OpenLearn

The Open University

Language and thought: Introducing representation

The Open University

Language and thought: Introducing representation

OpenLearn | Free learning from The Open University

About this free course

This OpenLearn course provides a sample of Level 3 study in Arts and Humanities: http://www.open.ac.uk/courses/find/arts-and-humanities.

This version of the content may include video, images and interactive content that may not be optimised for your device.

You can experience this free course as it was originally designed on OpenLearn, the home of free learning from The Open University: www.open.edu/openlearn/history-the-arts/culture/philosophy/language-and-thought-introducing-representation/content-section-0.

There you'll also be able to track your progress via your activity record, which you can use to demonstrate your learning.

The Open University, Walton Hall, Milton Keynes, MK7 6AA

Intellectual property

Head of Intellectual Property, The Open University

Designed and edited by The Open University

Contents

Introduction

This course introduces key questions about language and thought, such as how can language, which is public and accessible, be used to convey thoughts, which seem hidden from view.

This OpenLearn course provides a sample of Level 3 study in Arts and Humanities.

Learning Outcomes

After studying this course, you should be able to:

- discuss basic philosophical questions concerning language and thought
- understand problems concerning language and thought and discuss them in a philosophical way.

1 Introducing representation

1.1 Introduction

One of the most impressive but puzzling capacities we have is the ability to represent the world around us, both in talking about it among ourselves and in thinking about it as individuals. When someone utters the sentence, 'The German economy is bouncing back', for example, they are able to convey to their audience something about the German economy. Their utterance may be correct or it may be incorrect, but either way it is making a claim about how things are, and in this loose but intuitive sense they are using language to represent the world to someone else. Another example – this time of mental rather than verbal representation – is of someone believing that cinema tickets are half-price on Tuesdays. This is a belief about how things are. Things may be that way or they may not be that way; in either case, the believer is representing them as being that way to herself or to himself – once again in an intuitive sense of 'represent'.

This course provides a short introduction to some of the questions philosophers have asked (especially over the last hundred or so years) about our ability to represent, both in language and in thought.

1.2 Representation and language

Consider some of the many different things we can do with language: express ourselves in metaphor, issue commands, ask questions, fill in crosswords, write shopping lists and diary entries, repeat nursery rhymes by rote, solve logical or arithmetical problems, make promises, tell stories, sign our names, etc. Impressive though it is, this variety in the uses of language is a potential distraction from our main interest, which is in the use of language to represent. It will therefore help if we abstract away from this diversity by focusing on a paradigmatic use of language: the use of words, spoken or written, *to transfer knowledge*. An utterance of 'The German economy is bouncing back' would ordinarily have this purpose.

One reason for focusing in this way is that a significant proportion of what we know comes to us via language. Sharing our thoughts about the world by talking and writing to one another allows us to pool our cognitive resources. Our task will be to come to understand how this pooling process takes place. That is, we will attempt to answer, at least in broad outline, a question posed by the English philosopher John Locke (1632–1704):

> ### Locke's question
>
> How is it that 'the thoughts of men's minds [can] be conveyed from one to another' (Locke [1706] 1997, III.1.2)?

Locke's description of linguistic communication makes it sound like a kind of telepathy, and perhaps that is a useful comparison. By using language we are able to 'read each other's minds' quite effortlessly. Effortless it may be, but linguistic communication is a magnificent human accomplishment, every bit as peculiar as telepathy but more

interesting because it actually occurs. What is it about producing a particular sound or pattern of ink that allows one person, the speaker or writer, to share information about the world with another, the hearer or reader?

An intuitive way of describing what happens when 'the thoughts of men's minds' are 'conveyed from one to another' is that the speaker invites his or her audience to accept that the world *is as it is represented to be* by the speaker's utterance, and the audience takes up this offer. This way of putting it can be elaborated into a simple theory intended to answer Locke's question:

> ### The simple theory of communication
>
> The successful communication of knowledge about the world is possible because speakers are able to produce utterances *with a specific meaning*, and recognition of that meaning by an audience enables them to appreciate what the speaker intends to communicate.

The key phrase is 'specific meaning' of an utterance. This is roughly equivalent to 'how the utterance represents the world as being', whatever that involves.

Saying what it involves is precisely what we must do if we wish to save this simple theory from the charge that it is hopelessly empty. For compare the simple theory with the theory parodied by the seventeenth-century French playwright Molière in *The Imaginary Invalid*. In this play, a doctor offers a spurious account of what gives opium its power to induce sleep in those who ingest it: it has this power, he says, because of its *virtus dormitiva*, i.e. its 'soporific virtues, the tendency of which is to lull the senses to sleep' (Molière [1673] 1879?, 73–4). The point of the parody is that the doctor has merely introduced a fancy-sounding name for the thing that needs to be explained, without having advanced us towards a genuine explanation. His putative explanation comes to this: opium induces sleep because it has the power to induce sleep. The simple theory of communication *as it stands* is little better than the doctor's theory. Like Molière, we can ask whether a term like 'meaning' as it figures in the simple theory (or 'represents' insofar as it plays the same role) is any more than a name for what we are trying to explain. Before we can sign up to the simple theory, we need to know how it differs from the following satirised version of it:

> ### The empty theory of communication
>
> The successful communication of knowledge about the world is possible because speakers are able to produce utterances with communicative powers, and audiences are induced by these powers to appreciate what the speaker intends to communicate.

Of course, if we supplement the simple theory with a substantial, plausible and independent theory of what meaning is – a theory that amounts to more than the claim that meaning is what makes communication possible – all will be well. That is our task in this course, which considers attempts to develop a theory of the meaning of utterances, thereby rescuing the simple theory from the charge that it is empty as an explanation of how 'the thoughts of men's minds [can] be conveyed from one to another'.

1.3 Representation and thought

It would be surprising if the meaning of our utterances turned out not to derive, in part at least, from the thoughts and other mental states that these utterances express. Were that so, language would be failing in one of its main functions. Ordinarily, an utterance of the

sentence, ‘The German economy is bouncing back’, is intended to express *the thought that the German economy is bouncing back*, typically so that the audience will come to adopt this same thought. It is hard to see how this could be so unless the meaning of the utterance did not derive, in part at least, from the representational properties – the ‘content’ as it is often put – of the thoughts and other mental states of the speaker.

Understanding the nature of mental content is taken by many to be equivalent to understanding how – presumably by virtue of possessing a brain, a complex physical organ – humans are able to think about the world around them. How can a state of the brain be *about* the world outside the skull of the person whose brain it is? This is the mental equivalent of Locke's question about language, and equally daunting.

Recent developments in theories of human cognition have added impetus to the search for an answer to this question. Many philosophers and cognitive scientists have been impressed by the explanatory benefits of claiming that mental activity in humans is akin to the operations of a computer. Crudely put, computers operate by transforming symbols within them in a blindingly fast but rule-governed manner. According to advocates of the computational theory of mind, the same is true of us. On most versions of the theory, for a human being to be in a particular mental state is for their brain to contain symbols of a kind of brain language, ‘Mentalese’ as it is usually called. The alleged attractions of thinking of the human brain as populated by symbols of a language come to this: the computational theory of mind promises to explain how rationality is possible in a purely physical entity, as a living human body is assumed to be. Such an explanation has been a dream for many philosophers at least since Hobbes.

Not everyone accepts the analogy of human thinking with the operations of computers, but among those who do, the question arises of what gives the symbols of Mentalese their meaning. How can ‘words’ in a brain be about anything? How can they represent the world outside the skull? The meaning of the symbols of an actual computer – what makes it appropriate to call them ‘symbols’, in fact – derives from the interpretation imposed on them by computer designers and operators. The meaning of words in spoken or written language is also imposed, this time by the people using the language for the purpose of communication. But the source of the meaning of sentences hidden inside the human skull cannot be the interpretation imposed on them by an external interpreter, since there does not seem to be any such interpreter. So anyone who accepts the computational theory of mind is under an obligation to say what gives the symbols in the human brain their meaning. Many are sceptical of the computational theory of mind precisely because it is hard to see how this obligation could ever be discharged.

Discussion of mental representation, then, is often framed in terms of the meaning of inner symbols. But most of the difficulties that arise for those who accept the computational theory of mind also arise for *anyone* who (i) agrees that humans are capable of representing the world around them, but also (ii) wishes to claim that humans are in some sense essentially physical creatures subject to the laws of physics like other objects in the universe, and apt for study using scientific methods. Critics of the materialist world view are keen to stress how hard it is to show how both these assumptions could be true.

1.4 Three characteristic difficulties in discussions of representation

I have hinted that accounting for the nature of representation – whether it be the meaning of utterances or the content of our mental states – is not easy. There are several reasons for this, and it is as well to take note of some of them from the outset.

One is that there seem to be several different senses of 'meaning', 'represents' and related terms like 'stands for', 'being about', 'expresses' – differences that have been glossed over here but will need to be distinguished. Moreover, there are different *kinds* of thing that can ordinarily be said to represent or mean or stand for or express. Sometimes we talk of *people* meaning something, at other times we talk of their *words* meaning something, and often we talk of their *utterances* (i.e. the actions people produce *using* words) as meaning something. Different again is the meaning of a mental state, i.e. what is represented in thought rather than through spoken or written language.

Another source of difficulty is that meaning is not quantifiable in the way that, for example, temperature or humidity are. Generalising from this point, ordinary scientific methods do not seem suited in any straightforward way to assist us in our efforts to understand either what representation is or how it is possible for human beings to represent. This is unfortunate because scientific methods have, without doubt, been extremely effective in other spheres of enquiry.

A further reason for thinking that science and the topic of representation are not well suited to one another has to do with the close connection between representation and the notion of *correctness*. Consider representation as it figures in belief. One way of distinguishing a belief *that cats have kittens* from a belief *that dogs have puppies* is that these two beliefs have different correctness conditions from one another. The first belief is correct if, and only if, cats have kittens. The second belief is correct if, and only if, dogs have puppies. But correctness and incorrectness – being right and being wrong – are normative or evaluative properties. As such, science seems ill-suited to describing them. Physicists and geologists, for example, do not include evaluative notions in their theories: a microscopic particle or a rock formation would never be described as 'correct'. It is true that geologists might evaluate a rock as a 'nice sample', but this is only an indication of its usefulness to the interests of the scientists, perhaps for illustrative purposes. Being 'nice', like being 'correct', is not part of any actual geological theory.

Finally, *represents* differs in strange ways from ordinary relations such as *bumps into* or *is late for*. To illustrate, consider our ability to think and talk about – to represent – things that do not exist. It is not possible to *bump into* things that do not exist. Similarly one can only swim in, read or kick things that actually exist. How is it possible that we can be related to a non-thing? Yet '*x* represents *y*' seems to be just such a relation. Were it not so, we would not be able to think about or talk about Santa Claus or God unless they both existed, and it seems plain that we can think about and talk about both without making any such assumption. For example, one can think about and discuss *whether* they exist. A phenomenon related to our capacity to think and talk about things that do not exist is our capactity to think and talk about one thing without thereby thinking or talking about a second thing, even though that 'second' thing is in fact identical with the first thing. Examples bring out the contrast with ordinary relations. With ordinary relations like *bumps into*, if *x* bears that relation to *y*, and *y* = *z*, then *x* also bears that relation to *z*. If by accident I bump into Mr Jones on the bus, and Mr Jones is identical with Plastic Freddy the performance artist, then it is also true that I have bumped into Plastic Freddy.

Representing, or *thinking/talking about*, is different in this respect. When Oedipus thought longingly about Jocasta without realising that Jocasta was his mother, there is a sense in which he was not thinking longingly about his mother. Representing someone as Jocasta is not the same as representing them as one's mother even when Jocasta *is* one's mother. (The example comes from the ancient Greek tragedy, *Oedipus Rex*. Oedipus unwittingly kills his father so as to marry Jocasta, taking out his own eyes on discovering his mistake.)

1.5 Some useful terminology and a convention

It will be useful to end this section by establishing a simple convention and introducing some terminology.

The convention has already been at work in this chapter, but has yet to be made explicit. It is a convention for marking the difference between *using* a word and *mentioning* it. Italy has a capital city, and the English language contains a word for that city, but the word and the city are distinct entities. When we are talking about the word rather than what the word is about, we are *mentioning* that word. When we want to talk about the city itself, we are very likely to *use* the word and not merely mention it. In oral language, this distinction between using a word and mentioning it is not explicitly marked. Spoken aloud, the first and third sentences below are indistinguishable, as are the second and fourth. In writing, however, there is a helpful convention of surrounding a word with inverted commas whenever it is being mentioned rather than used. Thus:

> ‘Rome’ has four letters. (*correct*)
>
> ‘Rome’ is the capital city of Italy. (*incorrect*)
>
> Rome has four letters. (*incorrect*)
>
> Rome is the capital city of Italy. (*correct*)

The convention of marking the distinction in this way will be adopted throughout this book, and it is used by the authors of the associated readings.

The terminology to be introduced concerns the use of ‘subject’, ‘attitude’ and ‘content’ in discussing mental states. The last of these has appeared already. What a mental state is about (its *content*) can be distinguished from both the person whose mental state it is (the *subject* or alternatively the *agent*) and the kind of mental state it is (the *attitude*). The distinctions are easily conveyed through examples of mental states:

> **Same subject; different attitude; different content**
>
> *John* believes that there are some biscuits in the biscuit tin.
>
> *John* hopes that he can watch television undisturbed.
>
> **Different subject; same attitude; different content**
>
> John *believes* that there are some biscuits left in the biscuit tin.
>
> Mary *believes* that she can watch television undisturbed.
>
> **Different subject; different attitude; same content**
>
> John believes that *there are some biscuits left in the biscuit tin*.
>
> Mary hopes that *there are some biscuits left in the biscuit tin*.

This terminology assumes that every mental state involves three elements: a subject, an attitude and a content.

Contents are often expressed in the so-called ‘that-clause’ (underlined) of sentences used to attribute mental states: ‘So-and-so thinks (or hopes, intends, imagines, expects, decides, etc.) that such-and-such.’ This is true of the six examples above. When a content can be expressed in a that-clause, it is said to be a *propositional* content or, for short, a proposition. Propositional contents are capable of being true or false. But sometimes a content is not, on the surface at least, propositional. Being in love, for example, is a mental state, but one loves a thing, a person usually, rather than a propositional content. Someone can be correctly said to love *someone or something*, but it sounds peculiar to talk of loving *that* such-and-such, where the such-and-such is capable of being true or false. Contents of attitudes like ‘love’ are called *objectual*. Other mental attitudes relate a subject to a content of either kind. Someone can desire a person or a thing, but they can also desire that something be the case.

Differences between the attitudes will not be a large part of our discussion, but it is worth being aware that there are differences. The attitudes are traditionally classified as cognitive (e.g. believing or thinking), affective (e.g. being upset), or conative (e.g. intending, desiring or deciding). All three kinds involve representing, but in slightly different ways. The differences between them turn on the function of the attitude. Very crudely, we might say that the function of believing that *p* is to represent the world as it is, and the function of intending that *p* is to represent the world as it will become through one's own intervention.

Finally, though I will often talk of ‘speakers’ and ‘hearers’, the philosophical point being made always generalises to cover modes of linguistic communication other than oral. In particular, it applies to both sign language and written language.

1.6 Further reading

For an advanced general introduction to the philosophy of language, see Blackburn 1984. Lycan 1996 is pitched at a more accessible level. Pinker 1994 is an informal but informative discussion of the hypothesis that much of our linguistic ability is innate, an important topic that has had to be left out of this course.

2 Is the speaker's mind the source of an utterance's meaning?

2.1 Introduction

The distinction noted in section 1 between the representational properties of a linguistic utterance (its 'meaning') and the representational properties of a mental state (its 'content') gives rise, naturally enough, to the suspicion that one of these might be more fundamental than the other. In this section I will look at a theory, most closely associated with the British philosopher H.P. Grice (1913–88), to the effect that the source of an utterance's meaning is the speaker's mind, i.e. the content of their mental states. On this view, when someone produces an utterance, the meaning of this utterance can ultimately be traced back to the content of their intentions in producing it.

An alternative possible source of an utterance's meaning is the meaning of the words used, conceived of as dependent on the characteristics of the linguistic community the person belongs to and possibly other factors that are 'outside' the mind of the utterer. Both views of the source of an utterance's meaning are appealing in their different ways. So a pair of questions that will stretch across this section is (i) whether the two views are genuinely in competition or merely apparently so, and (ii) if there is genuine competition, which of the views is more successful in locating the source of the meaning of utterances?

2.2 The source of an utterance's meaning: the words used or the speaker's mind?

How are we able to use language to communicate knowledge? Locke's question, introduced in section 1, was recast as the obligation to spell out what 'meaning' amounts to as it figures within a simple theory of communication, repeated here:

> ### The simple theory of communication
>
> The successful communication of knowledge about the world is possible because speakers are able to produce utterances *with a specific meaning*, and recognition of that meaning by an audience enables them to appreciate what the speaker intends to communicate.

This theory is, we saw, only genuinely explanatory if we can supplement it with a non-vacuous statement of what it is for an utterance to have the meaning it does. Our task in this section will be to explore ways of doing this.

There are two different and potentially incompatible approaches to carrying out this task, distinguished according to what they take as the source of an utterance's meaning. According to one view, the source is the meaning of the expressions uttered according to the language to which they belong. If a speaker utters the sentence, 'The German economy is bouncing back', what the utterance means is that the German economy is bouncing back. On this first view, it means this because *that is what these words mean in*

English. If the speaker had unusual ideas about what the words ‘German’ and ‘economy’ mean – for example, that they mean what we ordinarily mean by ‘folding’ and ‘bed’ respectively – this would not (on this first view) change the meaning of the utterance. The speaker would still have said that the German economy is bouncing back, not that the folding bed is bouncing back.

Many find this thought intuitively appealing. To others, it seems unnecessarily prescriptive. Why, they ask, can we not use language in ways of our own choosing? If someone wants to use ‘jealous of’ to mean *envious of* rather than *possessive of* (its traditional or ‘proper’ meaning), then why shouldn't they? This second view treats the speaker's mental states as the source of the meaning of the utterances they produce. Supporters of the first view often respond to the liberal, speaker-centred perspective by pointing to the disastrous consequences of treating the actual speaker rather than the words the speaker uses as the arbiter of an utterance's meaning. They are fond of alluding to Humpty Dumpty (as he figures in Lewis Carroll's *Alice through the looking-glass* rather than as he figures in the nursery rhyme) to make their point. Below is a passage from the book in which Humpty proclaims his right to mean whatever he wishes by the words he uses.

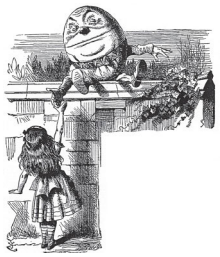


Figure 1 Alice and Humpty from Lewis Carroll, ‘Through the looking-glass and what Alice found there’, illustrations by John Tenniel, first published by Macmillan & Co., London, 1872

> ‘As I was saying [said Humpty Dumpty], that *seems* to be done right – though I haven't time to look it over thoroughly just now – and that shows that there are three hundred and sixty-four days when you might get un-birthday presents –’
>
> ‘Certainly,’ said Alice.
>
> ‘And only *one* for birthday presents, you know. There's glory for you!’
>
> ‘I don't know what you mean by “glory”,’ Alice said.
>
> Humpty Dumpty smiled contemptuously. ‘Of course you don't – till I tell you. I meant “there's a nice knock-down argument for you!”’
>
> ‘But “glory” doesn't mean “a nice knock-down argument”,’ Alice objected.
>
> ‘When I use a word,’ Humpty Dumpty said, in rather a scornful tone, ‘it means just what I choose it to mean – neither more nor less.’
>
> ‘The question is,’ said Alice, ‘whether you *can* make words mean so many different things.’
>
> ‘The question is,’ said Humpty Dumpty, ‘which is to be master – that's all.’
>
> Alice was too much puzzled to say anything; so after a minute Humpty Dumpty began again. ‘They've a temper, some of them – particularly verbs: they're the

> proudest – adjectives you can do anything with, but not verbs – however, *I* can manage the whole lot of them! Impenetrability! That's what I say!'
>
> 'Would you tell me please,' said Alice, 'what that means?'
>
> 'Now you talk like a reasonable child,' said Humpty Dumpty, looking very much pleased. 'I meant by "impenetrability" that we've had enough of that subject, and it would be just as well if you'd mention what you mean to do next, as I suppose you don't mean to stop here all the rest of your life.'
>
> 'That's a great deal to make one word mean,' Alice said in a thoughtful tone.
>
> 'When I make a word do a lot of work like that,' said Humpty Dumpty, 'I always pay it extra.'
>
> 'Oh!' said Alice.
>
> (Carroll 1893, 113–15)

We can distil from this dialogue a claim – call it *Humpty's thesis* – regarding the extent to which what we mean is under our own control rather than dependent on the meaning of the words used, where this latter is independent of our wishes:

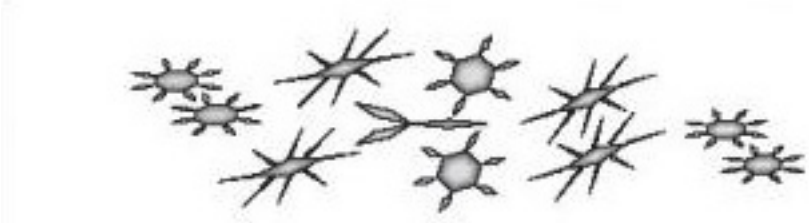
Humpty's thesis

What we mean when we utter a word or sentence is under our own control; we can mean whatever we want and choose.

Though it is difficult to see exactly why, Humpty's thesis seems deeply misguided. Grice's theory, to which I now turn, sees the mental states of the speaker as the primary source of the meaning possessed by the utterances the speaker produces. He even claims that the relevant mental states are the speaker's *intentions*. This sounds as if he is just providing a 'Humpty-Dumpty' theory according to which what we mean is up to us. Whether this unflattering comparison can be made to stick is something to consider after seeing the details of his influential theory.

Activity 1

We will be examining a position according to which the meaning of utterances is inherited from the content of the speaker's mental states. Having just read Humpty's embarrassingly extreme statement of this view, you may be deeply unsympathetic to that idea. So, to counterbalance your sympathies, consider a reason to be suspicious of views at the opposite extreme, which treat the meaning of utterances as entirely independent of mental states. Suppose some rocks are found arranged in a pattern on the dusty surface of Mars, as below:



The rock pattern could be the equivalent in some alien language of a piece of clumsy but meaningful handwriting. Alternatively it could be a meaningless cluster of tiny fallen meteorites, any discernable pattern in it being a random accident. What would make

you treat this pattern of rocks as a meaningful utterance – not a verbal utterance obviously, but an utterance in the same sense that a written letter is an utterance?

Arguably, in order to be meaningful the cluster would have to be judged as having been produced by some intelligence, perhaps with the intention of communicating with another intelligent being, or with us, or with God, or (as in a diary or a doodle or an arithmetical calculation) with itself. Considered merely as an unintended and accidental pattern in the dust on the Martian surface, it has no meaning whatsoever. If by outlandish chance a cluster of meteorites fell into a pattern that spelled out the English sentence 'Lo, Earthlings!', it would still not mean anything – though in that case we could be excused for incorrectly reading meaning into a meaningless event. This is a reason of sorts for suspecting that the meaning of utterances is dependent on the psychological states, perhaps even the intentions, of the producer of the utterance.

2.3 Grice on natural and non-natural meaning

Ironically, the word 'meaning' has many different meanings. There are four occurrences of 'mean' (or 'meaning' or 'meant', etc.), italicised, in the following paragraph:

> Roberto's instructor had been *mean* to put it so bluntly, but she was probably correct that his short legs *meant* he would never be a great dancer. He turned into the narrow alleyway, *meaning* to take a shortcut home. His life no longer seemed to have any *meaning*.

Here is the paragraph again, with each of the four occurrences of 'mean', 'meant' and 'meaning' replaced with an appropriate synonym.

> Roberto's instructor had been *cruel* to put it so bluntly, but she was probably correct that his short legs *were bound to result in* his never being a great dancer. He turned into the narrow alleyway, *intending* to take a shortcut home. His life no longer seemed to have any *purpose*.

There is no good reason to demand of a theory of meaning that it give an account of every kind of meaning. It *may* turn out that there is some underlying unity to the uses of 'mean' (or 'means', 'meant', 'meaning') on display here, but that is not something on which we should insist.

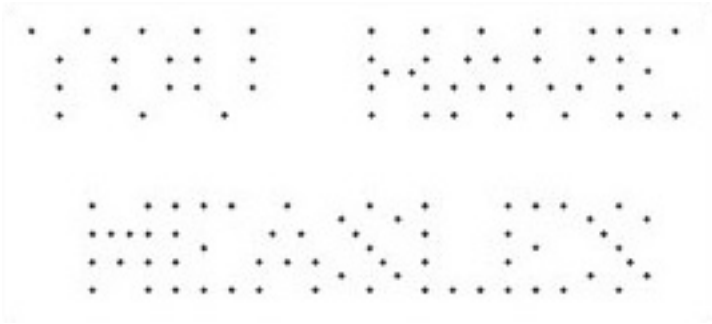
In view of the plethora of meanings of 'meaning', Grice proposes to set aside those that are not his immediate concern and to focus on understanding the nature of those that are. One kind of meaning that is left over is defined by Grice in terms of the speaker's intentions. This is a good candidate for being the kind of meaning *we* are interested in, i.e. the meaning utterances have that accounts for the role they play in communication. But before coming to what he says about this kind of meaning, we need to see which kinds of meaning he sets to one side as confusing distractions.

Grice begins his paper (Reading 1) by making an important distinction between two species of meaning that it is particularly easy to confuse, which he labels *natural meaning* and *non-natural meaning*. The kind of meaning he later defines in terms of speakers' intentions is non-natural meaning. Natural meaning is the kind being attributed in claim (a):

(a) 'Those spots on your face mean *you have measles*.

This claim could be true only if the italicised sub-sentence is true, i.e. only if you really do have measles. If you had spots but you didn't have measles, the spots would not mean that you had measles. They would have to have some other source. Contrast this with claim (b), which uses 'mean' in its non-natural sense:

(b) 'The spots in the arrangement below mean *you have measles*.'



This whole assertion would be true even if the italicised sub-sentence was false. That is, an assertion of (b) would still be correct even if you did not have measles. Generalising, the difference between the two kinds of meaning is this: it is consistent with something's having non-natural meaning that what it non-naturally means is false; but it is not consistent with something's having natural meaning that what it naturally means is false. (In the paper, Grice notes other differences, but this is the main one.)

Grice's purpose in making this distinction is merely to avoid confusion. He sets natural meaning (or 'meaning$_n$' as he calls it) to one side and moves on to developing a theory of non-natural meaning ('meaning$_{nn}$'), the kind he is more interested in. His partiality has to do with the fact that examples of meaning that involve language are typically cases of meaning$_{nn}$, and no one has so far come up with a good theory of meaning$_{nn}$. Natural meaning, by contrast, does not really have much to do with the meaning of words or utterances, and is in any case relatively non-mysterious. '*X* means$_n$ that *p*' can be understood as a substitute for one or other of various simple phrases, including:

- *X* causes it to be the case that *p*
- *X* is conclusive evidence for *p*
- *X* is not possible unless *p* is true
- *X* entails that *p*

Natural meaning is mentioned later in his paper, but only in order to clear up potential confusions, not because Grice is especially interested in it.

Activity 2

Read Part I of Grice's paper, 'Meaning'. The original paper is not actually divided into parts; they are my addition (indicated by the square brackets) to facilitate guided reading. Grice distinguishes natural from non-natural meaning. He also rejects an attempt (by C.L. Stevenson) to define non-natural meaning in terms of natural meaning, prior to offering his own theory of non-natural meaning in the rest of the paper.

Click to view Part I 'Meaning' by H.P. Grice:
'Natural meaning distinguished from non-natural meaning'

Which of the following claims about meaning are most plausibly interpreted as claims about natural meaning, and which are most plausibly interpreted as claims about non-natural meaning?

(i) John is sneezing. This means he has a sinus infection.

(ii) The French sentence, ‘Pierre aime les chats’, means that Pierre likes cats.

(iii) In saying what he did, John meant that he would be late.

(iv) Failure to bring an accurate map with him meant that John would be late.

(i) Natural. If John has no sinus infection, his sneezing could not possibly mean that he had a sinus infection. If it means anything, it would have to mean something else, e.g. that there is pepper in the air.

(ii) Non-natural. The sentence would mean what it does even if Pierre hates cats.

(iii) Non-natural. John's utterance (whatever it was – perhaps it was ‘I will be late’ or ‘start without me’) would have had this meaning even if he in fact ends up arriving on time.

(iv) Natural. Suppose John arrived on time. This would lead us to reject the claim that his failure to bring an accurate map meant that he would be late.

The distinction between natural and non-natural meaning is, Grice notes, not always clear cut. The same entity can sometimes have both natural and non-natural meaning. Here is an illustration:

> The canyon-dweller's shout of ‘here comes an echo’ meant that we would hear an echo a few seconds later.

This is plausible on both readings of ‘meant’. But in most cases the distinction seems to be reasonably easy to apply. Let us move on, then, to Grice's theory of non-natural meaning. (Henceforth in this section, ‘meaning’ should be read as ‘non-natural meaning’ unless specified otherwise.)

2.4 The meaning of expressions versus the meaning of individual utterances

I drew a contrast at the beginning of the chapter between those approaches to the meaning of utterances that look to the meaning of the words used, and those approaches that look instead to the content of the mental or psychological states of speakers. Grice belongs to the second camp. He aims to show that the meaning of an expression (e.g. a word or a sentence) is derivative, definable in terms of how that expression is typically used in meaningful utterances. The meaning of individual utterances is, he concludes, more fundamental than the meaning of expressions. More fundamental than both, though, are the contents of the speakers' minds, and in particular the intentions that give rise to the production of utterances.

With this agenda in mind, Grice draws a distinction within the category of non-natural meaning. What an utterance means ‘timelessly’ is tied to the meaning of the words and sentences used in making it. The word ‘timeless’ is used to allude to the fact that expressions can be used over and over again with the same meaning, but we can follow

standard practice and talk of *expression* meaning – the meaning of the sentences and words that occur in utterances. But more fundamental than expression meaning, in Grice's view, is another kind of non-natural meaning: the meaning of an individual utterance (or, as he sometimes puts it, of an expression as it is used 'on a specific occasion'). This more fundamental notion can be defined (Grice thinks) in terms of the psychological states of the speaker. Grice's working hypothesis, then, is that both kinds of meaning ultimately have their source in the content of psychological states, something he seeks to show in two steps:

> *Step One*: give a definition of the meaning of single, isolated utterances couched entirely in terms of *what the speaker intends to bring about*.
>
> *Step Two*: give a definition of expression meaning ('timeless meaning') couched entirely in terms of the definition of the meaning of single isolated utterances given in Step One.

Activity 3

Read Part II of Grice's paper. This is where he sets out the agenda just described. The crucial phrase is 'this might reasonably be expected to help us with'. Grice is asserting the priority of the meaning of individual utterances (or of expressions as used by someone on 'specific occasions' as he puts it) over the meaning of expressions (used 'timelessly'). He will go on to define the former in terms of the intentions of the utterer (what I have called his Step One) and the latter in terms of the former (Step Two).

Click to view Part II of 'Meaning' by H.P. Grice:
'A strategy for elucidating non-natural meaning'

2.5 Why intentions?

Most of the rest of Grice's paper is dedicated to spelling out a way of identifying the meaning of an individual utterance 'on an occasion' with the content of the utterer's intentions (Step One). The hard task he faces is to say what type of intention creates meaning. If someone shouts 'I saw a film last night' extremely loudly at their brother with the intention of making this brother fall off his bike, this 'utterance' (if that is the right word) does not thereby mean *fall off your bike, brother*. So Grice must distinguish between different kinds of intentions. To appreciate his efforts it will help to understand what is driving Grice's choice of the utterer's *intention*, rather than some other type of psychological state, as the source of an utterance's meaning. The motivation for this choice is important but is left largely implicit in this early groundbreaking paper.

Many of the examples Grice gives of events with (non-natural) meaning are non-linguistic. Ringing the bell on a bus is a case in point. The existence of non-linguistic but meaningful acts leads Grice to the view that it is *as acts* that linguistic acts have meaning – their being linguistic is something of a side issue. Grice even goes so far as to stretch ordinary usage of the term 'utterance' to include *any* event that has (non-natural) meaning. But what is it about meaningful acts, linguistic or otherwise, that gives them the meaning they have? One thing all acts seem to have in common is that they are performed with an intention, even if that intention is not always fulfilled. Grice's hunch is that it is the intention behind the production of a meaningful act, linguistic or non-linguistic, that gives it its meaning.

He attempts to vindicate this hunch by spelling out in detail the precise form the intention must take. Merely being intended does not make an act meaningful since all acts are intended but not all acts are meaningful. When a tree surgeon saws off a branch, this does not normally have any meaning. There must be something special about the intention behind utterances (i.e. meaningful acts) that sets them apart from acts that lack meaning. But what is that special something?

2.6 Which intentions?

Grice makes three attempts to answer this last question. The second builds on the first; the third, which he proposes to adopt, builds on the second. In the next three activities, you will be asked to extract these attempts in turn, and appreciate the alleged shortcomings of the first two.

Activity 4

Read Part III of 'Meaning', in which Grice makes a first attempt to specify which form an intention must have if the resulting act is to be meaningful. He then quickly dismisses the attempt. What is the definition, and why does he find it wanting?

Click to view Part III of 'Meaning' by H.P. Grice: 'A first proposal'

According to Grice's first suggestion, an utterance means whatever it is that the utterer is trying to get his or her audience to believe. (He is talking only about assertions, setting aside questions and orders.) Call this first definition 'Grice 1'.

> *Grice 1*: A specific utterance *U* means that *p* if, and only if, in performing *U*, the utterer intends an audience to come to believe that *p*.

Grice quickly dismisses this as insufficient. Doing something with the intention merely of getting one's audience to believe that *p* does not amount to meaning that *p*. He gives a simple example to show this. The example is a non-linguistic one, which fits with his hunch that the source of an individual utterance's meaning is not, ultimately, the words used.

The example involves someone, call them *A*, secretly leaving *B*'s handkerchief at a murder scene with the intention of getting the detective to believe that *B* is the murderer. Intuitively, we would not really say of A's act that it *means* that *B* is the murderer. (It certainly doesn't mean_{nn} this, anyway, and this is the only kind of meaning Grice is interested in. But in fact it does not really seem to mean_{n} this either. At most, the detective may be led to *think* it means_{n} that *B* is the murderer.)

A common reaction to this counter-example is that leaving a handkerchief at a crime scene is not verbal, so is not an utterance, so is irrelevant. But Grice's use of 'utterance' is meant to be stipulative and artificial. Using 'utterance' in this self-consciously broad way is a reminder of the fact that verbal utterances are not the only kind of meaningful act. (Think of nodding, or miming breaststroke behind a boss's back to mean that a co-worker should not cave in.) But in any case Grice's counter-example could easily have been a verbal one, where by 'verbal' is meant something like 'involving use of the voice'. Suppose that, for some reason, *C* wants *D*, a monolingual English-speaker, to believe that *D*'s house is haunted by the ghost of the Russian émigré who used to live in it. At night, from inside a

cupboard in *D*'s bedroom, *C* produces Russian-sounding nonsense using a guide to Russian enunciation. We would not say that *C*'s utterances mean that *D*'s house is haunted as Grice 1 requires. They do not mean anything, not even in Russian.

Grice uses the failings of this first definition to develop a better one.

Activity 5

Read Part IV of 'Meaning', in which Grice modifies his earlier definition. He also argues that the modified version still falls short. What is the modified version?

Click to view Part IV of 'Meaning' by H.P. Grice: 'A second proposal'

The new element, required for meaning but missing in the handkerchief scenario, is *openness of intent*. *A* placed *B*'s handkerchief secretly, since if the detective was aware that *A* had put it down, he would not have been led to suspect *B*. If we add an openness requirement to Grice 1, it should rule out this example and others like it. Grice incorporates this new element into his second attempt at a definition, which we can express as follows:

> *Grice 2*: A specific utterance *U* means that *p* if, and only if, in performing it, the utterer intends:
>
> (a) that an audience will come to believe that *p*, and
>
> (b) that this audience will recognise intention (a).

The new clause, (b), is not met in the handkerchief case or in the haunted-cupboard case. This lends support to this formulation of the Gricean proposal.

But even Grice 2 is inadequate. In one of several counter-examples, Grice imagines himself supplying Mr *X* with a photograph of Mr *Y* 'showing undue familiarity to Mrs *X*'. (Note: Grice was writing before the manipulation of photographs became commonplace.) He supplies it with the intention that Mr *X* will come to believe that 'there is something between Mr *Y* and Mrs *X*'. If we take '*U*' to be the act of supplying the photo to Mr *X*, and 'p' to be that there is something between Mr *Y* and Mrs *X*, clause (a) is satisfied. Suppose moreover that Grice, in this imaginary scenario, wishes his intention to be recognised. Perhaps he wishes this because Mr *X* is powerful and supportive of those who are loyal to him. So rather than posting the photo anonymously, Grice hands it to Mr *X* in person. This means that clause (b) is also satisfied. But we would not really want to say that Grice's act *means* that Mr *Y* and Mrs *X* are joined in some illicit union. We might be tempted to say that this is what the photo itself means, but that is at best meaning$_n$, not meaning$_{nn}$, the kind we are interested in. The photographic image is evidence for the existence of an illicit union, in the same way that red spots on a person's face are evidence of measles; if there were no illicit union then the claim that this is what it meant would have to be withdrawn, and this is the mark of meaning$_n$. (This is where the danger of confusion between the two kinds of meaning is at its greatest, which is why Grice was keen to make the distinction explicit early on.)

Activity 6

Read Part V of 'Meaning'. In this passage, Grice adds a further clause to get to what we can call Grice 3. What is the new requirement?

Click to view Part V of 'Meaning' by H.P. Grice: 'A third proposal'

The problem with Grice 2, illustrated in the photograph case, is used by Grice to draw a moral. Even though clause (a) and clause (b) are both met in the photo scenario, there is *no connection between* the intention required by clause (a) and the intention required by clause (b). In particular, the prospects for success of intention (b), i.e. the intention to be recognised as having intention (a), is inessential to the prospects for success of intention (a). Grice (in the photograph scenario) could easily have succeeded in intention (a) without succeeding in intention (b). The photo, so to speak, tells its own story without Grice's (b)-intention playing any essential role. To see this, suppose for contrast that the 'undue familiarity' were represented by a drawing by Grice of matchstick figures rather than by a photo. Such a drawing could serve as a warning, or as a strange fantasy, or as a reminder of what is happily not the case, or any number of purposes other than as the recording of a witnessed event. What would settle which of these possible messages was the drawing's actual meaning would turn on Grice's (b)-intention – his intention to be recognised by Mr *X* as producing the drawing for a particular purpose. That, thinks Grice, is why drawings have meaning$_{nn}$ and photos have mere meaning$_{n}$. This thought feeds into his final theory of the meaning of individual utterances. What would rule out the photo case, thinks Grice, is the requirement of a connection between intentions (a) and (b).

> *Grice 3*: A specific utterance *U* means that *p* if, and only if, in performing it, the utterer intends:
>
> (a) that an audience will come to believe that *p*, and
>
> (b) that this audience will recognise intention (a), and
>
> (c) that the recognition in (b) will cause the belief in (a).

This final version is complex, but that is unsurprising given the implausibility of the crude version of the approach advanced by Humpty, according to whom what the speaker means is just what the speaker chooses to mean. We will have to decide shortly whether Grice's theory is a genuine improvement over Humpty's or merely hides the same basic misconceptions behind increased complexity.

2.7 Expression meaning as defined by Grice

Recall Step Two in the Gricean agenda: to define the meaning of expressions in terms of the meaning of individual utterances. Carrying out this strategy successfully would lend strong support to the thought that it is the mental states of speakers, rather than the meaning of expressions, that are the ultimate source of utterances' meaning.

Activity 7

Read the rest of Grice's paper (Part VI). Pay particular attention to (1)–(3) on pp. 185–6, which state his overall theory. (1) and (2) are a summary of his theory of the meaning of individual utterances. (Be warned: Grice is not kind to his readers here. For example, his theory of the meaning of utterances is given twice over. He gives a definition of what it is for a speaker ('*A*') to mean something by producing an utterance. He then gives a definition of what it is for an expression ('*x*'), as it is used by a speaker

to produce an utterance 'on a specific occasion', to mean something. But according to him these notions are equivalent. That is why, in an effort to simplify his position, I represented him as offering a theory of the meaning of just one thing, the utterance itself, *U*.) (3) is his theory of the meaning an expression ('*x*') has 'timelessly', i.e. independently of the meaning it has in the context of any particular utterance. (This is what I have been calling 'the meaning of an expression'.)

Before coming to (1)–(3), Grice extends his account to cover utterances that are not intended to convey information, for example questions or orders; after (1)–(3) he attempts to deal with some potential problems with his overall account.

What is expression meaning, according to Grice (i.e. how does he define the meaning an expression, *x*, has 'timelessly')?

Click to view Part VI of 'Meaning' by H.P. Grice:
'Timeless meaning defined in terms of meaning on an occasion'

The meaning of an expression is defined by Grice in a way that takes for granted the meaning of individual utterances made using it, since this notion is already given in his (1)–(2). He claims (though he is not very clear at this point) that an expression – a sentence, in effect – means that *p* within a loosely circumscribed linguistic community if, and only if, members of that community tend to use that expression in utterances that mean that *p*. So, for example, the expression 'the train is late again' means what it does among the people of Britain, America and so on, because they tend to use it in individual utterances that mean the train is late again.

The theory of the meaning of expressions in (3) on page 186 comes to this:

> Sentence *s* means that *p* in the language of a specific population if and only if most utterances of *s* by members of that population mean that *p*.

Unfortunately, there seems to be an important lacuna in Grice's account of expression meaning: at best it defines sentence meaning, since it says nothing about word meaning. Words are typically uttered meaningfully only in combination with other words, typically within sentences. (There are a few exceptions, such as 'Fire!' or 'Help!') But they can appear in many *different* sentences, associated with *a huge variety of distinct* communicative intentions. So even if Grice is right to claim that a sentence means what it does because of a regularity in the meaning of the utterances made using it, the meaning of a word cannot be defined in terms of such a regularity.

But there is a problem with this definition even as it applies to just sentences. Some sentences have never been uttered before. Consider:

> It is easier to dry-clean umbrellas that have been soaked in giraffe saliva than it is to inhale freshly plucked Namibian goose down.

If Grice is right that a sentence's meaning is a matter of how it is regularly used, then this novel sentence, never used (I presume) before I wrote it down just now, ought to be meaningless. But that seems not to be the case. Its meaning is unusual, but it has one.

This second problem seems to have its source in the first. If we had a theory of word meaning, we could attempt to see how the meaning of a novel sentence is built up from the meaning of the words it contains. Grice's predicament is not quite as desperate as it appears. We will see a theory of word meaning that is in keeping with Grice's ambitions towards the end of this section.

2.8 The Gricean Programme

Before considering any further potential criticisms of Grice's position, let us step back and consider his wider importance to philosophy: his contribution to what is often called *The Gricean Programme*. Grice himself was not really a Gricean in this sense, since he was not committed to all elements of the programme that bears his name. But Grice's influence has been as great as it has in part because of the way in which his ideas have been co-opted into this broader programme.

The goal of the programme is to show that all species of representation (including the various kinds of meaning associated with language, as well as mental content) are reducible to the natural sciences. That is to say, Griceans aim to restate all uses of phrases like ‘means’, ‘represents’, ‘stands for’ and so on using only phrases drawn from the language of science. As it is sometimes put, the aim is to ‘naturalise’ the notion of representation. Grice (according to Griceans) shows us how to reduce the meaning associated with language to the content of mental states, and if the content of mental states can be reduced in turn to the natural sciences, the programme will have succeeded. One motive behind this reductionist enterprise is to give science a comprehensiveness it would lack if the notion of representation, very useful in its various guises, fell outside the main fold of science; another is to confer legitimacy on the notion itself by integrating it with the rest of science, much as biology, chemistry and physics have been increasingly integrated with one another.

The structure of the Gricean Programme is reflected in Figure 2, which I discuss below, starting at the top and following the arrows in order. The diagram looks complex, but most of it is already familiar.

We have seen already that Grice makes a plausible case for the existence of an ambiguity in the word ‘meaning’ between natural meaning and non-natural meaning (arrows 1 and 2). Natural meaning, we saw, can be thought of as a causal relation or necessary or lawlike connection (section 2.3), where these are notions that sit happily with the natural sciences (arrow 3). If we want to know whether a particular cloud configuration means$_n$ that rain is on its way, our best bet is to ask a meteorologist.

Non-natural meaning – the kind still left mysterious after natural meaning is set aside as definable in terms of causal connections, etc. – can be distinguished into two sorts (arrows 4 and 5). This is the familiar distinction between the meaning of expressions and the meaning of individual utterances. We have just seen how Grice proposes to define the former in terms of the latter, and this is captured in arrow 6. (The outstanding matter of accounting for word meaning as opposed to sentence meaning will be returned to at the end of this section.)

We saw that Grice uses the term ‘utterance’ in a broad and slightly artificial way to mean any action with non-natural meaning. Utterances in this sense can include not only linguistic actions but many other kinds of non-linguistic action, such as the ringing of a bell on a bus. The distinction between linguistic and non-linguistic meaningful acts is, he thinks, misleading if it is taken to be a fundamental one, since all meaningful acts have something in common: the source of their meaning is the agent's intention. In the diagram, this is why the downward flow bifurcates temporarily into linguistic and non-linguistic utterances (arrows 7 and 8) but immediately comes together into a single point (arrows 9 and 10), the intention of the utterer. The content of intentions is but one kind of mental content (arrow 11).

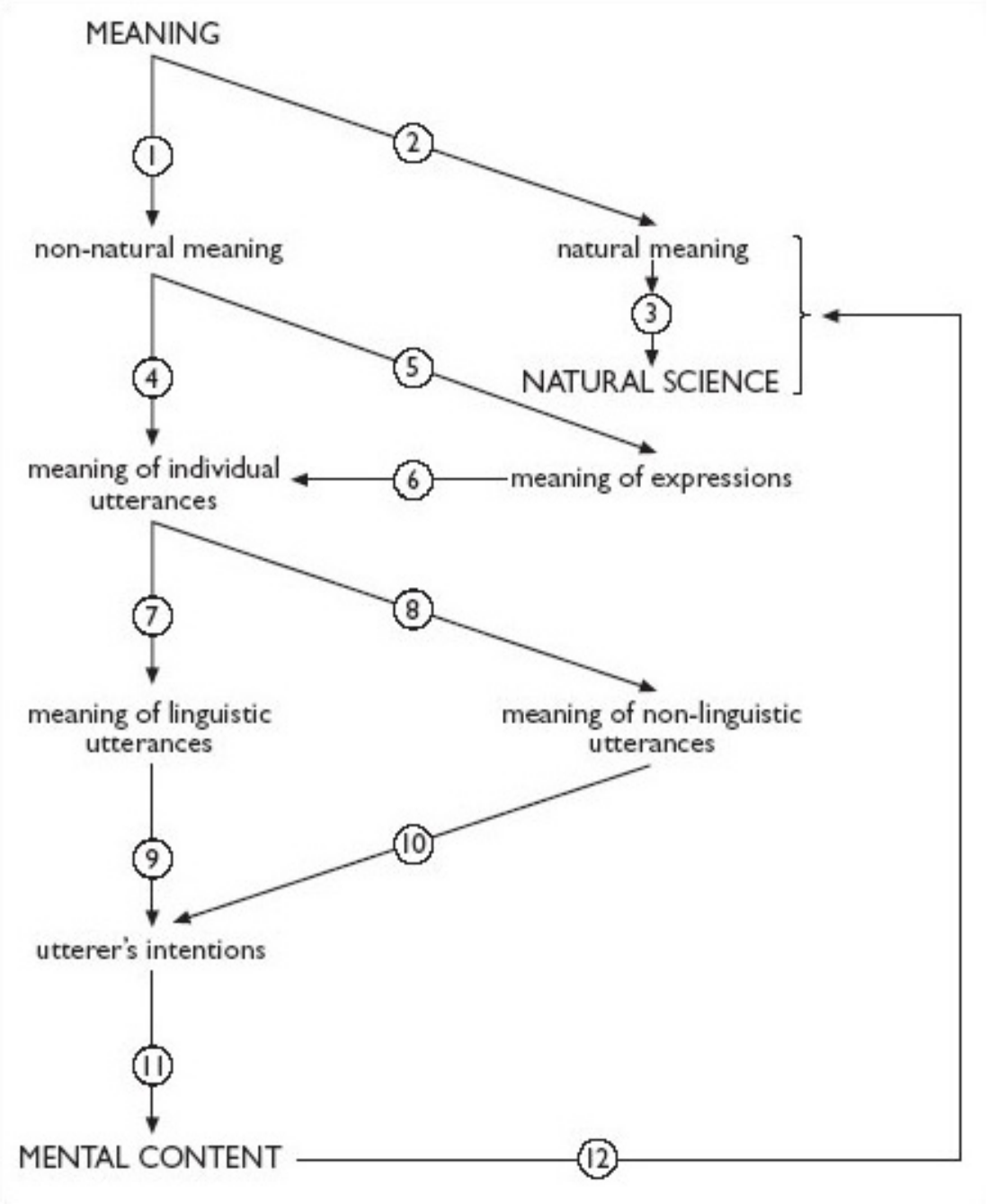


Figure 2 The Gricean Programme. Arrows represent the direction of flow in the analysis of representational notions (meaning and mental content) into, ultimately, properties belonging to the natural sciences

Everything recounted so far is more or less resonant with Grice's position as it is manifested in his 1957 paper. But if the Gricean Programme stopped here it would be incomplete. For what, in scientific terms, is mental content? The final piece in the jigsaw would be a theory of mental content (including the content of intentions) in terms that are familiar from other areas of science (arrow 12). This final part of the programme – following the course of arrow 12 – is not covered in this course. But perhaps worth noting is that one approach to the content of mental states involves trying to define it in terms of what Grice has called *natural meaning* (hence the bracket in the diagram at the end of arrow 12).

2.9 How successful is Grice's theory of the meaning of utterances?

I turn now to difficulties for Grice's account of the meaning of utterances, beginning with a concern over his methodology. By focusing on examples, real or imagined, Grice attempts to draw out our intuitions and so lead us, as he has been led, to Grice 3. But perhaps our intuitions are wildly inaccurate, or wildly irrelevant. We need to check that Grice's notion of meaning, mined out of his and our intuitions, delivers what we were after when we turned to him for a theory of the meaning of utterances.

In fact, the definition Grice arrives at does seem to be well suited to our needs once we consider the model of communication we get when Grice's final statement of what it is for an individual utterance to mean what it does (Grice 3) is plugged into the simple theory of communication. Think of a communicative predicament. *A* is with *B* in the jungle, and sees a leopard nearby, hidden from *B*'s view. *A* wants *B* to appreciate the leopard's presence but is not able to point it out. *A* must therefore communicate the leopard's presence to *B*. By uttering an appropriate sentence, *A* intends that *B* will come to believe there is a leopard nearby. Moreover, it is through *B*'s recognition of this intention that *A* intends this to come about. When *B* hears *A*'s utterance and recognises it as having been performed with the intention that *B* believes there is a leopard nearby, *B* obliges by forming just this belief, exactly as anticipated by *A*. This elaborated version of the simple theory seems to go a long way towards capturing what is going on in the communication of knowledge. And unlike the unelaborated version, it does not rest on an unanalysed notion of meaning that can be compared unflatteringly to the *virtus dormitiva* in Molière's satire (see section 1.2). So, Grice's theory is a reasonable candidate for being what we want from a theory of the meaning of utterances. But before celebrating, we need to consider other potential weaknesses.

One is the extreme structural complexity of the intentions we would need to have for Grice 3 to be correct. This structural complexity seems to be at odds with the ease with which we speak. If Grice is right, every time we open our mouths or pick up our pens to communicate, we form a complex triple-parted intention. Is this plausible as a description of our psychology as speakers? Ordinary speakers would be hard pressed to give a verbal statement of the content of the intentions that Grice is saying they form at high speed in everyday conversation. Such complex intentions also seem to be far more than we are capable of recognising in others in the real-life communication of knowledge.

Whether we have these complex intentions is a difficult topic. In his paper Grice does say something in his own defence (pp. 186–7), denying that the intentions he is describing are ‘explicitly formulated linguistic (or quasi-linguistic) intentions’. The fact that these communicative intentions – unlike the intention to, say, hail a taxi – cannot be verbalised does not show that they do not exist, he is suggesting. But the problem seems to go deeper than he realises, even if he is right to insist that there can be unverbalisable intentions. The communicative intentions Grice posits are not only unverbalisable, they are unconscious. When we speak we are aware of an intention to produce a particular sentence or to communicate in some loose way, but we are not aware of ourselves as having complex triple-parted intentions. (Or at least I am not. You are invited to decide for yourself the next time you communicate.) But the idea of an unconscious intention has struck many as an odd one. Grice could reply that intentions *can be* unconscious. He could draw an analogy with Freud's claim that we have unconscious drives, or with the way in which learning to type or to play a musical instrument is deliberate and conscious

at first but gradually becomes unconscious and fluent. But many are left with a nagging doubt that this is an account of utterance meaning that fails to do justice to our actual experience of using and understanding language.

A different worry turns on the adequacy of the content of the intentions rather than on their structural complexity. The development of Grice 3 out of Grice 2 and, before that, Grice 1 involved *adding* clauses. But there are some reasons to worry that the first clause, (a), demands too much, not too little. There are plenty of meaningful utterances in which there is no intention to get an audience to believe anything, for the simple reason that there is no audience.

Activity 8

Are communicative intentions of the kind set out in Grice 3 (or Grice 2 or Grice 1 for that matter) genuinely necessary for an utterance to be meaningful? Think about this question in relation to:

(i) writing a diary;

(ii) talking to a dog – not as in 'Fetch!' but as in, 'Oh Rover, how I wish I were just a dog like you, with no worries';

(iii) Hamlet's 'To be or not to be' soliloquy.

In each of these cases, Grice would have to demonstrate that, contrary to appearances, there is an intended audience. Perhaps he could reason as follows. (i) The intended audience of a diary is often the wider world, since not all diaries are intended to be secret. For the secret ones, Grice could say that the audience is just a future version of oneself, as it is with shopping lists. For (ii) the audience might be some constructed fantasy, a dog that understands English. (i) and (ii) could each also be thought of as involving an element of soliloquy like (iii). Soliloquy might be understood as an attempt to communicate with a god, or with posterity in the abstract.

2.9.1 The Humpty objection

The Humpty objection to the Gricean position can be formulated as the charge that Grice is too much like Humpty for his own good. Humpty, we saw, seems to be advocating the following thesis:

> What we mean when we utter a word or sentence is under our own control; we can mean whatever we want and choose.

The charge against Grice can be formulated in a two-premise argument:

Premise One: Humpty's thesis is false.

Premise Two: Grice's position entails Humpty's thesis.

Conclusion: Grice's position is false.

To assess this charge, let us first consider Premise One. There does seem to be something terribly wrong with Humpty's thesis, but before getting to what is wrong with it, it is as well to appreciate why it could be attractive, to us and not merely to a self-important egg. Words do not have their meaning intrinsically. What they mean is *arbitrary*. The word 'dog' might have meant what the word 'cat' in fact means, and vice versa. Equally, nothing

about the letters ‘i’, ‘m’, ‘p’, ‘e’, ‘n’, ‘e’, ‘t’, ‘r’, ‘a’, ‘b’, ‘i’, ‘l’, ‘i’, ‘t’, and ‘y’ requires that, taken together, they must mean what they do – roughly, a lack of susceptibility to physical infiltration. They could easily have been combined to mean something quite different. What is wrong with assuming that this arbitrariness is resolved by a decision of the utterer to mean one thing rather than something else? Nothing at all, Humpty would no doubt insist (see Alice and Humpty's conversation).

He could support his case by pointing to the benign practice, common in philosophy as in any sphere of life, of deciding to introduce a new term with a specific meaning laid down in a definition, or else of taking an old term and stipulating that it will be used in a novel way, or with one from among several established usages. We have an example of this in Grice's potentially insightful departure from the normal usage of ‘utterance’. Is this practice of stipulation not proof that we can choose to use a term however we want and to bestow on it any meaning whatsoever?

Against this, there are strong reasons for thinking that Humpty's thesis is mistaken, as Premise One asserts. A quick reason for thinking this is to follow Grice's own methodology of appealing to intuitions. Suppose Humpty were to utter: ‘There's glory for you.’ We would judge that his utterance means that some salient object or event is glorious. Humpty, following his own thesis, regards it as meaning that he has just given a nice knockdown argument. So, our considered intuitions do not accord with Humpty's thesis.

A second and more sophisticated reason for denying that meanings are entirely in our control depends not on any appeal to brute intuitions but on consideration of the role of meaning in the communication of knowledge. Utterer-controlled meaning, unless it is accompanied by uptake on the part of the audience, would not lead to communicative success. Evidence for this is provided by Alice's confusion when Humpty uses a word in a particular way without announcing to Alice beforehand that he is going to use it this way. What our utterances mean must be available to the audience if this meaning is to be any use. To that extent, what words mean is *not* entirely up to the speaker. Grice himself seems to be working towards this point when he writes, ‘the intended effect must be something which in some sense is within the control of the audience’ (Reading 1, p. 186).

Moreover, Humpty's thesis is not clearly supported by the phenomenon of stipulation. When we depart from common usage we generally have to announce this beforehand, using terms whose meaning is already familiar. Humpty fails to make a prior announcement, which is why Alice is left confused. It is only because he falsely believes his thesis that he is led to suppose that he can mean whatever he wishes with no prior announcement.

The falsity of Humpty's thesis can also be reconciled with the arbitrariness of language. There are other ways than individual legislation for words to come to mean what they do. Perhaps they mean what they do because we have beliefs about what they mean, not because we *decide* what they mean. So let us accept Premise One – that Humpty's thesis is false – and turn to the second premise, which is that Grice's theory entails Humpty's thesis. This is where the Humpty objection looks vulnerable.

Humpty's thesis asserts that our utterances and the words in them mean whatever we want or choose them to mean. Grice's theory is that what our utterances mean is a matter of what we intend to communicate. Intending, wanting and choosing may sound as though they come to much the same thing, but reflection suggests that they may be quite different.

Activity 9

Are any of the following states of mind equivalent?

(i) Wanting to deposit $8,000,000 in US Government bonds in your bank vault.

(ii) Intending to deposit $8,000,000 in US Government bonds in your bank vault.

(iii) Choosing to deposit $8,000,000 in US Government bonds in your bank vault.

(i) seems to be possible without either (ii) or (iii). (i) might well be true, but I can only intend something if I expect that I will be successful when I act on that intention, and I have no reason to expect that I will ever successfully deposit bonds of this value in my bank vault, even if I had one. Similarly, I can only choose to do something that is in my power, and depositing valuable bonds in my personal vault is, sadly, outside my power. (ii) and (iii) seem to differ from one another, too. I could mistakenly believe that I am in a position to make a large deposit and go to the bank intending to do just this; but I can only choose to make a deposit if I genuinely have the funds available.

If the subtleties noted in the previous activity are taken seriously, the second premise in the argument against Grice looks weak. Humpty's thesis is about choosing and wanting, neither of which is the equivalent of intending.

The main difference between Grice and Humpty seems to be this. Intention formation, the notion at the heart of Grice's theory but not Humpty's, requires an expectation of success. This was a lesson of the bank vault example: in general, we cannot form intentions to act unless we expect there is some chance that acting on the intention will lead to its fulfilment. In Grice's theory, the intention is to bring about a change of belief in an audience, and to bring it about through having them recognise this very intention. If we selected words according to personal whim, there would be no reasonable expectation of success. When Humpty used ‘impenetrability’, he had no reason to think that he would be successful in getting Alice to recognise him as meaning what, after the fact, he insists he did mean. After all, he had not warned her beforehand that he would be ascribing this unusual meaning to the word. With no such reason, he could not *even have formed the intention* to communicate using those words with that meaning. So Grice would not allow that Humpty meant what he later claims he meant.

Premise Two, then, seems to be false. Humpty's thesis may be mistaken, but Grice's theory looks like it is consistent with this falsity. In fact, once the distinctions between choosing and intending and between wanting and intending are granted, Grice's theory even seems to *explain* why Humpty's thesis is false. Because we cannot expect to be interpreted however we wish, we cannot intend to be interpreted however we wish, and so cannot mean whatever we wish.

These considerations may see off the Humpty objection, but they should also remind us that Grice still owes us a viable theory of what expressions mean. Grice, unlike Humpty, does not hold that what our utterances mean is entirely up to us. This is because what our utterances mean is a matter of what we intend to communicate, and we cannot intend to communicate that *p* unless we expect to succeed in communicating that *p*. This expectation of success is clearly tied to our choice of words. A speaker will have to feel that she has chosen *appropriate* words, words that her audience will interpret accordingly. For example, she will choose ‘tiger’ rather than ‘apricot’ if she intends to communicate information about a tiger (unless she and her audience have a special code). But to talk of ‘appropriate’ words, here, is to talk of words that *have a particular meaning*, i.e. ones that accord with what the speaker intends the audience to come to believe, etc. Grice's theory

of the meaning of utterances therefore needs to be supplemented with a theory of the meaning of expressions that is more plausible than the one he provides in Reading 1. We saw earlier (section 2.7) how that theory does not extend to words or novel sentences.

The meaning of words, and of the sentences built out of words, also figures in the next objection, due to John Searle. In the subsequent discussion of Searle's objection I will sketch a theory of word meaning that seems to be compatible with Grice's project. This theory has in fact been around since long before Grice wrote his paper.

2.9.2 Searle's objection

In 'What is a speech act?', John Searle introduces a memorable example of an utterance in which Grice's conditions are all met for it to mean one thing, but where the words used suggest that the utterance means something quite different, if it means anything at all. The conclusion Searle invites us to draw is that what our utterances mean is not exhausted by the speaker's intentions alone. An additional consideration is the meaning of the expressions used. If they don't match the intention, then nothing is meant.

Activity 10

Read the extract from Searle's article, 'What is a speech act?'. Searle summarises Grice's theory and then offers what he claims is a counterexample, the example of a captured American soldier in wartime Italy. Why is Searle's example a potential problem for Grice?

Click to view the reading by J.R Searle: 'What is a speech act?'

In the example, the American soldier's utterance of the German sentence does not mean what, according to Grice's theory, it should mean. On Grice's account, the utterance means *that the speaker is a German officer*. This is what the American soldier intends his Italian captors to believe, and he intends them to arrive at this belief through recognising his intention that this is what the Italians will come to believe. But intuitively either the American soldier's utterance means nothing at all, or else it means just that the speaker wishes to know whether his Italian captors are familiar with the land where the lemon trees blossom (i.e. what the German sentence he uses means).

What should we make of Searle's American soldier example? Grice could just dig in his heels and insist that what the American soldier's utterance means *really is* that he is a German officer, notwithstanding any intuitions we may have to the contrary over this interpretation. In philosophy we are often called on to reject the pre-theoretical intuitions we have about situations. But it is usually better to avoid doing so unless forced.

Searle's own response is to suggest that Grice has overlooked the importance of community-wide linguistic conventions, the 'rules of language', to the meaning of an utterance. Searle proposes that the speaker's intentions are *the main* source of what our utterances mean, but he adds a further condition: these intentions must accord with the meanings of the expressions we use, where this is fixed by community-wide conventions, the rules of the language from which the expressions are drawn.

Searle thinks of himself merely as extending the Gricean analysis, correcting it for an oversight. He does not think he is overthrowing the Gricean project. But it may be that adding this new constraint – that what the speaker means depends on community-wide conventions – would in fact leave Grice's project deeply damaged. This would depend on

whether the meanings of expressions, fixed as they are by community-wide linguistic conventions, can be reduced to the mental states (the beliefs, intentions, etc.) *of those who make up that community*. If they can, then the reduction of linguistic meaning to mental content will have succeeded, albeit not quite in the way Grice envisaged. On this new picture, but not on Grice's original picture, the meaning of an individual utterance would depend on more than the mental states of the speaker alone. It would depend also on the mental states of those in the speaker's community, since these fix the meanings of the expressions used, and these meanings in turn are a factor in the meaning of the individual utterance.

Two questions arise at this point. It is clear that following Searle's suggestion or persisting with Grice's original theory will require a viable theory of how the meaning of expressions reduces to the content of mental states. So what is that theory? A different question is whether the relevant mental states are those of the speaker alone, or whether they also include those of members of the wider linguistic community as Searle implies. Let us start with this second question.

Grice could endorse a slightly weakened version of Searle's additional requirement, and in doing so persist with his claim that only the individual speaker's mental states are relevant to what an utterance means. Searle requires that the speaker's intentions *actually* accord with the community-wide meaning of the sentence used. But Grice could require merely that the speaker's intentions accord with what the speaker *thinks* is the community-wide meaning of the sentence used – or even with what the speaker thinks the specific audience takes the sentence's meaning to be. Adding such a weak requirement is enough to explain our intuitions about the American soldier example. The American soldier does not think that the established meaning of 'Kennst du das Land, wo die Zitronen blühen?' is that the speaker is a German officer, so Grice would not be committed to saying that this is what the utterance means. Moreover, Grice's attempt to define the meaning of utterances in terms of the individual speaker's psychological states would still be on track. The weakened condition, unlike Searle's condition, is couched in terms of the beliefs of the individual utterer alone.

We might wonder whether this weaker condition deals with counterexamples as well as Searle's condition does. For example, suppose the American soldier really did believe that 'Kennst du das Land, wo die Zitronen blühen?' meant the same as 'I am a German officer'. Perhaps he was told this in military training. Would it now be fair to say that the American soldier's utterance meant that he was a German officer? Grice's new condition (and all the old ones) would be met but Searle's condition would not, so an affirmative answer would favour Grice's weaker condition and a negative answer would favour Searle's stronger one. I will not try to settle this issue here, except to note that people's intuitions about this new version of the scenario differ. Some think that the American soldier's utterance *does* now mean the same as 'I am a German officer'. They are likely to be happy with the weaker condition. Others deny that it means this. They are likely to be sympathetic to Searle's stronger condition.

Now consider the challenge, facing both Searle and Grice, of how to reduce the meaning of expressions to the meaning of mental states. Even adopting the weaker condition leaves this challenge in place. Grice could not treat beliefs about what expressions mean as fundamental, because 'what expressions mean' is itself a kind of non-natural meaning, and as such needs to be reduced to the content of mental states.

2.10 Section summary

After setting aside ‘natural’ meaning as largely irrelevant to language (section 2.3), Grice attempts to define the (non-natural) meaning of utterances in terms of the content of the speakers’ psychological states, and in particular in terms of their intentions in performing those utterances. He reaches a final definition, which we called Grice 3, after two false starts (section 2.6). The meaning of expressions, or of sentences at least, is derivative, defined by him in terms of the meaning of typical utterances.

Several problems confront Grice's proposal, including the phenomenon of apparently meaningful but audience-less speech, and the massive complexity of the intentions Grice attributes to ordinary speakers (section 2.9). Another common objection is that he seems to be committing the same errors as Humpty Dumpty (section 2.9, ‘The Humpty objection’). Replies to each of these objections are available, though whether these replies are ultimately successful was not resolved here.

Perhaps the most complex issue in Grice's discussion in ‘Meaning’ concerns the place in it of the meaning of expressions. One source of difficulty is that he provides no theory of the meaning of words or novel sentences (section 2.7). And Searle thinks that he pays insufficient notice to the contribution of community-wide linguistic conventions to the meaning of expressions and hence of utterances (section 2.9, ‘Searle's objection’).

One thing we have learnt in the course of this section is that the key debate is not as simple as was implied in the introduction to this section. The core debate over the source of the meaning of utterances is not between those who look to the mental states of the speaker and those look to the meaning of the expressions used. It turns out that everyone, even those who look to the mental states of the speaker, must provide a theory of the meaning of expressions. Moreover, Searle and others claim that the community-wide meaning of the expressions used is a key factor in what an utterance means, but they often wish to claim that the meaning of expressions itself reduces to the content of mental states – not the mental states of the individual speaker but the mental states of members of the linguistic community. The real debate is between those who think that all species of linguistic meaning – the meaning of utterances and the meaning of words and sentences – reduce to the content of the mental states of those who use the language, and those who deny this. Searle and Grice are on the same side on this matter, even if they disagree over strategy.

2.11 Further reading

Grice's writing on the philosophy of language, including the 1957 paper ‘Meaning’, is collected in Grice 1989. Discussion of the issues raised in ‘Meaning’ can be found in Avramides 1997 (the most accessible), Blackburn 1984, Miller 1998, and Taylor 1998. A defence of the Gricean Programme can be found in Schiffer 1972, later retracted in Schiffer 1987. Searle's position is developed at length in Searle 1969.

Conclusion

This free course provided an introduction to studying the arts and humanities. It took you through a series of exercises designed to develop your approach to study and learning at a distance and helped to improve your confidence as an independent learner.

Keep on learning

Study another free course

There are more than **800 courses on OpenLearn** for you to choose from on a range of subjects.

Find out more about all our free courses.

Take your studies further

Find out more about studying with The Open University by visiting our online prospectus.

If you are new to university study, you may be interested in our Access Courses or Certificates.

What's new from OpenLearn?

Sign up to our newsletter or view a sample.

For reference, full URLs to pages listed above:

OpenLearn – www.open.edu/openlearn/free-courses

Visiting our online prospectus – www.open.ac.uk/courses

Access Courses – www.open.ac.uk/courses/do-it/access

Certificates – www.open.ac.uk/courses/certificates-he

Newsletter – www.open.edu/openlearn/about-openlearn/subscribe-the-openlearn-newsletter

References

Avramides, A. (1997) 'Intention and Convention', in C. Wright and B. Hale (eds) *A Companion to the Philosophy of Language*, Oxford, Blackwell.

Blackburn, S. (1984) *Spreading the Word: Groundings in the Philosophy of Language*, Oxford, Clarendon Press.

Carroll, L.(1893) 'Through the looking-glass and what Alice found there', in *Alice's Adventures in Wonderland*, London, Macmillan.

Grice, H.P. (1957) 'Meaning', *Philosophical Review*, 66, 377–88.

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, Mass., Harvard University Press.

Locke, J. (1997) *An Essay Concerning Human Understanding*, ed. R. Woolhouse, Harmondsworth, Penguin (based on the posthumously published 5th edn, 1706).

Lycan, W.G. (1996) *The Philosophy of Language*, London, Routledge.

Molière, J.B. de (1879?) *The Imaginary Invalid* (*Le Malade Imaginaire*), trans. Roscoe Mongan, London, James Cornish and Sons (first published 1673).

Miller, A. (1998) *Philosophy of Language*, London, UCL Press.

Pinker, S. (1994) *The Language Instinct*, London, Allen Lane.

Schiffer, S.R. (1972) *Meaning*, Oxford, Clarendon Press.

Schiffer, S.R. (1987) *Remnants of Meaning*, Cambridge, Mass., MIT Press.

Searle, J.R. (1969) *Speech Acts: An Essay in the Philosophy of Language*, London, Cambridge University Press.

Searle, J.R.(1971) 'What is a speech act?', in J.R. Searle (ed.) *The Philosophy of Language*, Oxford, Oxford University Press, 39–53. Originally published in M. Black (ed.) (1965) *Philosophy in America*, London, Allen & Unwin, 221–39.

Taylor, K. (1998) *Truth and Meaning: An Introduction to the Philosophy of Language*, Oxford, Blackwell.

Acknowledgements

This course was written by Dr Alex Barber

Grateful acknowledgement is made to the following sources for permission to reproduce material in this course:

Course image: James St. John in Flickr made available under
Creative Commons Attribution 2.0 Licence.

Grice, H.P., (1957) 'Meaning', in *The Philosophical Review*, Volume 66, no.3, pp. 357-388. Copyright, 1957, Duke University Press. All rights reserved. Used by permission of the publisher. http://www.dukeupress.edu

Searle, J.R., (1971) 'What is a speech act?', in J.R.Searle (ed.) *The Philosophy of Language*, Oxford, Oxford University Press

Don't miss out:

If reading this text has inspired you to learn more, you may be interested in joining the millions of people who discover our free learning resources and qualifications by visiting The Open University - www.open.edu/openlearn/free-courses