance property, but possession of the former is *immediately explained* by possession of the latter" (Horwich 2005: 32, emphasis in original).

Horwich argues that insofar as "linguistics is an empirical science – standing alongside psychology, neurology, biology, physics, etc.", then such acceptance-laws "should be testable against concrete observable events" (Horwich 2008: 315). Thus, "the semanticist of a given language ought to be looking, concerning each word, for the basic law governing its use" (Horwich 2008: 319), and if such laws are forthcoming and explanatorily fruitful, Horwich believes that "[s]emantics would then somewhat resemble fundamental physics" (Horwich 2008: 318). The phenomena of sentence acceptance is supposed to cohere with phonology, syntax, and pragmatics to yield a science of language use. Horwich argues that

truth-theoretic semantics cannot yield such a science but that his use-based semantics can. However, both are problematic if construed as science. One problem with sentence acceptance, a main tenet of Horwich's theory, is that it is unclear whether it can be generalised beyond the examples that Horwich gives. As Schiffer (2000) argues, meaning-constituting properties are supposed to be acceptance properties, but it is not even clear whether relatively simple words like *dog* have acceptance properties. As he puts it, there are no plausible candidates for "a kind of 'dog'-containing sentence *K* and a kind of circumstance *C* such that a speaker for whom 'dog' means DOG will be disposed to accept a sentence of kind *K* in circumstances of kind *C*, and that fact will belong to the explanation of his accepting any other sentence that contains 'dog'" (Schiffer 2000: 534). For instance, "'[d]og' may mean DOG for someone who is blind or who does not know what a dog looks like, so it cannot be required that anyone who understands 'dog' must be disposed to accept 'That's a dog' when confronted with a paradigm dog" (Schiffer 2000: 534).

Even granting the validity of acceptance properties, it is unclear whether sentence meaning can be reduced to sentence acceptance because the latter involves much more than what is traditionally thought of as sentence meaning. Consider the following example from Gupta (1993). Suppose that a predicate G has the following stipulative definition: a thing is G if and only if it is red and round. Given the close connection between the acceptance properties and meaning properties of G, it may appear that the two can be regarded as the same thing. But such a definition, argues Gupta, may play only a minimal role in the explanation of a person's acceptance of sentences containing G. The fundamental role in such an explanation may be played by, for example, the authority of some expert (if, say, the person trusts the expert's colour reports). There is thus "little reason to think [...] that 'explanatorily basic patterns [of sentence acceptance]'" in Horwich's use theory of meaning "provide the meaning of a word", for "plainly, the acceptance of sentences depends not just on the meanings of words but also on the methods of obtaining information (and misinformation) about the world." In other words, "we should distinguish general ideas such as 'meaning is use' and 'meaning explains use' from Horwich's particular claim. The former may express truisms, the latter does not" (Gupta 1993: 666) (see also Gupta 2003).

Gupta's remark that the acceptance of sentences depends not just on the meanings of words but also on the methods of obtaining information about the world hints at the main reason why externalist theories such as Horwich's cannot serve as a foundation upon which to construct a science of semantics: such theories have a problematic choice of subject matter. The scope of semantic theories was

discussed by Katz & Fodor (1963) in the early days of internalist semantics and it's worth briefly revisiting here, for it bears directly on the problems with the scope of externalist theories of meaning. Katz and Fodor ask the reader to compare the following three sentences: *Should we take junior back to the zoo? Should we take the lion back to the zoo? Should we take the bus back to the zoo?* They then remark that, for example, "[i]nformation which figures in the choice of the correct readings for these sentences includes the fact that lions, but not children and busses, are often kept in cages." After listing a handful of other examples, they note that the "reader will find it an easy matter to construct an ambiguous sentence whose resolution requires the representation of practically any item of information about the world he chooses." But "a complete theory of setting selection must represent as part of the setting of an utterance any and every feature of the world which speakers need in order to determine the preferred reading of that utterance", and "practically any item of information about the world is essential to some disambiguations" (Katz & Fodor 1963: 179). If this is so then a number of conclusions follow.

The first conclusion is that a theory that insists (as externalism does) on including the mind's relations to the external world in a theory of language cannot hope to find reliable relations of this sort (let alone systematising them into a fruitful explanatory theory). Second, as Katz and Fodor note, "such a theory cannot in principle distinguish between the speaker's knowledge of his language and his knowledge of the world, because, according to such a theory, part of the characterization of a LINGUISTIC ability is a representation of virtually all knowledge about the world that speakers share" (Katz & Fodor 1963: 179, emphasis in original). Thirdly, Katz and Fodor remark that "since there is no serious possibility of systematizing all the knowledge of the world that speakers share, and since a theory of the kind we have been discussing requires such a systematization, it is ipso facto not a serious model for semantics" (Katz & Fodor 1963: 179). The same is true of externalism. Moreover, despite the efforts of Horwich and others, due to the creative aspect of language use there is little chance of constructing a science of language use (Chomsky 1966; McGilvray 2001; 2005; Asoulin 2013).

But there is a deeper reason. The laws of language use, if they can be formulated at all, at best tell us what a language user *does*, they do not tell us why that is the case nor explain the underlying ability to do so. It is the latter that science seeks to uncover. Scientific laws *describe* the data in question, not explain the data. As Cummins (2000) discusses, there is now a consensus that the deductive nomological (DN) sense in which laws are explanatory is a myth and that "the suspicion grows that it *cannot* be done successfully" because there is no dif-

ference between laws and data: "No laws are explanatory in the sense required by DN" (Cummins 2000: 119, emphasis in original). Cummins (2000: 120) quips that in psychology "we are overwhelmed with things to explain, and somewhat underwhelmed by things to explain them with." The same is true of theories of meaning that take an externalist approach but still claim to be psychological or non-hermeneutic. We are indeed overwhelmed with things to explain in externalism: phenomena of meaning ascription, sentence acceptance properties, truth-evaluable judgements and the like are fascinating phenomena of language use. But they are data of language use, not scientific explanations of language use. This conflation, according to Cummins, "derives from a deep-rooted uncertainty about what it would take to really explain a psychological effect" (Cummins 2000: 121).

We saw above that Horwich claims that if the science of semantics is done the way he proposes then semantics would "somewhat resemble fundamental physics" (Horwich 2008: 318), but this reflects the very conflation that Cummins points to. Let us see why. Semantics cannot resemble fundamental physics any more than geology can, for they are what Fodor (1974) famously called special sciences. Unlike fundamental physics, the special sciences do not yield general laws of nature but only "laws governing the special sorts of systems that are their proper objects of study." Laws of psychology are laws in situ, which "specify effects—regular behavioral patterns characteristic of a specific kind of mechanism" (Cummins 2000: 121). But notice the crucial difference here: the laws in question *describe the effects* of the specific kind of mechanism which is their subject matter. But in order to move from a description to an explanation we need an account of the mechanism itself. Notice that this is not the case at the level of fundamental physics, where "laws are what you get because, at a *fundamental* level, all you can do is say how things are." That is, the "things that obey the fundamental laws of motion (everything) do not have some special constitution or organization that accounts for the fact that they obey those laws" because the "laws of motion just say what motion is in this possible world" (Cummins 2000: 122, emphasis in original). As Cummins argues (Cummins 1975; 1983; 2010; Roth & Cummins 2014), special sciences like psychology (and, of course, like a semantics construed as science) "should seek to discover and specify the effects characteristic of the systems that constitute their proprietary domains, and to explain those effects in terms of the *structure* of those systems, that is, in terms of their constituents (either physical or functional) and their mode of organization" (Cummins 2000: 122, emphasis in original).

I would like to briefly return to the discussion of Davidon's truth-theoretic

semantics to show that the problematic nature of the explanations in use-based semantics qua scientific explanations is also present in the Davidsonian program (and, indeed, in any externalist theory of meaning of this sort). Lepore and Ludwig remark that the "centrepiece and nexus of Davidson's philosophy" is the project of the radical interpreter, and that "the stance of the radical interpreter of another speaker [… is] methodologically basic in understanding language and connected matters" (Lepore & Ludwig 2005: viii, 2). This stance stems from the problem confronting linguists constructing a grammar of a language they do not yet understand. The problem is how to choose amongst competing theories of meaning for the language under investigation. In other words, "given a theory that would make interpretation possible, what evidence plausibly available to a potential interpreter would support the theory to a reasonable degree?" (Davidson 1973: 125). Davidson claims that the problem of interpretation is "domestic as well as foreign: it surfaces for speakers of the same language in the form of the question, how can it be determined that the language is the same?" (Davidson 1973: 125). That is, even in cases of everyday communication by speakers of the same language, the speakers are in a sense theorising interpreters. Two speakers of English, say, who successfully communicate to each other are thus each possessors of a theory of interpretation.

Davidson takes his theory of meaning to model what interpreters are doing in this form of theorising, and he thus uses the terms *interpretation* and *understanding* as if they were interchangeable (see Mulhall 1987). A meaning theory, says Davidson, must allow the interpreter "to understand any of the infinity of sentences the speaker might utter" in the language (Davidson 1973: 127), and thus "someone who knows the theory can interpret the utterances to which the theory applies" (Davidson 1973: 128). The theory Davidson has in mind, of course, is a truth-theoretic theory of meaning, and thus the project of the radical interpreter becomes that of "confirming a truth theory for a speaker's language that can be used to interpret correctly the speaker's utterances" (Lepore & Ludwig 2005: 3).

As already noted, this is a conflation of understanding as speaker and understanding as theorist. The project of the radical interpreter that aims at constructing a compositional meaning theory for a natural language is a hermeneutic project of interpretation because it takes for granted the underlying mechanisms of language. If we take scientific explanations as seeking to unearth the mechanisms responsible for the observed phenomena, then truth-theoretic semantics leaves unexplained the abilities it purports to explain. Davidson's truth-theoretic semantics is a case in point in regard to an externalist approach being ill suited to a scientific semantics. The conflation of the two senses of understanding arises

in part from the often implicit assumption that the investigation of the underlying mechanisms will not be philosophically illuminating. As Davidson notes in regard to his theory: "The point of the theory was not to describe how we actually interpret, but to speculate on what it is about thought and language that makes them interpretable." Furthermore, he remarks that "[a]ll that is lacking at the start [of constructing a theory of meaning] is a shared language and prior knowledge of each other's attitudes." Therefore, since "the theory and the official story of how it can be applied are already remote from actual practice, we must expect that the theory will throw only the most oblique light on the acquisition of a first language, and less still on the origins of speech" (Davidson 1995: 128). Such statements clearly distinguish Davidson's project from the internalist (and biolinguistic) project. It is clear that Davidson does not attempt to explain the mechanisms that underlie language production and comprehension, for his theory explicitly assumes a shared language and prior knowledge of other people's propositional attitudes.

The quest to construct an interpretive truth theory is seen by Davidsonians as the distinguishing mark of a semantic theory: this sets the subject area of their theory apart from an internalist theory of meaning that seeks to explain the very abilities that externalist theories use in constructing their own theories. Davidson (1995: 133) is explicit about this:

> I want to know what it is about propositional thought – our beliefs, desires, intentions, and speech – that makes them intelligible to others. This is a question about the nature of thought and meaning which cannot be answered by discovering neural mechanisms, studying the evolution of the brain, or finding evidence that explain the incredible ease and rapidity with which we come to have a first language.

In other words, the nature of the underlying mechanisms by which language is acquired, produced, and comprehended in the heads of speakers cannot be unearthed via the radical interpreter project. If this is so, then Davidson's claims that his project is scientific displays the conflation between the two senses of understanding (*verstehen* and *erklären*).

Another way to understand the conflation is by considering the difference between a traditional grammar and a generative grammar. This is the distinction between, on the one hand, a descriptive or interpretive theory of language and, on the other hand, an explanatory theory of language qua science. The latter is a grammar in the I-language sense, which is an account of the ideal speaker/hearer's competence. Furthermore, "[i]f the grammar is [...] perfectly

explicit – in other words, if it does not rely on the intelligence of the understanding reader but rather provides an explicit analysis of his contribution – we may (somewhat redundantly) call it a *generative grammar*" (Chomsky 1965: 4, emphasis in original). Recall that a generative grammar is understood intensionally, meaning that it focuses on the specific procedure encoded in the mind that generates the strings of the language (as opposed to the E-language conception that focusses on the strings themselves). A generative grammar is distinct from both an E-language grammar and a traditional prescriptive grammar. Traditional grammars describe how a particular language is or should be used. However, as Chomsky (1980: 237) remarks, traditional grammars "do not provide an analysis of the qualities of intelligence that the reader brings to bear on the information presented." The same is true in the case of externalist theories such as truth-theoretic semantics. Traditional descriptive grammars and interpretive truth-theoretic theories of meaning, whatever their merits (and one should not be dismissive of their accomplishments), provide only examples and hints as to the underlying nature of the language. That is, the success of traditional grammars (and the success of externalist semantics) rests entirely on their pairing with "an intelligent and comprehending reader" (Chomsky 1962: 528). Indeed, the remark of Chomsky (1962: 529) almost 60 years ago that "[r]eliance on the reader's intelligence is so commonplace that its significance may easily be overlooked", is still pertinent today in the case of externalist theories of meaning.

My criticism of descriptive or interpretive theories of language does not stem from the claim that they omit certain facts. Considering their subject matter, such theories haven't left out the mechanisms in the head in virtue of which language is made possible, for that is not their aim. From the perspective of scientific explanations, however, they give an incomplete picture of the nature of language because they assume, indeed they build upon and would be unusable without, the abilities of language users. It is this ability on which internalism and biolinguistics would like to shed light. Lepore & Ludwig (2005: 11) appear to concur when they remark that "Davidson treats compositional meaning theories as empirical theories, theories of particular speakers or natural languages, which must be confirmed on the basis of public evidence." These theories belong "in the context of a theory of interpretation of human action in general", and thus "from this perspective, the role of a theory of interpretation is to identify and systematize patterns in the behaviour of speakers in relation to their environment."

***

Fodor (2000a: 21) notes approvingly in regard to the modern form of the rep-

resentational theory of mind that it has "shed a feature that traditional versions of the doctrine, rationalist and empiricist, invariably took for granted: that RTP [representational theory of perception] must provide not just a psychology of perception but an epistemology, too." Fodor distinguishes what psychology does, which is, among others, to explain the mechanisms that underpin belief fixation, from what epistemology does, which is, among others, to investigate whether one is justified in having a particular belief or meaning, or whether one's particular representation of, say, an object in the world, is a true representation or a misrepresentation. He refers to the attempt to incorporate within one theory both how one has a particular visual perception and how one is justified in believing that particular visual perception as a "double burden" and as the core of what was wrong with the traditional representational theories of perception. However, Fodor fails to see that this same double burden is carried by externalist theories of meaning such as his own.

A different double burden in regard to theories of meaning is discussed by Lepore (1983; 1983a), who argues that the problem of how to characterise what is in the speaker's head is a "non-issue" (Lepore 1983: 185, fn. 5). That is, the problem of unearthing the underlying mechanisms of language in the head "arises only if, as many do, one views semantics as a subfield of psychology." In other words, "[i]f we assume that questions about knowledge and understanding of language are psychological questions, then semantics should be a subfield of psychology." Lepore disagrees with this conception of semantics and argues for an externalist conception according to which semantics "properly understood, is not a subfield of psychology but of epistemology." Thus, he concludes, since semantics and questions about knowledge and understanding of language are not psychological questions but are instead epistemological questions, "we need not worry about what's in the speaker's head – whatever that may mean" (Lepore 1983: 185, fn. 5). If the goals of a theory of meaning are understood as part of epistemology, then perhaps one can make the case that psychology is beside the point. But if this is the case then what is one to make of the claims of Burge, Davidson, Fodor, and Horwich that their externalist theories of meaning are scientific? In other words, if psychology aims to discover the mechanisms in virtue of which language is made possible, and if that means that psychology is part of science, then, *pace* Lepore, if one's research program is scientific then we do need to worry about what's in the head, we do need to have a semantics that is informed by (and in turn informs) cognitive psychology.

Either an externalist theory of meaning is scientific and should thus be answerable to or at least in principle be able to be integrated with the other sciences,

or it is a hermeneutic theory and thus, as Lepore argues, it isn't answerable to and can remain impartial in regard to science. I argued in this book that externalist theories are indeed hermeneutic and thus are not explanatory theories in the scientific sense.

References

Armour-Garb, Bradley P. & J. C. Beall. 2005. Deflationism: The basics. In Bradley P. Armour-Garb & J. C. Beall (eds.), *Deflationary truth*, 1–30. Illinois: Open Court Publishing.

Asoulin, Eran. 2013. The creative aspect of language use and the implications for linguistic science. *Biolinguistics* 7. 228–248.

Asoulin, Eran. 2016. Language as an instrument of thought. *Glossa: a journal of general linguistics 1(1):* 46. 1–23.

Asoulin, Eran. 2019. Phrase structure grammars as indicative of uniquely human thoughts. *Language Sciences* 74. 98–109.

Asoulin, Eran. 2020. Why should syntactic islands exist? *Mind & Language*.

Bechtel, William. 2008. *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Routledge.

Bechtel, William. 2009. Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology* 22(5). 543–564.

Bechtel, William. 2011. Mechanism and biological explanation. *Philosophy of Science* 78. 533–557.

Bechtel, William & Adele Abrahamsen. 2005. Explanation: A mechanist alternative. *Studies in History and Philosophy of Biology & Biomedical Science* 36(2). 421–441.

Bechtel, William & Robert C. Richardson. 1993. *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

Ben-Menahem, Yemima. 2005. Introduction. In Yemima Ben-Menahem (ed.), *Hilary Putnam*, 1–16. Cambridge: Cambridge University Press.

Berwick, Robert C. & Noam Chomsky. 2016. *Why only us: Language and evolution*. Cambridge, MA: MIT Press.

Brentano, Franz. 1995 [1874]. *Psychology from an empirical standpoint*. London: Routledge.

Burge, Tyler. 1979. Individualism and the mental. *Midwest Studies in Philosophy* 4. 73–121.

Burge, Tyler. 1986. Individualism and psychology. *The Philosophical Review* 95(1). 3–45.

Burge, Tyler. 1989. Wherein is language social? In Alexander George (ed.), *Reflections on Chomsky*, 175–191. [Reprinted in his 2007. Foundations of Mind (Philosophical Essays, Vol. 2), pp. 275-290, Oxford: Oxford University Press, from which I quote.] Oxford: Blackwell.

Burge, Tyler. 2003. Reply to Chomsky. In Martin Hahn & Bjørn Ramberg (eds.), *Reflections and replies: Essays on the philosophy of Tyler Burge*, 451–470. Cambridge, MA: MIT Press.

Burton-Roberts, Noel. 2011. On the grounding of syntax and the role of phonology in human cognition. *Lingua* 121. 2089–2102.

Cann, Ronnie. 1993. *Formal semantics: An introduction*. Cambridge: Cambridge University Press.

Chalmers, David J. 2003. The nature of narrow content. *Philosophical Issues* 13. 46–66.

Cherniak, Christopher. 1994. Philosophy and computational neuroanatomy. *Philosophical Studies* 73. 89–107.

Cherniak, Christopher, Zekeria Mokhtarzada & Uri Nodelman. 2002. Optimal-wiring models of neuroanatomy. In Giorgio A. Ascoli (ed.), *Computational neuroanatomy: Principles and methods*, 71–82. Totowa, New Jersey: Humana Press.

Chomsky, Noam. 1962. Explanatory models in linguistics. In Patrick Suppes Ernest Nagel & Alfred Tarski (eds.), *Logic, methodology and philosophy of science: Proceedings of the 1960 International Congress*, 528–550. Stanford: Stanford University Press.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1966. *Cartesian linguistics: A chapter in the history of rationalist thought*. New York: Harper & Row.

Chomsky, Noam. 1975. *Reflections on language*. New York: Pantheon Books.

Chomsky, Noam. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger Publishers.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995a. Language and nature. *Mind, New Series* 104(413). 1–61.

Chomsky, Noam. 2000. *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.

Chomsky, Noam. 2000a. Minimalist inquiries: The framework. In Juan Uriagereka Roger Martin David Michaels & Samuel Jay Keyser (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2002. *On nature and language*. Cambridge: Cambridge University Press.

Chomsky, Noam. 2003. Reply to Egan. In Louise M. Antony & Norbert Hornstein (eds.), *Chomsky and his critics*, 268–274. Oxford: Blackwell Publishing.

Chomsky, Noam. 2003a. Reply to Ludlow. In Louise M. Antony & Norbert Hornstein (eds.), *Chomsky and his critics*, 287–295. Oxford: Blackwell Publishing.

Chomsky, Noam. 2003b. Internalist explorations. In Martin Hahn & Bjørn Ramberg (eds.), *Reflections and replies: Essays on the philosophy of Tyler Burge*, 259–288. Cambridge, MA: MIT Press.

Chomsky, Noam. 2013. Problems of projection. *Lingua* 130. 33–49.

Chomsky, Noam. 2016. *What kind of creatures are we?* New York: Columbia University Press.

Chomsky, Noam. 2016a. Minimal computation and the architecture of language. *The Tsuru University Graduate School Review* 20. 7–17.

Chomsky, Noam, Ángel J. Gallego & Dennis Ott. 2019. Generative grammar and the faculty of language: insights, questions, and challenges. *Catalan Journal of Linguistics Special Issue*. 229–261.

Collins, Chris. 2017. Merge(X,Y) = {X,Y}. In Leah Bauke & Andreas Blümel (eds.), *Labels and roots*, 47–68. Berlin: De Gruyter.

Collins, Chris & Edward Stabler. 2016. A formalization of minimalist syntax. *Syntax* 19. 43–78.

Collins, John. 2004. Faculty disputes. *Mind & Language* 19(5). 503–533.

Collins, John. 2007. Syntax, more or less. *Mind* 116(464). 805–850.

Craver, Carl F. 2007. *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press.

Cummins, Robert. 1975. Functional analysis. *Journal of Philosophy* 72. 741–764.

Cummins, Robert. 1983. *The nature of psychological explanation*. Cambridge, MA: Bradford/MIT Press.

Cummins, Robert. 2000. "How does it work?" versus "What are the laws?": Two conceptions of psychological explanation. In Frank C. Keil & Robert A. Wilson (eds.), *Explanation and cognition*, 117–145. Cambridge, MA: MIT Press.

Cummins, Robert. 2002. Truth and meaning. In Michael O'Rourke Joseph Keim-Campbell & David Shier (eds.), *Meaning and truth: Investigations in philosophical semantics*, 175–197. [Reprinted in his 2010, The world in the head, pp. 152-

173. Oxford: Oxford University Press, from which I quote]. New York: Seven Bridges Press.

Cummins, Robert. 2010. *The world in the head*. Oxford: Oxford University Press.

D'Ambrosio, Justin. 2019. Semantic verbs are intensional transitives. *Mind* 128(509). 213–248.

Davidson, Donald. 1967. Truth and meaning. *Synthese* 17(3). 304–323.

Davidson, Donald. 1973. Radical interpretation. *Dialectica* 27(3-4). [Reprinted in his 2001, Inquiries into truth and interpretation, pp. 125-140, Oxford: Oxford University Press, from which I quote]., 331–328.

Davidson, Donald. 1973a. In defence of convention T. *Studies in Logic and the Foundations of Mathematics* 68. [Reprinted in his 2001, Inquiries into truth and interpretation, pp. 65-75. Oxford: Oxford University Press, from which I quote.], 76–86.

Davidson, Donald. 1975. Thought and talk. In Samuel D. Guttenplan (ed.), *Mind and language: Wolfson College Lectures 1974*, 7–23. Oxford: Clarendon Press.

Davidson, Donald. 1982. Rational animals. *Dialectica* 36(4). 317–328.

Davidson, Donald. 1987. Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association* 60(3). 441–458.

Davidson, Donald. 1995. Could there be a science of rationality? *International Journal of Philosophical Studies* 3. [Reprinted in his 2004, Problems of rationality, pp. 117-134. Oxford: Oxford University Press, from which I quote.], 1–16.

Davidson, Donald. 2005. *Truth and predication*. Harvard: Harvard University Press.

Deacon, Terrence W. 1997. *The symbolic species: The co-evolution of language and the brain*. New York: W.W Norton.

Devitt, Michael. 1984. Thoughts and their ascription. *Midwest Studies in Philosophy* 9(1). 385–420.

Devitt, Michael. 2003. Linguistics is not psychology. In Alex Barber (ed.), *Epistemology of language*, 107–139. Oxford: Oxford University Press.

Devitt, Michael. 2006. *Ignorance of language*. Oxford: Oxford University Press.

Devitt, Michael. 2006a. Intuitions in linguistics. *British Journal for the Philosophy of Science* 57. 481–513.

Devitt, Michael & Kim Sterelny. 1989. Linguistics: What's wrong with "The Right View". *Philosophical Perspectives* 3. 497–531.

Dummett, Michael. 1989. Language and communication. In Alexander George (ed.), *Reflections on Chomsky*, 192–212. Oxford: Blackwell.

Dummett, Michael. 1993. *The seas of language*. Oxford: Clarendon Press.

Dummett, Michael. 1994. *Origins of analytic philosophy*. Harvard: Harvard University Press.
Egan, Frances. 1999. In defence of narrow mindedness. *Mind & Language* 14(2). 177–194.
Farkas, Katalin. 2003. What is externalism? *Philosophical Studies* 112. 187–208.
Farkas, Katalin. 2006. Semantic internalism and externalism. In Barry C. Smith & Ernest Lepore (eds.), *The Oxford Handbook of the philosophy of language*, 323–340. Oxford: Oxford University Press.
Farkas, Katalin. 2008. *The subject's point of view*. Oxford: Oxford University Press.
Ferguson, Kenneth G. 2009. Meaning and the external world. *Erkenntnis* 70. 299–311.
Fodor, Jerry A. 1968. *Psychological explanation*. New York: Random House.
Fodor, Jerry A. 1974. Special sciences (or: The disunity of science as a working hypothesis). *Synthese* 28. 97–115.
Fodor, Jerry A. 1975. *The language of thought*. Cambridge, MA: Harvard University Press.
Fodor, Jerry A. 1994. *The elm and the expert: Mentalese and its semantics*. Cambridge, MA: MIT Press.
Fodor, Jerry A. 1998. *Concepts: Where cognitive science went wrong*. Oxford: Clarendon Press.
Fodor, Jerry A. 2000. It's all in the mind: Noam Chomsky and the arguments for internalism. *Times Literary Supplement June* 23. 3–4.
Fodor, Jerry A. 2000a. A science of tuesdays [review of The threefold cord: Mind, body and world by Hilary Putnam]. *London Review of Books (July 20)* 22(14). 21–22.
Fodor, Jerry A. 2003. *Hume variations*. Oxford: Clarendon Press.
Fodor, Jerry A. 2007. Semantics: An interview with Jerry Fodor. *Revista Virtual de Estudos da Linguagem – ReVEL* 5(8). 1–13. http://www.revel.inf.br/files/entrevistas/revel_8_interview_jerry_fodor.pdf.
Fodor, Jerry A. & Zenon W. Pylyshyn. 2015. *Minds without meanings: An essay on the content of concepts*. Cambridge, MA: MIT Press.
Frege, Gottlob. 1980 [1892]. Function and concept. In *Translations from the philosophical writings of Gottlob Frege*, 323–340. Oxford: Blackwell.
Gallistel, Randy Charles. 1990. Representations in animal cognition: An introduction. *Cognition* 37. 1–22.
Gallistel, Randy Charles. 1991. *Animal cognition*. Cambridge, MA: MIT Press.
Gallistel, Randy Charles. 2011. Prelinguistic thought. *Language Learning and Development* 7(4). 253–262.

Georgalis, Nicholas. 2015. *Mind, language and subjectivity: Minimal content and the theory of thought*. Cambridge, MA: Routledge.

Glennan, Stuart. 1996. Mechanisms and the nature of causation. *Erkenntnis* 44. 49–71.

Gupta, Anil. 1993. A critique of deflationism. *Philosophical Topics* 21(2). 57–81.

Gupta, Anil. 2003. Deflationism, the problem of representation, and Horwich's use theory of meaning. *Philosophy and Phenomenological Research* 67(3). 654–666.

Hale, Mark & Charles Reiss. 2008. *The phonological enterprise*. Oxford: Oxford University Press.

Halle, Morris. 1983. On distinctive features and their articulatory implementation. *Natural Language & Linguistic Theory* 1(1). 91–105.

Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26. 1–46.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell Publishing.

Higginbotham, James. 1991. Remarks on the metaphysics of linguistics. *Linguistics & Philosophy* 14. 555–566.

Hinzen, Wolfram. 2006. Internalism about truth. *Mind & Society* 5. 139–166.

Hinzen, Wolfram. 2013. Narrow syntax and the language of thought. *Philosophical Psychology* 26(1). 1–23.

Hornstein, Norbert & Paul M. Pietroski. 2009. Basic operations: Minimal syntax-semantics. *Catalan Journal of Linguistics* 8. 113–139.

Horwich, Paul. 1998. *Meaning*. Oxford: Oxford University Press.

Horwich, Paul. 2001. Deflating compositionality. *Ratio (New Series)* 14(4). 369–385.

Horwich, Paul. 2005. *Reflections on meaning*. Oxford: Oxford University Press.

Horwich, Paul. 2008. What's truth got to do with it? *Linguistics & Philosophy* 31. 309–322.

Jackendoff, Ray. 1989. What is a concept, that a person may grasp it? *Mind & Language* 4(1). 68–102.

Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1991. The problem of reality. *Noûs* 25(4). 411–433.

Kadmon, Nirit. 2001. *Formal pragmatics: Semantics, pragmatics, presupposition, and focus*. Oxford: Blackwell Publishing.

Katz, Jerrold & Jerry A. Fodor. 1963. The structure of a semantic theory. *Language* 39(2). 170–210.

Katz, Jerrold & Paul Postal. 1964. *An integrated theory of linguistic description*. Cambridge, MA: MIT Press.
Kemp, Gary. 2012. *Quine versus Davidson: Truth, reference, and meaning*. Oxford: Oxford University Press.
Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Oxford: Blackwell.
King, Jeffrey C. 2007. *The nature and structure of content*. Oxford: Clarendon Press.
King, Jeffrey C. 2018. W(h)ither semantics!(?) *Noûs* 52(4). 772–795.
Kripke, Saul. 1980. *Naming and necessity*. Oxford: Blackwell Publishing.
Lakoff, George. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
Langacker, Ronald. 1987. *Foundations of cognitive grammar, vol. 1*. Chicago: University of Chicago Press.
Lassiter, Daniel. 2008. Semantic externalism, language variation, and sociolinguistic accommodation. *Mind & Language* 23(5). 607–633.
Lepore, Ernest. 1983. What model theoretic semantics cannot do? *Synthese* 54(2). 167–187.
Lepore, Ernest. 1983a. The concept of meaning and its role in understanding language. *Dialectica* 37. 133–139.
Lepore, Ernest & Kirk Ludwig. 2004. Donald Davidson. *Midwest Studies in Philosophy* 28. 309–333.
Lepore, Ernest & Kirk Ludwig. 2005. *Donald Davidson: Meaning, truth, language, and reality*. Oxford: Oxford University Press.
Lepore, Ernest & Kirk Ludwig. 2007. *Donald Davidson's truth-theoretic semantics*. Oxford: Oxford University Press.
Lewis, David. 1969. *Convention: A philosophical study*. Harvard: Harvard University Press.
Lewis, David. 1970. General semantics. *Synthese* 22. 18–67.
Lewis, David. 1975. Languages and language. In Keith Gunderson (ed.), *Language, mind and knowledge*, 3–35. Minnesota: University of Minnesota Press.
Lohndal, Terje & Hiroki Narita. 2009. Internalism as methodology. *Biolinguistics* 3(4). 321–331.
Lohndal, Terje & Paul Pietroski. 2018. Interrogatives, instructions, and I-languages: An I-semantics for questions. In Terje Lohnda (ed.), *Formal grammar: Theory and variation across English and Norwegian*, 319–367. New York: Routledge.

Ludlow, Peter. 2003. Referential semantics for I-languages? In Louise M. Antony & Norbert Hornstein (eds.), *Chomsky and his critics*, 140–161. Oxford: Blackwell Publishing.

Machamer, Peter, Lindley Darden & Carl F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67(1). 1–25.

Majors, Brad & Sarah Sawyer. 2005. The epistemological argument for content externalism. *Philosophical Perspectives* 19(1). 257–280.

Malcolm, Norman. 1972. Thoughtless brutes. *Proceedings and Addresses of the American Philosophical Association* 46. 5–20.

Malpas, Jeff (ed.). 2011. *Dialogues with Davidson: Acting, interpreting, understanding*. Cambridge, MA: MIT Press.

McDowell, John. 1994. *Mind and world*. Cambridge, MA: MIT Press.

McGilvray, James. 2001. Chomsky on the creative aspect of language use and its implications for lexical semantics studies. In Federica Busa & Pierrette Bouillon (eds.), *The language of word meaning*, 5–27. Cambridge: Cambridge University Press.

McGilvray, James. 2002. MOPs: The science of concepts. In Wolfram Hinzen & Hans Rott (eds.), *Belief and meaning: Essays at the interface*, 73–103. Frankfurt: Ontos Verlag.

McGilvray, James. 2005. Meaning and creativity. In James McGilvray (ed.), *The Cambridge Companion to Chomsky*, 204–222. Cambridge: Cambridge University Press.

McGinn, Colin. 1989. *Mental content*. Oxford: Basil Blackwell.

McLaughlin, Brian P. & Michael Tye. 1998. Externalism, twin earth, and self-knowledge. In Crispin Wright, Barry C. Smith & Cynthia Macdonald (eds.), *Knowing our own minds*, 285–320. Oxford: Oxford University Press.

Mendola, Joseph. 2008. *Anti-externalism*. Oxford: Oxford University Press.

Millikan, Ruth Garrett. 1984. *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.

Montague, Richard. 1974. *Formal philosophy: Selected papers of Richard Montague, edited and with an introduction by Richmond Homason*. New Haven, CT: Yale University Press.

Mulhall, Stephen. 1987. Davidson on interpretation and understanding. *The Philosophical Quarterly* 37(148). 319–322.

Murphy, Gregory L. 2002. *The big book of concepts*. Cambridge, MA: MIT Press.

Nuccetelli, Susana. 2003. Introduction. In Susana Nuccetelli (ed.), *New essays on semantic externalism and self-knowledge*, 1–22. Cambridge, MA: MIT Press.

Peacocke, Christopher. 1993. Externalist explanation. *Proceedings of the Aristotelian Society* 93. 203–230.
Peacocke, Christopher. 1994. Content, computation, and externalism. *Mind & Language* 9(3). 303–335.
Pietroski, Paul M. 2008. Minimalist meaning, internalist interpretation. *Biolinguistics* 2(4). 317–341.
Pietroski, Paul M. 2010. Concepts, meanings and truth: First nature, second nature and hard work. *Mind & Language* 25(3). 274–278.
Pietroski, Paul M. 2018. *Conjoining meanings: Semantics without truth values*. Oxford: Oxford University Press.
Portner, Paul H. 2005. *What is meaning? Fundamentals of formal semantics*. Oxford: Blackwell Publishing.
Putnam, Hilary. 1975. The meaning of "meaning". *Minnesota Studies in the Philosophy of Science* 7. 131–193.
Riemer, Nick. 2005. *The semantics of polysemy: Reading meaning in English and Warlpiri*. Berlin: Mouton de Gruyter.
Riemer, Nick. 2010. *Introducing semantics*. Cambridge: Cambridge University Press.
Riemer, Nick (ed.). 2015. *The routledge handbook of semantics*. London: Routledge.
Riemer, Nick. 2019. Cognitive linguistics and the public mind: Idealist doctrines, materialist histories. *Language & Communication* 64. 38–52.
Roth, Martin & Robert Cummins. 2014. Two tales of functional explanation. *Philosophical Psychology* 27(6). 773–788.
Ryle, Gilbert. 1968. A puzzling element in the notion of thinking. In Peter F. Strawson (ed.), *Studies in the philosophy of thought and action*, 7–23. Oxford: Oxford University Press.
Schiffer, Stephen. 2000. Review: Horwich on meaning. *The Philosophical Quarterly* 50(201). 527–536.
Segal, Gabriel M. A. 2000. *A slim book about narrow content*. Cambridge, MA: MIT Press.
Sigurðsson, Halldór Ármann. 2004. Meaningful silence, meaningless sounds. *Linguistic Variation Handbook* 4(1). 235–259.
Slezak, Peter. 1990. Man not a subject for science? *Social Epistemology* 4(4). 327–342.
Slezak, Peter. 2002. Thinking about thinking: Language, thought and introspection. *Language & Communication* 22. 353–373.

Slezak, Peter. 2004. Fodor's guilty passions: Representation as Hume's ideas. In *Proceedings of the 26th annual meeting of the Cognitive Science Society*, 1255–1260. Chicago: Lawrence Erlbaum Associates Publishers.

Slezak, Peter. 2014. Intuitions in the study of language: Syntax and semantics. In Lisa M. Osbeck & Barbara S. Held (eds.), *Rational intuition: Philosophical roots, scientific investigations*, 362–394. Cambridge: Cambridge University Press.

Slezak, Peter. 2018. *Spectator in the Cartesian Theatre: Where theories of mind went wrong since Descartes*. Unpublished manuscript. University of New South Wales.

Smith, Barry C. 1992. Understanding language. *Proceedings of the Aristotelian Society, New Series* 92. 109–141.

Soames, Scott. 1984. Linguistics and psychology. *Linguistics & Philosophy* 7(2). 155–179.

Soames, Scott. 2009. *Philosophical essays: Natural language: What it means and how we use it, vol. 1*. Princeton: Princeton University Press.

Sprouse, Jon & Norbert Hornstein (eds.). 2013. *Experimental syntax and island effects*. Cambridge: Cambridge University Press.

Stich, Stephen P. 1972. Grammar, psychology, and indeterminacy. *The Journal of Philosophy* 69(22). 799–818.

Tarski, Alfred. 1956. The concept of truth in formalized languages. In *Logic, semantics, metamathematics: Papers from 1923 to 1938*, 152–278. Oxford: Oxford University Press.

Thagard, Paul. 2012. *The cognitive science of science: Explanation, discovery, and conceptual change*. Cambridge, MA: MIT Press.

Underhill, James W. 2009. *Humboldt, worldview and language*. Edinburgh: Edinburgh University Press.

Volenec, Veno & Charles Reiss. 2020. Formal generative phonology. *Radical: A Journal of Phonology*.

de Waal, Frans. 2016. *Are we smart enough to know how smart animals are?* New York: W. W Norton & Company.

Wikforss, Åsa. 2008. Semantic externalism and psychological externalism. *Philosophy Compass* 3(1). 158–181.

Yli-Vakkuri, Juhani & John Hawthorne. 2018. *Narrow content*. Oxford: Oxford University Press.

Language and scientific explanation: Where does semantics fit in?

This book discusses the two main construals of the explanatory goals of semantic theories. The first understands semantic theories in terms of a hermeneutic explanatory project, this is often referred to in philosophy of language as externalism. This construal, often implicit, is the standard one in linguistics, but it is far from being the only one. The second construal understands semantic theories in terms of the internalist study of the cognitive psychological mechanisms in virtue of which meaning production and comprehension is made possible. This book compares the internalist and externalist approach to semantics, describing their different motivations and theoretical assumptions. It is argued that a fruitful scientific explanation is one that aims to uncover the underlying mechanisms in virtue of which the observable phenomena are made possible, and that a scientific semantics should be doing just that. If this is the case, then a science of semantics is unlikely to be an externalist one based on hermeneutic and interpretive principles, for reasons having to do with the subject matter and form of externalist theories. Externalist explanations of meaning are concerned with ascription and description of meaning rather than the mechanisms of meaning. They are not concerned with the mental mechanisms in virtue of which humans produce and comprehend meaning. Therefore, despite the claims of leading externalists and formal semanticists to be doing science, externalist explanations are not part of the cognitive psychological explanation of the mechanisms in virtue of which meaning is made possible. Rather, externalist explanations are a hermeneutic explanatory project in that they are an inherently interpretive project. It is argued that semantics construed hermeneutically is nevertheless a valuable explanatory project.

In regard to linguistic science and the way in which linguists think and work, sorting out what the domain of a semantic theory is and what explanatory goals it has are paramount in assessing the success or otherwise of the theory.