1 Introduction

For externalists, semantic content – meaning – consists in the connection between language and the external environment: words’ meaning, roughly speaking, is their reference to the world (see Chapter 1). Classic externalist arguments notwithstanding (Putnam 1975; Burge 1979; Kripke 1980), there clearly must be mechanisms internal to people that make it possible for us to use language to refer. Equally clearly, those mechanisms should be sought in the head. Beyond that point, however, consensus is lacking: not only is the internalism/externalism question the subject of ongoing debate within cognitive science and philosophy (Farkas 2008; Mendola 2008; Egan 1992, 1995), but the details of the various internalist accounts of meaning within linguistics itself are all contested.

Nevertheless, internalist linguistic semantics generally speaking assumes that mental representations called concepts are key parts of the mechanisms linking mind to world. This apparently innocuous assumption is foundational to lexical semantics in particular and is widely shared among researchers who agree on little else (Langacker 1997; Lakoff 1987; Bierwisch and Schreuder 1992; Allan 2001; Jackendoff 2002; Wierzbicka 1999; Sperber and Wilson 1995; Evans and Green 2006; Pietroski 2003). As we will see, however, it represents only one of the theoretical possibilities within a broadly internalist approach to meaning. There is a sense in which, construed internalistically, meaning might not have anything to do with concepts at all, despite current assumptions.

We will shortly clarify just what is involved in the claim that meaning is a matter of concepts. First, however, it is worth asking where the very idea of a concept comes from. In the case of semantics, as elsewhere, paying attention to the history of basic notions will allow us to avoid naive overconfidence in the obviousness of our starting assumptions. An earlier philosophical tradition associated with British Empiricism spoke not of concepts but of “ideas” as the relevant mental entities underlying thought and communication (Locke 1691 [1979]): “concept”, indeed, comes from the Latin verb meaning “comprehend intellectually, imagine, think”. The postulation of “ideas” and later of “concepts” as part of the explanation of human intelligence reflects a particular psychological folk-taxonomy which appears to not
be universal by any means. This taxonomy assumes a categorical division between cognitive and affect-based internal processes—“thought” and “feeling”—with concepts, obviously, falling uniquely on the side of “thought”. However, as documented by Lillard (1998), among others, even a non-exhaustive survey reveals a number of cultures in which “thought” and “feeling” are either not distinguished, or in which “thought” does not exist as a named psychological state at all. Outlandish as this strikes Western ears, the fact that a thinking–feeling contrast cannot always be found outside the bubble of our own first-world cultures reminds us that the theoretical cuts we inherit from our pretheoretical background should not be taken for granted. This is even more the case since, as Waskan (2006) emphasizes, the traditional folk-psychology familiar to speakers of “standard Average European” provides nothing like a satisfying explanation of a host of common mental phenomena, and should not therefore be the object of unconditional theoretical allegiance. (In section 3.3 below we will consider the relation of cognition and emotion in the generation of linguistic meaning.)

For the sake of clarity, two uses of the term “concept” in semantics need to be kept apart. In one sense, “concept” is used simply to refer to whatever psychological structures support meaning. This “weak” use entails no commitments about the nature of the structures in question. We know that whatever is discovered about the nature of human mentality, we will always be able to attribute “concepts” to language users as part of the explanation of their linguistic capacities—just as we will always be able to describe as a “disease” any disorder that compromises an organism’s health, whether it is the result of an infection, hereditary condition, organ damage, psychological condition, or some new kind of physical dysfunction yet to be discovered.

This weak sense of “concept” is not the one of interest here. Because our goal is to clarify the psychological assumptions underlying specific proposals about the nature of semantic content, we have to operate with a substantive definition of “concept” that will stand or fall empirically, and that is not merely a cover-all term for “whatever it is in the head that constitutes meaning”. In section 2 we will delineate the characteristics of concepts as they are mostly assumed in linguistic semantics, and illustrate the way in which concept-attribution functions as an explanation of semantic facts. Section 3 discusses criticisms of this conception. Section 4 considers alternatives to concepts and indicates possible future directions for conceptualist research in semantics.

2 Features of concepts

Word meanings are the clearest examples of concepts. However, language is just one among the many cognitive capacities that concepts are intended to explain. On the standardly assumed picture, concepts are the “units” of rational thought and so at the origin of our most elementary intellectual capacities: unless we possessed the appropriate concepts, we could not identify, sort, name, or imagine the objects of our environment, or reason, plan, or communicate about them. In this section we will explore some of the properties traditionally attributed to concepts. These properties are often not explicit in linguistic semantic theories. Nevertheless, they are presupposed by them, and they are typically the objects of direct theorization in related disciplines within the cognitive sciences.

2.1 Conceptualization and perception

In the most general terms, concepts can be thought of as mental instructions or rules for binding the representations of different properties together. The concept RAINBOW, for instance, binds
together such properties as “coloured”, “in the sky”, “striped”, “curved”, and “occurs after rain” into a single representation. Only if someone possesses the concept rainbow can they identify actual rainbows and reason and communicate about them. In contrast, there is no mental instruction to bind together the properties “coloured”, “observed last week”, and “reminiscent of the art of Anish Kapoor” – even though such properties might well present themselves at the same time in the same context – since there is no unitary concept that contains those properties (although, as with any arbitrary concatenation of properties, we could always create such a concept and stipulate a name for it). In binding properties in this way, concepts establish categories which group individual diverse objects together as instances of the same kind of thing. Individual rainbows will differ in intensity, length, and other phenomenal properties but they will all belong to the category rainbow because they possess the necessary properties for the satisfaction of that concept.

This way of thinking about concepts immediately raises an important issue, namely the question of how perceptual signals – in the case of a rainbow, patterns of retinal stimulation – make their way into the conceptual system. What mechanisms, in other words, govern the “transduction” of perceptual information in such a way that it can interface with thought? Without a detailed answer, accounts of conceptualization have a major explanatory gap, since they fail to specify how the cognitive system “decides” when the various features that constitute a concept are sufficiently present. Consider the concept bottle. The kinds of perceptual information that can trigger this concept are staggeringly diverse: the objects we call “bottles” are of an almost unlimited variety of appearances and can be apprehended from many different angles and under very different visual conditions. Two people talking about the same bottle on a table in front of them will necessarily be observing it from different perspectives and will therefore be in qualitatively different perceptual states. Yet these enormously divergent kinds of perceptual information all somehow get connected to the concept bottle in a reliable enough way for people to reason and communicate about bottles in the way they do.

Just how perceptual signals are converted (“transduced”) into conceptual information is poorly understood, even for simple objects (see e.g. Sudre et al. 2012; Hummel 2013; Peterson and Kimchi 2013; Di Lollo 2012; Palmer 1999). What is clear is that in the absence of a solid explanation of transduction, traditional accounts of concepts remain indeterminate in significant respects. We might imagine, for example, that one of the properties bound in the concept bottle is “has a neck”. But unless we can operationalize recognition of this property in a way that does not tacitly rely on the very cognitive capacities we are trying to understand, we have not achieved any explanation worthy of the name. A genuine solution of the transduction problem would give us a description of the process by which the conceptual system recognizes bottles that could be simulated computationally. Like our minds, a computer would token the concept bottle if and only if it was presented with an actual, neck-possessing bottle – precisely by being able to automatically recognize the presence of the relevant bottle-constituting properties. We are a very long way from being able to do this. When analysing bottle as involving the property “has a neck”, we are tacitly relying on our own conceptual abilities to do all the work for us by identifying what does and does not count as neck-possession. Without a formal account of this property, our “theory” is therefore entirely beholden to the very conceptual ability it purports to explain (see Chapter 8).

This explanatory deficit is particularly serious when concepts are enlisted as part of a theory of human linguistic ability. One of the tasks of a semantic theory is to explain acts of reference. Yet without an account of transduction, we are missing a crucial part of the explanation of how speakers and hearers make the link between what they see and the language they use to refer to it.
2.2 Concepts, information, and invariance

As we have already noted, concepts are traditionally seen as uniquely belonging to the intellectual, rational dimension of human subjectivity. This has a significant consequence on theories of linguistic meaning: since meanings are identified with concepts, they are always taken to be constituents of thoughts, not constituents of feelings or emotions. The validity of this assumption will only be confirmed if it turns out that concepts as traditionally conceived are able to provide a complete account of the human intellectual capacities they are intended to explain and that non-rational factors are consequently explanatorily redundant. Since we do not have anywhere near a complete account of human intellectual capacities, this demonstration has certainly not yet been made.

We cannot capture the rationality of concepts without referring to their informational or representational role. The reason that concepts represent such a key psychological resource in those organisms that possess them is arguably that they allow the representation of the organism’s environment (Fodor 1975, 1981; Pylyshyn 1984). For the purposes of this discussion, \( x \) represents or conveys information about \( y \) if \( x \) is a veridical indicator of \( y \) — if, given \( x \), one can draw many appropriate conclusions about \( y \) that one could not otherwise draw. A passport photograph represents its subject because it licenses many true inferences about it: if the photograph shows a woman’s face, then we are justified in concluding that its subject is a woman. Concepts stand in a relation of accurate representation to, and thus constitute information about, their referents.

The ability to manipulate mental representations is a crucial psychological advantage for any organism fortunate enough to possess it. The correct analysis of this ability forms the essential topic of controversy between externalists and internalists in the philosophy of mind and language (see e.g. Wikforss 2008; Egan 1992, 1995), and is rich in consequences for linguistics. Externalists about concepts believe that how a concept is individuated, and what it represents, is a matter of the way it is connected to its external referents. According to one influential strain of externalist theorizing, \( x \) represents \( y \) if \( x \) is triggered by the presence of \( y \) (Fodor 1987, 1990). The full story is considerably more complicated than this, since we can of course both entertain mental representations (concepts) of things that are not immediately present, and sometimes mistakenly entertain concepts that ought not to have been triggered. For instance, the concept strawberry jam might be triggered by raspberry jam when I mistakenly have the thought “this is strawberry jam” in the presence of raspberry jam. Nevertheless, the idea that meaning consists in an essentially causal or other law-like link between mental symbols and their referents is a key commitment of externalist semantics (see Chapter 1).

For internalists, by contrast, the meaningfulness of mental representations does not derive from any causal or other law-like connection to real-world referents, but from the representations’ own intrinsic properties. For example, proponents of one species of internalism, the computational theory of mind (Pylyshyn 1984), hold that it is the formal, computational properties of the way that mental symbols are coded that underlies their ability to convey information about the world. Consider for example the concept fire. This concept must be of such a kind as to license the inference if this is fire, then it is hot. On an internalist account, this could either be done because the concept fire contains the concept hot, as is assumed in most decompositional theories of conceptual content in linguistics (see Chapter 12), or because the conceptual system licenses the inference from fire to hot, as in so-called inferential role theories (for both stories, see the discussion in Fodor 2008). What makes these stories internalist is that neither depends on any intrinsic connection between the concept fire...
and actual instances of fire. Certainly, speakers use the word fire to refer to actual fire and, as it happens, their concept fire does reflect the actual properties of fire itself. But on both the decompositional and the inferential role account, these representational properties are the result of purely formal characteristics of the underlying concept: the fact that one concept is contained in another, or the fact that a particular chain of inference is marked as valid. This is exactly analogous to the way in which the purely formal arrangement of zeros and ones in strings of software code support that code’s ability to symbolize the functions visible to the user of a computer application.

Internalists and externalists therefore both place central importance on concepts’ representational or informational character. For externalists, representation is constituted by concepts’ causal or other law-like connection to the actual, objective aspects of the world they pick out. For internalists, representation occurs because concepts intrinsically symbolize those aspects without having any necessary causal or other relation to them. In both cases, the concept is objectively correlated with its referent – either by being caused by or otherwise nomologically linked to it, or by symbolizing it in such a way that the concept supports veridical inferences about its referent.

In both cases, the properties of concepts reflect the way things actually are: it would be unthinkable to propose that our fire concept functioned in such a way as to license the inference if this is fire, then it is a solid. Concepts are, and must be, beholden to the objective state of the actual world, because only in that case can we deploy them to successfully reason about and plan our actions. If the information that fire is hot isn’t made available in my conceptual structure, then I won’t be able to plan my interactions with fire accurately: I may, for instance, think that I can touch it safely, or that I cannot use it to cook food. Clearly, there must be a basic fit between the conceptual system and the environment its possessors must survive in.

An important consequence of the representational character of concepts is that they are taken to be invariant – in all essential respects – from one individual to another, and within the one individual over time. Since concepts reflect the actual way the world objectively is, they must, like the world itself, be the same for everyone. Since the body’s physical structure determines a similar array of affordances (interactional possibilities), it is plausible to imagine that the psychological structures that govern action in healthy humans must be essentially similar, at least at the level of grain on which they control action. Whatever the differences in how two different agents do so, they must both represent fire as non-solid and (other things being equal) hot. The environment is the same for everyone, so the conceptualizations with which we control our interactions with it must also ultimately be identical.

What goes for non-linguistic cognition also goes for language. Speakers can only coordinate their linguistic action on the world successfully if they share a similar enough way of using words to refer to aspects of the environment, and if the factual knowledge they have of referents is, in essential respects, shared. Different individuals must not only have congruent internal knowledge of fire (the thing); they must use the word “fire” in highly similar patterns of reference and inference if it is to serve as a useful tool of coordinated action. The assertion the house is on fire will only serve its purpose if the thoughts it produces in the hearer are sufficiently similar to the thoughts that prompted it in the speaker. The possibilities of successful, coordinated action in the world require, in other words, an identity in language-world relations between individuals (and, for similar reasons, within the same individual over time).

A subsidiary question raised by our received picture of conceptualization concerns the extent to which it is possible to distinguish any distinctively linguistic information within the general stock of concepts with which we think. Words, it is standardly accepted, label
(configurations of) concepts. But many investigators have thought that there is something wrong about declaring that the meaning of a given word – say, rainbow – includes all the conceptual information that we have stocked in long-term memory about the word’s referents. Everyone has had different experiences of different rainbows, and will therefore access slightly different sets of concepts. But it seems wrong to say that we do not still share a single meaning for the word “rainbow”. So perhaps it is possible to separate out an essential core of information which all speakers must share in order to count as having the linguistic meaning rainbow, as distinct from the world knowledge we have about rainbows? A now rather old-fashioned way of describing this putative purely linguistic information is as dictionary meaning, in contrast to the broader stock of encyclopaedic concepts associated with the referent. If such dictionary meaning exists, no one has yet managed to define it to the satisfaction of competent experts – either in the case of rainbow or in that of any other word (see the Introduction to the present volume and Chapter 12 for further remarks).

On the other hand, if it is impossible to make a principled distinction between “dictionary” and “encyclopaedic” information, then we must relax the standard definition of “communication” as the exchange of the same thoughts from speaker to hearer. If your and my concepts of “rainbow” are slightly different, then we can no longer claim to be sharing them perfectly when we use the word rainbow to communicate. This relaxation may appear innocent, but it is rich in consequences: once we admit that the conceptualizations underlying a single word may not fully coincide between members of the same speech community, we are no longer entitled to assume that apparent identities of language use reflect the same underlying conceptualizations. You and I may use rainbow to refer to the same kind of thing, but perhaps we do so in virtue of slightly different underlying conceptualizations – a possibility that there are other, more general reasons to take seriously.

2.3 Explanatory autonomy

The last feature of concepts we will address is not one that we know they possess: it is a potential feature that they may or may not turn out to have. Nevertheless, it is worth exploring since it concerns the place of meaning as a phenomenon within the science of the mind, and reveals some fundamental characteristics of semantic theorizing. The feature is explanatory autonomy. The postulation of concepts is justified by the hypothesis that our rational capacities are open to analysis on a distinctive, autonomous cognitive level of explanation – that is, a level fundamentally different from that of electrochemical brain processes on the one hand, and that of the common-sense, folk-psychological categories we use to talk about meaning and the mind informally on the other (on levels in psychological explanation, see Marr 1982). A commitment to the explanatory autonomy of concepts involves acknowledging that the facts that we characterize in conceptual terms on the cognitive level will not prove to be reducible to (or to supervene on) facts of any other kind. As it happens, many linguists apparently assume the opposite: they believe that the language-related phenomena presently characterized using the tools of theoretical linguistics will in the future yield to an account couched in quite different terms – those of neurophysiology. Linguistics will, as a result, be entirely superseded, and therefore decidedly not explanatorily autonomous. However, it is not at all clear that such a reduction is possible. We will explore this issue in the following paragraphs.

It is occasionally the case in science that an existing theoretical description can be entirely superseded by one of a greater degree of detail and power on a lower level. In this case we speak of the first theory being reduced to the second (Hooker 1981; Churchland 1986). Many
facts about the behaviour of light, such as its propagation, reflection, and refraction, can be explained by classical optics (for details see chapter 7 of Churchland 1986). However, it turns out that all these effects completely depend on facts about electromagnetic radiation. After the advent of Maxwell’s theory of electromagnetism, optics need no longer be invoked, since everything we can explain using the theoretical tools of optics can also be explained electromagnetically. Furthermore – and crucially – the theory of electromagnetism relates to microstructural properties of the physical world and therefore reveals more fundamental determinants of the facts previously described in optical terms. Classical optics is therefore said to be reduced to electromagnetism. (Another example of reduction is that of thermodynamics to statistical mechanics.)

There are, however, many explanatory situations where no reduction is possible. This is regularly the case with the so-called special sciences (Fodor 1974); that is, those sciences that only apply to some particular aspect of nature – as opposed to physics, which applies to everything. Geology, neuroscience, and organic chemistry are all special sciences. Linguistics – if it is a science at all (see Introduction) – is clearly a special science, since it only applies to language. To appreciate why reduction isn’t usually feasible for the special sciences, we need to recognize the role that generalization plays in scientific explanation. Science works by identifying certain general classes of phenomena that obey distinctive regularities, which are stated in the “laws” of the science in question. Philosophers refer to these general classes of phenomena that figure in laws as natural kinds. Planets are natural kinds for the purposes of Kepler’s laws of planetary motion, genes are natural kinds for Mendel’s laws of inheritance, and certain types of liquid are natural kinds for Poiseuille’s equation (which describes Newtonian liquid-flow through pipes).

The possibility of reduction turns on the question of whether the natural kinds that figure in higher-level theories can also be identified in the lower-level theories to which they are to be reduced. Whether this is possible in science is often a matter of debate. There are, however, many cases where it is obvious that any such reduction, if attempted, would be unlikely to succeed. To illustrate this, let us consider the prospects for reducing a future theory of human behaviour to physics. We will confine ourselves to a very narrow subset of human behaviour – that manifested on tennis courts. To play tennis, participants must consciously follow the rules of the game. We can therefore say an enormous amount about what happens on tennis courts by referring to the explicit rules of tennis. These rules can be interpreted as a partial theory of tennis-court behaviour, since they allow us to state quite robust regularities: the server only has two chances to hit the ball over the net; if the ball bounces twice on the court then it is served again, and certain adjustments are made to the scores of the players; if the ball lands outside the line of the court then either the server serves again or the game is over, and so on. These rules certainly don’t exhaustively describe how people behave on tennis courts – they don’t, for example, have anything to say about when a player will cough, or trip, or wipe their brow, or about the myriad activities that occur on tennis courts other than games of tennis. They are also subject to all kinds of conditions: for instance, if the server is injured after a first serve, there won’t be a second one. Nevertheless, they express something essential to the game of tennis. No explanation of what is happening on a tennis court could avoid appealing to them.

Note, however, that this aspect of a future theory of human behaviour could in no way be reduced to the laws of physics. There are a large number of physically different actions that constitute, for example, holding a racquet: this will differ in any number of physical parameters (the hand used, the precise configuration and force of the grip, the physical differences between different players’ hands). Holding the racquet is a natural kind for the purposes
of the rules of tennis – which require that the racquet be held horizontally in one hand, not vertically in both, as one holds a cricket bat – but it certainly isn’t for those of physics: there is no single physical description, in terms of the fundamental constituents of matter, that we could possibly advance which would capture all and only those events that constitute holding a tennis racquet.

As argued by Fodor and others, psychological abilities, including language, are like this. It seems not to be possible, at this stage at least, to localize semantic representation to any particular brain region (Hinzen and Poeppel 2011), even though the angular gyrus of the left inferior parietal lobe appears to be particularly implicated (Binder et al. 2009).

Another way of putting this is to say that psychological states have the property of multiple realizability (Funkhouser 2007). This means that there are a variety of different brain states which instantiate any one of them. There may be no unique neurophysiological signature for the mental state one is in when, for instance, understanding the word spoon: the brain mechanisms corresponding to this psychological state differ from individual to individual, and within one individual over time. Whether psychological states are in fact multiply realizable is currently controversial (Bechtel and Mundale 1999; Aizawa and Gillett 2009).

If they are, we have a strong argument for the explanatory autonomy of the conceptual level of explanation. Once we have identified a particular semantic phenomenon – say, the fact that a particular verb means “hear” – we can draw various conclusions about other aspects of its semantics – for instance, we can conclude that it may develop an extra sense meaning “touch”, but not one meaning “see” (Viberg 1984) – but we cannot – at the moment, and perhaps ever – reduce it to any kind of neurophysiological fact. If this situation persists, semantic facts as characterized on the cognitive level will retain explanatory autonomy. If it does not, then semantics will eventually melt away into neurophysiology, gloriously martyring any remaining linguistic semanticists on the altar of scientific progress.

3 Critiques of internalist semantics

In this section we discuss some of the serious critiques of internalist semantics that any proponent of concepts or other psychological supports of meaning must address.

3.1 Two traditional pseudo-critiques

First, however, we can dismiss two pseudo-critiques, albeit well-known ones. First, a traditional objection to the conceptualist identification of meaning is that it makes meanings private mental entities which cannot be shared. Since communication is thought of as the transfer of the same thoughts or concepts from speaker to hearer, the hypothesis that meanings are essentially private might seem to require us to jettison our view of communication. However, as long as we assume that there is some overlap in the mental content expressed by words, we can retain the traditional view of communication while upholding a conceptualist theory of meaning. To identify meanings with concepts internal to each language-user does not exclude the possibility that there is a type-identity between concepts from one user to another, just as there is a type-identity between different individuals’ kidneys. (Type-identity is the kind of identity that holds between two instances of the same category. In the previous sentence, the third word and the third last word are type-identical because they are both occurrences – “tokens” – of the word “the”. Similarly, your concept red and my concept red are also type-identical, even though they are different tokens or instances of the concept.) Concepts can be private and shared because they are, by hypothesis, type-identical from one speaker to another.
Second, John Stuart Mill criticized the idea that meanings might correspond to mental entities on the grounds that speakers routinely presuppose that words refer to actual objects. “When I say that fire causes heat,” he asked (1875: 98), “do I mean that my idea of fire causes my idea of heat?” “No,” he answered: “I mean that the natural phenomenon, fire, causes the natural phenomenon, heat” (see Adams 2003 for a contemporary restatement of this critique directed against Jackendoff). Mill is right: speakers obviously do not take their language to refer to the world of ideas. However, this truism does not constitute any argument against internalism. Mill’s critique rests on a classic error. Of course the speaker means that the natural phenomenon, fire, causes heat – but it is the meaning of the word fire that, for internalists, explains the speaker’s referential behaviour; that meaning itself is not the object of any predication, as Mill mistakenly seems to think. The situation in other domains is exactly equivalent. The statement France is beautiful does not mean that the six-letter word France is beautiful, any more than it means that the meaning of the word France is beautiful: rather, it predicates beauty of France, the actual country. Mill’s critique only works by conflating the linguistic viewpoint of the speaker and the metalinguistic viewpoint of the theorist. Clearly, these two must be kept separate, at the price of major confusion.

3.2 Explanatory inertia

Many critics have drawn attention to the minimal explanatory force of conceptualist explanations (Kamp and Reyle 1993; Givón 1995). Concepts are the internal structures we take to instantiate meaning. The attribution of a meaning to an expression allows us to capture a generalization about its linguistic behaviour by identifying the regularities of reference and inference in which it participates, but, once hypothesized, the meaning does not in itself hold any further explanatory power. To simply relabel meanings as “concepts” is not therefore an explanatory advance: as noted by Murphy (2011: 395):

in order to have a cognitive theory of something, it is not sufficient to add the word cognitive to the name of one’s theory. Empirical evidence from cognitive psychology and psycholinguistics should inform an account that is supposed to be based on psychological principles.

However, to say that spoon means “hand-held implement with handle and bowl for use with liquids”, and to flesh out the concept spoon in equivalent ways, is to do no more than state a generalization about the way the word spoon is used without bringing in any of the empirical evidence Murphy rightly requires. Generalizing about patterns of use is, of course, a necessary first step towards an explanation, but it is not yet an explanation in itself, since we have not said anything about the underlying structures in virtue of which this behaviour is manifested. All we have done is describe a generalization about the word’s use, attributed it to the psychology of the language-user, and called this description a “meaning” or a “concept”. There is a striking analogy here with the classic “dormitive virtue” explanation for the soporific power of opium. Before the chemical agent responsible for this effect – morphine – had been identified, a standard explanation of opium’s power to induce sleep was that it possessed a “dormitive virtue” – clearly not an explanation of any kind, but simply a relabelling of the sleep-inducing property in question (see Riemer 2013 for details).

What makes this objection possible is the lack of constraints facing analysts in hypothesizing the existence of concepts. Linguists have felt entitled to engage in quite straightforward
kinds of armchair, notional analysis when making hypotheses about the conceptual structures underlying meaning. The conceptual structures hypothesized by most semanticists ultimately look a lot like ordinary dictionary definitions. This is something of a worry, since the scientific study of meaning should do more than simply relabel as concepts the descriptions lexicographers give of word-meaning. Lexicography is, after all, a branch of applied, not theoretical linguistics: our suspicions should be aroused if there is a strong commonality between the two. Perhaps the conceptual structures of interest to semantics are more complex, messy, and cross-cutting than the one-to-one word-concept model that still dominates most semanticists’ thinking (apart, of course, from cases of polysemy and synonymy, which are, on the traditional picture, precisely of interest since they are departures from this basic case; see Chapter 13). Cummins (2000: 131) draws a comparison with the “almost irresistible temptation in biology to believe that the morphological traits of importance and interest to us must correspond to our genes in some neat way”.

3.3 Explanatory priority

Classical cognitive science, linguistics included, makes two important assumptions about concepts, each of which is open to question. First, it assumes that concepts in the strong sense introduced above are always involved in the mental operations underlying language, and that they are the principal mental entity of importance in the study of the mind. Let’s call this the assumption of explanatory priority. It may well be, however, that a variety of other kinds of mental process underlie our intelligent abilities, language included. Take for example the conjunction “or”. The natural hypothesis is that “or” corresponds to an underlying concept – that, in other words, the “language of thought” contains a concept for disjunction. However, there are other, non-conceptual ways in which the cognitive system could represent disjunction (see Waskan 2006: 152). For instance, entertaining the thought “either it will rain or it will snow” could amount to entertaining the separate thoughts “it will rain” and “it will snow”, accompanied by an appropriate modal attitude to them – as for example when one holds two objects – different bottles of wine, say – in each hand, weighing up which to buy. This latter situation does not involve any explicit representation of an “or” operator, yet the two objects are related disjunctively, since either one or the other will be bought. This kind of attitudinal alternative to explicit conceptual representation has been explored in the context of simulation theories of conceptualization (Barsalou 1999; 2008).

Second, the received view of conceptualization presupposes that concepts are essentially static: they represent recurrent states of the cognizer which stand out as fixed islands of regularity in the surrounding flow of mental processes, themselves best understood as points of transition between distinct conceptual states. For Aristotle and Aristotelians like Augustine and Locke, nouns, for instance, are meaningful because

the speaker stops his process of thinking and the mind of the hearer acquiesces.

(Aristotle, On interpretation 16b III; italics added)

In this way of looking at things, it’s only by “freezing” the mental process that meaning is created. As Aquinas put it in the thirteenth century, the mind of a hearer “comes to rest” when they understand the meaning of a word (De Int. I.5.16–17; Aquinas 1962). For this reason, the concepts involved in understanding linguistic meaning can be characterized independently of this wider mental flow, using the various analytical frameworks developed by
semanticists to do so: Jackendovian or Wierzbickian primitives (Jackendoff 2002; Goddard and Wierzbicka 2002), Idealized Cognitive Models (Lakoff 1987), event-structure decompositions (see Chapter 22), and so on. In each of these formalisms, it is possible to represent the underlying conceptual content expressed by a lexeme in a static theoretical structure; these static structures recur in different combinations depending on the propositions being expressed. Various currents of research in cognitive science and the philosophy of mind have taken issue with the assumption of the static nature of conceptualization (Peirce 1868; Rumelhart et al. 1986; O’Regan and Noe 2001; Port and Van Gelder 1995; Van Gelder 1998; Barsalou 2008), on the grounds that “natural cognition happens in real time” (Van Gelder 1998: 622). For Van Gelder, cognitive processes are best thought of as “always ongoing, not starting anywhere and not finishing anywhere” (1998: 621), and the goal of the sciences of the mind “is not to map an input at one time to an output at some later time, but to constantly maintain appropriate change” (ibid.).

Correspondingly, many modern theories of linguistic behaviour (Langacker 1997) stress the extent to which meaning must be conceived of as a dynamic process. However, since all of them, even the most revisionist (Jasperson et al. 1994), still recognize a fixed, invariant core of meaning carried by expressions, there is something of a lack of fit between the machinery of their analysis and the accompanying theorizing. As discussed in Chapter 29 of this volume, studies of semantic processing (Sanford 2002) suggest that the mental representations involved in on-line language understanding are actually much less static than this traditional picture assumes, with the degree to which fully consolidated semantic structures are activated depending on a range of variables including, among other things, attention. Since other areas of language structure are known to display variation dependent on factors like these (Coupland 2001; Eckert 2000; Schilling-Estes 2007), there is no a priori reason why semantic structure should be different. We should expect that the kinds of meaning expressed by lexical items will vary as a function of variables like the degree of attention and planning on the part of the speaker, the mode (spoken or written), the register, and perhaps others.

A limiting case is, in fact, plausible in which no distinctly linguistic semantic information needs to be activated for understanding to be achieved. This is where the context itself supplies all the information that the conceptualizer needs to represent the intended meaning. Consider the passengers on a plane hearing the announcement fasten your seatbelts. On traditional models of sentence comprehension, the meaning of seatbelt would be processed exhaustively, with the word’s full semantic representation being activated to allow the hearer to decide which aspects of the current context satisfy it. Since the seatbelts satisfy the specifications of the meaning of seatbelt – say, “safety restraint secured at waist” – the hearer identifies the appropriate referent by realizing that the strap in front of her is a safety restraint secured at the waist.

In that context, however, none of this information is relevant. The hearer does not need to specifically represent the information that the strap on her seat is a waist-secured safety restraint; all she needs to realize is that it is the only object capable of being fastened in the current circumstances. There is therefore no need for the hearer to instantiate any of the information that appears in a conceptual definition of the word seatbelt. The structure of the environment itself means that the word seatbelt need not trigger any form of consolidated representation for the hearer to respond appropriately in the context in which it is uttered. The hearer can “understand” without having to decode the meaning of the word seatbelt, because the context only makes available a single reasonable response to the speaker’s request. For further details, see Riemer (2013).
3.4 Expressivity

The explanatory priority of conceptual explanation is also called into question by the possibility that non-cognitive factors play a more central role in cognition than has been acknowledged. In particular, emotion has traditionally been excluded from any deep or essential role in our picture of the operation of human psychology. As suggested at the start of this chapter, we can once again attribute a role to pretheoretical understandings in determining the direction of research on this point. As Schwarz-Friesel (2007: 92) notes, emotion episodes are often viewed in post-Enlightenment thinking as temporary disruptions of a permanent rational background – a “cold and neutral state of intellectual perception”, in James’ words (1884: 193) – which represents the default state of human subjectivity. This goes against the culturally widespread assumption that people are always experiencing some emotion or other (see e.g. Wierzbicka 1999: 17–18).

The secondary role attributed to emotion generally is reflected in the linguistic presupposition that the principal genus of meaning is descriptive – that is, conceptual or cognitive. (Indeed, it’s worth recalling the fact that, according to Aquinas, Aristotle’s De Interpretatione, arguably the founding document of European semantics, had been excluded by Andronicus from the Aristotelian canon on the grounds that it contained an assimilation of the cognitive and non-cognitive domains that, for Andronicus, confirmed its inauthenticity (see Couillaud and Couillaud 2004: 17–18).) A small class of the vocabulary is considered to be expressive, which means that the psychological states with which it is associated are emotions or feelings, not thoughts or concepts (Potts 2007). The clearest examples of expressive meaning are exclamations (damn!, yuck!, shit!, etc.), though the case has been made at different times that many other items are also expressive (Stevenson 1937; see Chapter 27). It is of course recognized that descriptive vocabulary can have emotion-related aspects, but they are taken to be accessory and accounted for as connotations without ever receiving significant study. Overall, then, expressive meaning is assumed to be numerically and theoretically marginal in the lexicon, and all existing semantic theories take descriptive, i.e. conceptual, meaning as basic.

There are at least two reasons to be suspicious about the claim that cognition is essentially concept-driven and that as a result meaning is mostly descriptive. First is the undeniable fact that emotion constitutes an intimate part of the subjective experience of language. Expressions’ subjective, emotional “loading” is often just as salient phenomenologically as any propositional, informational “content”. If we are sometimes aware of broadly conceptual characteristics of words as we speak (of, for instance, the fact that a logarithm is a particular kind of “arithmetic function”), we are just as, if not more often, aware of their personal emotional qualities (the feeling of dread inspired by the word logarithm in high school, perhaps). It is remarkable that this fact has been recognized for so long without being the object of serious theorizing. In the eighteenth century, Berkeley suggested that non-cognitive factors outweigh the importance of cognitive ones in the economy of speech:

Besides, the communicating of ideas marked by words is not the chief and only end of language, as is commonly supposed. There are other ends, as the raising of some passion, the exciting to, or deterring from an action, the putting of the mind in some particular disposition; to which the former is in many cases barely subservient, and sometimes entirely omitted, when these can be obtained without it, as I think doth not infrequently happen in the familiar use of language. I entreat the reader to reflect within himself, and see if it doth not often happen either in hearing or reading a discourse, that the passions
of fear, love, hatred, admiration, disdain, and the like arise, immediately in his mind upon the perception of certain words, without any ideas coming between.

(Principles of Human Knowledge, Introduction §20)

For their part, Vigliocco et al. (2009: 221) suggest that “emotion . . . may play a crucial role in the representation and processing of abstract concepts”. As a result, any theory in which meanings are identified with conscious or unconscious mental entities (like concepts) needs a good reason to exclude emotions from the characterization of these entities, since the emotional dimension seems to be on an equal footing with propositional/conceptual meaning in the makeup of the phenomenologically accessible states by which meaning is experienced. This is even more the case as the role of non-conceptual factors in cognition more generally is increasingly emphasized (Gunther 2003).

The second reason to doubt the soundness of the claim that cognition is wholly concept-driven is that the very basis for the distinction between concepts and emotions is itself unclear. As is frequently acknowledged, we have no good definition of what makes something cognitive or a concept (Hooijmans and Keijzer 2007; Prinz 2004). However, the very act of describing meaning in one of the recognized metalanguages of semantic analysis implies that these meanings are conceptual and therefore not emotion-based. The problem here is that anything can be described in conceptual terms: there is no meaning for which some kind of semantic analysis cannot be advanced. This means that the very representational medium used to characterize meaning excludes from the outset a non-conceptual definition of meaning, and thereby discounts the role of a whole dimension of human psychology. Semantic theorizing, as we have already noted, typically proceeds in something of a vacuum of independent psychological evidence of human conceptual capacities: semanticists develop semantic analyses on the basis of linguistic evidence, and then declare that these analyses correspond to stored conceptual mental representations. The latter claim is rarely argued for, and is the one that should excite scepticism.

3.5 Modality

Semantic analysis as traditionally undertaken in linguistics is predicated on the assumption of a “language of thought” – a neutral, language-like mental format or code that supports cognitive processes, like the software codes running a computer (Fodor 1975, 2008; Fortis 1996). This language of thought is often conceived of as the medium in which central mental processes unfold, neutral between the different perceptual channels that input to it. Simulation theories of cognition, however, hold that conceptual information is “couched in representational codes that are specific to our perceptual systems” (Prinz 2002: 119; see Chapter 8 of the present volume). On the basis of neurophysiological (Damasio 1994, 2000) and psychological (Barsalou 1999, 2008) research, simulation theorists argue that there is no single representational system or language of thought which cognition draws on in the planning and control of action, including language, and which justifies the kinds of semantic analysis typical of linguistics. Instead, the psychological architecture supporting reference and other cognitive tasks harnesses the perceptual modality appropriate to particular referents. If this thesis is confirmed, it will carry major implications for the methodology of linguistic semantic research.

3.6 The regress problem

One of the fundamental roles of meaning and hence of concepts in a theory of language is to explain the patterns of reference and inference which comprise use: words have the use they
have in virtue of the meanings they express. Along with syntax and pragmatics, semantics has the role of accounting for these patterns. A further challenge for conceptualism arises from the observation that there is nothing in any traditional conceptual representation which determines a particular pattern of use: any given concept, as the mental structure activated during language use, can support an indefinite range of different and incompatible uses.

This emerges most clearly when, in the context of reference, semanticists speak of representations “matching” or “corresponding” to their referents. English speakers are hypothesized to have a conceptual representation of the meaning of, say, fire, and match this representation to real-world situations of fire to achieve reference. We lack, however, an account of how this matching happens. What principles determine when a representation matches a real-world situation and how are these principles applied? This apparently innocent question opens a Pandora’s box which might cast doubt on the very explanatory viability of explicit representations like concepts in a theory of cognition. If concepts are attributed to cognizers as part of the explanation for their intelligent action, then they also must have the capacity to apply the concept to situations. This capacity in turn requires us to attribute to them a whole new set of principles specifying the way in which this application happens, which themselves then require further principles for their own interpretation, and so on ad infinitum. In a tradition that goes back at least to Wittgenstein (1953), many researchers have worried that this question opens an explanatory regress: the regress problem and ones like it are often advanced as an argument against the classical understanding of cognition as a symbolic, rule-governed process of symbol manipulation (cf. Kripke 1982; Dreyfus 1985, 1992; Block 1990; Searle 1992; Riemer 2005).

4 Future directions

Having surveyed some considerations pro and contra internalist semantics, we can ask what future directions internalist semantic research might take, and what alternatives may exist to a theory that places concepts – in the strong sense – at the centre of the explanation of meaning. Given the complexity and lack of resolution in the areas mentioned in this chapter, there is not a single one that would not benefit from sustained and foundational research. This is even more the case since various currents of enquiry offer completely different accounts of the internal states corresponding to meaning, and therefore constitute a decisive challenge to linguistic semantics. Prime among these would be the statistical approach to many of the phenomena addressed by traditional conceptualist theories, exemplified by Landauer and Dumais (1997). On these views, linguistic phenomena can be theorized without any appeal to internal conceptualist structures. Another important strand of research is constituted by simulation-based theories of cognitive processes (Barsalou 2008; Prinz 2002; see Chapter 8) mentioned in the previous section. Playing these theories off against those that involve concepts in the strong sense must constitute an urgent task for future investigations into meaning (see e.g. Chaterjee 2010).

Beyond that, the challenge posed by broadly phenomenologically inspired theories of semantics also demands the attention of the field. Semanticists have traditionally presupposed that the meanings they are analysing are essentially more principled analogues of dictionary definitions. However, work like that of Cadiot et al. (2006) and Zlatev (2010) leads us to ask whether a mode of representation that is less abstract and arguably closer to the grain of experience might not provide a better way of capturing facts about meaning. There needs to be a dialogue between traditional semantics and the kinds of alternative analysis proposed by researchers like these.
This challenge draws attention to the present lack of agreement on questions of semant
ic analysis. Current linguistic research into meaning is characterized by a profusion of
different theoretical approaches, with little or no effort being paid to questions of theory
evaluation or assessment of rival hypotheses. Of existing semantic theories, it is mainly
the Natural Semantic Metalanguage theory that has been the object of sustained challenge
on the level of its concrete analyses of meaning (see Geeraerts 2010 for references). While
there are plenty of critiques of research in, for instance, Cognitive Semantics, these have
largely been directed at the theory’s high-level background, and despite some notable
controversies like the debate over over, rather little effort has been devoted to assessing
the adequacy of the particular accounts of meaning it furnishes. If – if – semantic research
should aim at consensus, then an immediate priority for the discipline must be a concerted
attempt to compare different theories and to resolve the numerous differences of approach
presently practised. Whether such a project is feasible, however, is far from clear (see the
Introduction).

Among the most urgent other priorities must be clarification of the descriptive/expressive
distinction. This is, as we have seen, a fundamental distinction in semantic analysis, but one
that is currently ill understood. Illuminating the difference between the two genera of mean-
ing would constitute substantial explanatory progress (see Chapter 27).

Further reading
Outlines the role of concepts in a naturalistic theory of cognition.
Riemer, Nick 2013. Conceptualist semantics: explanatory power, scope and uniqueness. Language
Sciences 35: 1–19. Explores alternatives to concepts in semantic analysis and suggests a clarifica-
tion of the expressive/descriptive distinction.
logical research on conceptualization.
plete account of emotional factors in language. Chapter 4 includes an illuminating discussion of
the relation between cognition and emotion in language.

References
Aizawa, K. and C. Gillett 2009. The (multiple) realization of psychological and other properties in the
Aquinas, Thomas 1962. Aristotle: On Interpretation (Jean T. Oesterle, tr.). Milwaukee, WI: Marquette
University Press.
Philosophy of Science 66: 175–207.
University Press.
Binder, J.R., Rutvik H. Desai, William W. Graves and Lisa L. Conant 2009. Where is the semantic sys-
tem? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex
19: 2767–2796.


Nick Riemer


**Related topics**

Chapter 1, (Descriptive) Externalism in semantics; Chapter 5, Cognitive semantics; Chapter 7, Categories, prototypes and exemplars; Chapter 8, Embodiment, simulation and meaning; Chapter 12, Lexical decomposition; Chapter 27, Expressives.