Literal Meaning

François Recanati
Institut Jean-Nicod
1bis, avenue de Lowendal, 75007 Paris, France
recanati@ehess.fr
Contents

Foreword
Acknowledgments
Analytical table of contents

Chapter 1: Two approaches to 'what is said'
1.1 The basic triad
1.2 Minimalism
1.3 Literal truth-conditions v. actual truth-conditions
1.4 A problem for Minimalism
1.5 The availability of what is said
1.6 The availability-based approach
1.7 'Saying' as a pragmatic notion
1.8 Availability v. Minimalism

Chapter 2: Primary pragmatic processes
2.1 Enrichment, loosening, and transfer
2.2 Rejecting the Gricean picture
2.3 Accessibility
2.4 Objections and responses
2.5 Interactive processing
2.6 The role of schemata

Chapter 3: Relevance-theoretic objections
3.1 One or two systems?
3.2 Personal and sub-personal inferences
3.3 Implicature or enrichment?
3.4 Mutual adjustment of explicature and implicature
3.5 Implicated premisses
3.6 Personal-level inferences : occurrent v. dispositional

Chapter 4: The Syncretic View
4.1 Four levels?
4.2 Semantics and pragmatics: the literalist picture
4.3 Semantic underdeterminacy
4.4 The minimal proposition as 'common denominator'
4.5 Interaction between saturation and optional pragmatic processes
4.6 Do we need the minimal proposition?
4.7 The reflexive proposition

Chapter 5: Nonliteral uses
5.1 Nonliteral uses as nonminimal departures from literal meaning
5.2 Nonliteral uses and secondary meaning
5.3 Nonminimal departures without secondariness
5.4 The transparency condition
5.5 Varieties of nonliteral meaning
5.6 Internal v. external duality
5.7 Conclusion
Chapter 6: From Literalism to Contextualism
6.1 Five positions
6.2 Indexicalism
6.3 Contextualism
6.4 Literalist responses to the contextualist challenge
6.5 Where Indexicalism and Contextualism meet

Chapter 7: Indexicalism and the Binding Fallacy
7.1 Mandatory v. optional
7.2 Two criteria
7.3 The indexicalist challenge
7.4 Is the Binding Criterion reliable?
7.5 Variadic functions
7.6 The Binding Fallacy
7.7 Conclusion: the failure of Indexicalism

Chapter 8: Circumstances of evaluation
8.1 Modality
8.2 Time and tense
8.3 Situations
8.4 Saturation or enrichment?
8.5 Sub-sentential circumstances
8.6 Conclusion

Chapter 9: Contextualism: How far can we go?
9.1 The modulation of sense
9.2 The semantic relevance of modulation
9.3 Four approaches
9.4 Truth-conditionality instability: from Waismann's 'open texture' to Searle's 'background'
9.5 Ostensive definitions
9.6 Meaning Eliminativism
9.7 Conclusion

Conclusion
10.1 Alleged arguments against Contextualism
10.2 Remants of Literalism
10.3 Availability, Minimalism, and the dispositional/occurrent contrast
Chapter 1

Two approaches to 'what is said'

1.1. The basic triad

Anyone who has reflected on the sentence meaning/speaker's meaning distinction knows that a simple distinction is in fact insufficient. Two equally important distinctions must be made. First, there is the distinction between the linguistic meaning of a sentence-type, and what is said (the proposition expressed) by an utterance of the sentence. For example, the English sentence 'I am French' has a certain meaning which, qua meaning of a sentence-type, is not affected by changes in the context of utterance. This context-independent meaning contrasts with the context-dependent propositions which the sentence expresses with respect to particular contexts. Thus 'I am French', said by me, expresses the proposition that I am French; if you utter the sentence, it expresses a different proposition, even though its linguistic meaning remains the same across contexts of use.

Second, there is a no less important distinction between what is actually said and what is merely 'conveyed' by the utterance. My utterance of 'I am French' expresses the proposition that I am French, but there are contexts in which it conveys much more. Suppose that, having been asked whether I can cook, I reply: 'I am French'. Clearly my utterance (in this context) provides an affirmative answer to the question. The meaning of the utterance in such a case includes more than what is literally said; it also includes what the utterance 'implicates'.

'What is said' being a term common to both distinctions, we end up with a triad:

sentence meaning

vs

what is said

vs

what is implicated

The distinguishing characteristic of sentence meaning (the linguistic meaning of the sentence type) is that it is conventional and context-independent. Moreover, in general at least, it falls short of constituting a complete proposition, i.e. something truth-evaluable. In contrast, both 'what is said' and 'what is implicated' are context-dependent and propositional. The difference between 'what is said' and 'what is implicated' is that the former is constrained by sentence meaning in a way in which the implicatures aren't. What is said results from fleshing out the meaning of the sentence (which is like a semantic 'skeleton') so as to make it propositional. The propositions one can arrive at through this process of contextual enrichment or 'fleshing out' are constrained by the skeleton which serves as input to the process. Thus 'I am French' can express an indefinite number of propositions, but the propositions in question all have to be compatible with the semantic potential of the sentence; this is why the English sentence 'I

---

1 See Grice, Paul Studies in the Way of Words (Harvard University Press, 1989), p. 24: « I wish to introduce, as terms of art, the verb implicate and the related nouns implicature (cf implying) and implicatum (cf what is implied). The point of this maneuver is to avoid having, on each occasion, to choose between this or that member of the family of verbs for which implicate is to do general duty. »
am French' cannot express the proposition that kangaroos have tails. There is no such constraint on the propositions which an utterance of the sentence can communicate through the mechanism of implicature. Given enough background, an utterance of 'I am French' might implicate that kangaroos have tails. What's implicated is implicated by virtue of an inference, and the inference chain can (in principle) be as long and involve as many background assumptions as one wishes.

The basic triad can be mapped back onto the simple sentence meaning/speaker's meaning distinction by grouping together two of the three levels. There are two ways to do it, corresponding to two interpretations for the triad. The 'minimalist' interpretation stresses the close connection between sentence meaning and what is said; together, sentence meaning and what is said constitute the literal meaning of the utterance as opposed to what the speaker means:

```
<table>
<thead>
<tr>
<th>sentence meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>what is said</td>
</tr>
<tr>
<td>vs</td>
</tr>
</tbody>
</table>
```

speaker's meaning

The other, 'non-minimalist' interpretation of the triad stresses the commonality between what is said and what is implicated, both of which are taken to be pragmatically determined:

```
<table>
<thead>
<tr>
<th>sentence meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>vs</td>
</tr>
<tr>
<td>what is said</td>
</tr>
<tr>
<td>what is implicated</td>
</tr>
</tbody>
</table>
```

Essential to this interpretation is the claim that 'what is said', though constrained by the meaning of the sentence, is not as tightly constrained as is traditionally thought and, in particular, does not obey what I will refer to as the 'minimalist' constraint.

### 1.2. Minimalism

As I said above, what distinguishes 'what is said' from the implicatures is the fact that the former must be « closely related to the conventional meaning of the words (the sentence) [one] has uttered ». However, this constraint can be construed more or less strictly. What I call 'Minimalism' construes the constraint very strictly; 'what is said', in the minimalist framework, departs from the conventional meaning of the sentence (and incorporates contextual elements) only when this is necessary to 'complete' the meaning of the sentence and make it propositional. In other words, the distance between sentence meaning and what is said is kept to a minimum (hence the name 'Minimalism').

The crucial notion here is that of 'saturation'. Saturation is the process whereby the meaning of the sentence is completed and made propositional through the contextual assignment of semantic values to the constituents of the sentence whose interpretation is

---

context-dependent (and, possibly, through the contextual provision of 'unarticulated' propositional constituents, if one assumes, as some philosophers do, that such constituents are sometimes needed to make the sentence fully propositional). This process takes place whenever the meaning of the sentence includes something like a 'slot' requiring completion or a 'free variable' requiring contextual instantiation. Thus an indexical sentence like 'He is tall' does not express a complete proposition unless a referent has been contextually assigned to the demonstrative pronoun 'he', which acts like a free variable in need of contextual instantiation. Genitives provide another well-known example: an utterance including the phrase 'John's book' does not express a complete proposition unless a particular relation has been identified as holding between the book and John. Nominal compounds work the same way: 'burglar nightmare' means something like 'a nightmare that bears a certain relation R to burglars', which relation must be contextually identified. Other well-known examples of saturation include parametric predicates ('small', 'on the left'), definite null instantiation (i.e. the case where one of the arguments in the semantic structure of a lexeme, typically a verb, is not syntactically realized and must be contextually identified, as when someone says 'I heard' or 'I noticed'), and so on and so forth.

Whenever saturation is in order, appeal to the context is necessary for the utterance to express a complete proposition: from a semantic point of view, saturation is a mandatory contextual process. Other contextual processes — e.g. the inference process generating implicatures — are semantically optional in the sense that the aspects of meaning they generate are dispensable; the utterance would still express a complete proposition without them. According to Minimalism, those extra constituents of meaning which are not necessary for propositionality are external to what is said. The only justification for including some pragmatically determined constituent of meaning into what is said (as opposed to what is merely conveyed) is the indispensability of such a constituent — the fact that the utterance would not express a complete proposition if the context did not provide such a constituent.

1.3. Literal truth-conditions v. actual truth-conditions

Consider examples (1)-(6), often discussed in the literature:

(1) I've had breakfast
(2) You are not going to die
(3) It's raining
(4) The table is covered with books
(5) Everybody went to Paris
(6) John has two children

In all such cases, as we shall see, the minimalist constraint implies that what the utterance literally says is not what intuitively seems to be said.

From a minimalist point of view, the first sentence, 'I've had breakfast' expresses the proposition that S (the speaker) has had breakfast before t* (the time of utterance). Strictly speaking this proposition would be true if the speaker had had breakfast twenty years ago and never since. This is clearly not what the speaker means (when she answers to the question 'Do you want something to eat' and replies 'I've had breakfast'); she means something much more

---

^3 Even when saturation consists in contextually providing a constituent that is unarticulated in surface syntax (as the implicit argument in 'I noticed'), it is something in the sentence (here the predicate 'notice', which arguably denotes a two-place relation) which triggers the search for the contextual element and makes it obligatory. See §2.1 of my ‘Unarticulated Constituents’, in *Linguistics and Philosophy* 25 (2002), 299-345.
specific, namely that she's had breakfast on that very day (i.e. the day which includes t*). This aspect of speaker's meaning, however, has to be construed as external to what is said and as being merely conveyed, in the same way in which the utterer of 'I am French' implies, but does not say, that he is a good cook. That is so because the 'minimal' interpretation, to the effect that the speaker's life was not entirely breakfastless, is sufficient to make the utterance propositional. Nothing in the sentence itself forces us to bring in the implicit reference to a particular time span. Indeed we can easily imagine contexts in which a speaker would use the same sentence to assert the minimal proposition and nothing more.4

The same thing holds even more clearly for the second example. Kent Bach, to whom it is due, imagines a child crying because of a minor cut and her mother uttering (2) in response.5 What is meant is: 'You're not going to die from that cut'. But literally the utterance expresses the proposition that the kid will not die tout court — as if he or she was immortal. The extra element contextually provided (the implicit reference to the cut) does not correspond to anything in the sentence itself; nor is it an unarticulated constituent whose contextual provision is necessary to make the utterance fully propositional. Again, we can easily imagine a context in which the same sentence would be used to communicate the minimal proposition and nothing more.

What about (3)? John Perry and many others after him have argued as follows.6 Even though nothing in the sentence 'It's raining' stands for a place, nevertheless it does not express a complete proposition unless a place is contextually provided. The verb 'to rain', Perry says, denotes a dyadic relation — a relation between times and places. In a given place, it doesn't just rain or not, it rains at some times while not raining at others; similarly, at a given time, it rains in some places while not raining in others. To evaluate a statement of rain as true or false, Perry says, we need both a time and a place. Since the statement ‘It is raining’ explicitly gives us only the two-place relation (supplied by the verb) and the temporal argument (indexically supplied by the present tense), the relevant locational argument must be contextually supplied for the utterance to express a complete proposition. If Perry is right, the contextual provision of the place concerned by the rain is an instance of saturation, like the assignment of a contextual value to the present tense: both the place and the time are constituents of what is said, even though, unlike the time, the place remains unarticulated in surface syntax.

But is Perry right? If really the contextual provision of a place was mandatory, hence an instance of saturation, every token of 'It's raining' would be unevaluable unless a place were contextually specified. Yet I have no difficulty imagining a counterexample, i.e. a context in which ‘It is raining’ is evaluable even though no particular place is contextually singled out. In ‘Unarticulated Constituents’ I depicted an imaginary situation in which

rain has become extremely rare and important, and rain detectors have been disposed all over the territory (whatever the territory — possibly the whole Earth). In the imagined scenario, each detector triggers an alarm bell in the Monitoring Room when it detects rain. There is a single bell; the location of the triggering detector is indicated by a light on a board in the Monitoring Room. After weeks of total drought, the bell eventually rings in the Monitoring Room. Hearing it, the weatherman on duty in the adjacent room shouts: ‘It’s

---

raining!’ His utterance is true, iff it is raining (at the time of utterance) in some place or other.\footnote{Unarticulated Constituents’, p. 317.}

The fact that one can imagine an utterance of ‘It’s raining’ that is true iff it is raining (at the time of utterance) in some place or other arguably establishes the pragmatic nature of the felt necessity to single out a particular place, in the contexts in which such a necessity is indeed felt. When a particular place is contextually provided as relevant to the evaluation of the utterance, it is for pragmatic reasons, not because it is linguistically required. (Again, if it were linguistically required, in virtue of semantic properties of the sentence type, it would be required in every context.) If this is right, then the contextual provision of a place is not an instance of saturation after all: it's not something that's mandatory. It follows (by minimalist standards) that the place is not a constituent of what is strictly and literally said: when I say 'It is raining' (rather than something more specific like 'It's raining in Paris' or 'It's raining here'), what I literally say is true iff it's raining somewhere or other.\footnote{See Borg, Emma ‘Saying What You Mean: Unarticulated Constituents and Communication’ (forthcoming) for a defense of that claim.} That is obviously not what I mean, since what I mean involves a particular place. Appearances notwithstanding, the situation is similar to the case of 'I've had breakfast', where a restricted time-interval is contextually provided for pragmatic reasons, without being linguistically mandated.

Examples (4) and (5) are amenable to the same sort of treatment. According to standard Russellian analysis, a definite description conveys an implication of uniqueness: hence 'The table is covered with books' is true iff there is one and only table and it is covered with books. To make sense of this, we need either to focus on a restricted situation in which there is indeed a single table, or to expand the predicate 'table' and enrich it into, say 'table of the living-room' in order to satisfy the uniqueness constraint. Either way, the form of enrichment through which we make sense of the utterance is not linguistically mandated: it is only pragmatically required. If we don't enrich, what we get is an already complete proposition (albeit one that is pretty absurd): the proposition that the only existing table is covered with books. Similarly with example (5): without enrichment the utterance expresses a proposition that is true iff every existing person went to Paris. Such a proposition is unlikely to be true, but that does not make it incomplete. The enrichment process through which, in context, we reach the proposition actually communicated (to the effect that everybody in such and such group went to Paris) is not linguistically but pragmatically required; hence it is not an instance of saturation, but an optional process of 'free enrichment'. It follows that, in those examples as much as in the previous ones, the proposition literally expressed is different from, and more general than, the proposition actually communicated.

1.4. A problem for Minimalism

In general, the literal truth-conditions posited as part of the Minimalist analysis turn out to be very different from the intuitive truth-conditions which untutored conversational participants would ascribe to the utterance. This divergence between the intuitive truth-conditions of an utterance and the literal truth-conditions postulated by the theorist is particularly striking in connection with examples like (6). According to a fairly standard view,\footnote{See Horn, Larry The Natural History of Negation (Chicago University Press, 1989), pp. 205-16.} the proposition literally expressed by (6) is the proposition that John has at least three children, i.e., no less than three but possibly more. In certain contexts this corresponds to what the speaker actually means (as when I say, 'If John has three children he can benefit from lower rates on public
transportation') but in other contexts what the speaker means is quite different. Suppose for example that I am asked how many children John has and that I reply by uttering (6). Clearly, in this context, I mean that John has (exactly) three children — no more and no less. This is standardly accounted for by saying that the proposition literally expressed, to the effect that John has at least three children, combines with the 'implicature' that John has no more than three children (a generalized implicature that is accounted for in terms of the maxim of quantity); as a result of this combination, what is globally communicated — and what I actually mean — is the proposition that John has exactly three children. Now this is the only proposition I am conscious of expressing by my utterance; in particular, I am unaware of having expressed the 'minimal' proposition that John has at least three children. To account for this obvious fact, the minimalist claims that we are aware only of what is globally conveyed or 'communicated' by the utterance. Analysing this into 'what is literally said' and 'what is implied' is the linguist's task, not something that is incumbent upon the normal language user. Figure 1.1 illustrates this widespread conception.

[Figure 1.1 about here]

The problem with this conception is that it lacks generality. Recall the example I gave earlier — the utterance 'I am French' used to convey that I am a good cook. In the relevant situation of utterance, both the speaker and the listener are aware that the speaker says he is French, and thereby implies he is a good cook. This typical case of implicature is very different from a case like (6) in which the speaker is not only (like the hearer) unaware of the proposition literally expressed, but would strongly deny having said what the minimalist claims was actually said.

It turns out that there are two sorts of case. On the one hand there are prototypical cases of implied meaning, in which the participants in the speech situation are aware both of what is said and of what is implied, and also of the inferential connection between them. On the other hand, there are the cases illustrated by (1)-(6). Given his willingness to treat certain aspects of the intuitive meaning of (1)-(6) as conversational implicatures external to what is literally said, the minimalist must explain why those implicatures, unlike the prototypical cases (e.g. the French/cook example), do not have the property of conscious 'availability'.

The only explanation I have come across in the literature makes use of Grice's distinction between 'generalized' and 'particularized' conversational implicatures, i.e. between implicatures which arise 'by default', without any particular context or special scenario being necessary, and those which require such specific contexts. In contrast to the latter, the former are 'hard to distinguish from the semantic content of linguistic expressions, because such implicatures [are] routinely associated with linguistic expressions in all ordinary contexts'.

Generalized implicatures are unconsciously and automatically generated and interpreted. They belong to the 'micropragmatic' rather than to the 'macropragmatic' level, in Robin Campbell's typology:

---

10 As Grice puts it in one of his early papers, "one should not make a weaker statement rather than a stronger one unless there is a good reason for so doing" (Grice, Paul 'The Causal Theory of Perception', in Proceedings of the Aristotelian Society, Supplementary Volume 35 [1961], p. 132). Since the statement that John has (at least) three children is weaker than the statement that John has \( n \) children (for \( n > 3 \)), the maxim is obeyed only if John has no more than three children. (If John has more than three children, the statement that he has three is too weak and violates the maxim.) The statement 'John has three children' therefore implicates that John has no more than three children, in virtue of the presumption that the maxim is obeyed.

A macropragmatic process is one constituted by a sequence of explicit inferences governed by principles of rational cooperation. A micropragmatic process develops as a cryptic [= unconscious] and heuristic procedure which partially replaces some macropragmatic process and which defaults to it in the event of breakdown.¹²

But there are problems with this explanation. According to Horn, the generalized nature of an implicature does not entail its conscious unavailability — its 'cryptic' character.¹³ In other words, it is possible for an implicature to be both 'generalized' and intuitively accessible as an implicature distinct from what is said. Thus Horn insists that the generalized scalar implicature from 'some' to 'not all' is consciously available (in contrast to that from 'three' to 'exactly three'). A speaker saying 'Some students came to the meeting' normally implies that not all students came, and when this is so there is no tendency on the part of the interpreter to conflate the implicature with what is said. This is actually debatable, for the 'implicature' at issue can arise at sub-sentential level (e.g. 'He believes some students came'), and in such cases there are reasons to doubt that the availability condition is satisfied. Be that as it may, the 'generalization' of an implicature does not seem to be necessary for its unconscious character. Many particularized 'bridging' inferences are automatic and unconscious. To take an example from Robyn Carston, 'He went to the cliff and jumped' is readily interpreted as saying that the person referred to jumped over the cliff, even though this is only contextually suggested.

1.5. The availability of what is said

In earlier writings I put forward a conception diametrically opposed to that illustrated by Figure 1.1 above.¹⁴ 'What is said', I held, is consciously available to the participants in the speech situation (Figure 1.2). 'What is communicated' is not a distinct level where 'what is said' and 'what is implied' have been merged and integrated into a unified whole; it is merely a name for the level at which we find both what is said and what is implied, which level is characterized by conscious accessibility. On this picture, there are only two basic levels: the bottom level in which we find both the meaning of the sentence and the contextual factors which combine with it to yield what is said; and the top level at which we find both what is said and what is implied, both being consciously accessible (and accessible as distinct).

[Figure 1.2 about here]

The availability of what is said follows from Grice's idea that saying itself is a variety of nonnatural meaning. One of the distinguishing characteristics of nonnatural meaning, on Grice's analysis, is its essential overtness. Nonnatural meaning works by openly letting the addressee recognize one's primary intention (e.g. the intention to impart a certain piece of information, or the intention to have the addressee behave in a certain way), that is, by (openly) expressing that intention so as to make it graspable. This can be done in all sorts of

---

ways, verbal or nonverbal. Even if we restrict ourselves to verbal communication, there are many ways in which we can mean things by uttering words. *Saying* is one way; *implying* is another.

The view that 'saying' is a variety of nonnatural meaning entails that what is said (like what is meant in general, including what is implied) must be available — it must be open to public view. That is so because nonnatural meaning is essentially a matter of intention-recognition. On this view what is said by uttering a sentence depends upon, and can hardly be severed from, the speaker's publicly recognizable intentions. Hence my 'Availability Principle', according to which 'what is said' must be analysed in conformity to the intuitions shared by those who fully understand the utterance — typically the speaker and the hearer, in a normal conversational setting.

I take the conversational participants' intuitions concerning what is said to be revealed by their views concerning the utterance's truth-conditions. I assume that whoever fully understands a declarative utterance knows which state of affairs would possibly constitute a truth-maker for that utterance, i.e. knows in what sort of circumstance it would be true. The ability to pair an utterance with a type of situation in this way is more basic than, and in any case does not presuppose, the ability to report what is said by using indirect speech; it does not even presuppose mastery of the notion of 'saying'. Thus the proper way to elicit such intuitions is not to ask the subjects 'What do you think is said (as opposed to implied or whatever) by this sentence as uttered in that situation'? I therefore tend to agree with Bach's criticism of the experiments through which Gibbs and Moise attempted to support the availability-based approach.

[They] thought they could get their data about what is said, and thereby test the validity of Recanati's Availability Principle, by asking people what is said by a given utterance, or by asking them whether something that is conveyed by a given utterance is implicated or merely said. Evidently they assume that what people say about what is said is strongly indicative of what is said. In fact, what it is indicative of is how people apply the phrase « what is said »... It tells us little about what is said, much less about the cognitive processes whereby people understand utterances.

However, Bach himself uses what he calls the 'IQ test' to determine what is said, that is, he ties what is said to indirect speech reports of what is said. I find this procedure most objectionable, and that is not what I mean when I claim that what is said should be

---

15 Direct Reference, p. 248.
16 Michael Thau notes that speakers almost never explicitly think about the distinction between what they've said and what they've implicated. So the question of what a speaker takes himself to have said by some utterance will have to depend upon the answer he would give if he were asked. And it's very likely that in many circumstances there won't be a single answer, that the answer will differ depending on how the question is put. It's also very likely that the answer will vary from circumstance to circumstance. (Consciousness and Cognition [Oxford University Press, 2002], p. 148)

Contrary to what Thau thinks, however, this does not speak against the availability-based approach. The speaker's intuitions concerning what is said need not involve the very notion of what is said.


18 Bach, Kent ‘Seemingly Semantic Intuitions’, in Keim Campbell, Joseph, O'Rourke, Michael, and Schier, David (eds.) Meaning and Truth (Seven Bridges Press, 2002), p. 27.

individuated according to the intuitions of normal interpreters. Thus I strongly disagree with Cappelen and Lepore's surprising statement:

We ourselves don't see how to elicit intuitions about what-is-said by an utterance of a sentence without appealing to intuitions about the accuracy of indirect reports of the form 'He said that...' or 'What he said is that...' or even 'What was said is that...'.

I find this statement surprising, because there obviously is another way of eliciting truth-conditional intuitions. One has simply to provide subjects with scenarios describing situations, or, even better, with — possibly animated — pictures of situations, and to ask them to evaluate the target utterance as true or false with respect to the situations in question. That procedure has been used by several researchers to test speaker's intuitions about e.g. the truth-conditions of donkey sentences. Thus Bart Geurts describes his experimental set-up (inspired from earlier work by Yoon) as follows:

Twenty native speakers of Dutch were asked to judge whether or not donkey sentences correctly described pictured situations. Instructions urged subjects to answer either true or false, but they were also given the option of leaving the matter open in case they couldn't make up their minds.

This procedure presupposes that normal interpreters have intuitions concerning the truth-conditional content of utterances. On my view, those intuitions correspond to a certain 'level' in the comprehension process — a level that a proper theory of language understanding must capture. That is the level of 'what is said' (as opposed to e.g. what is implied).

1.6. The availability-based approach

From a psychological point of view, we can draw a helpful parallel between understanding what one is told and understanding what one sees. In vision, the retinal stimuli undergo a complex (multistage) train of processing which ultimately outputs a conscious perception, with the dual character noted by Brentano: the subject is aware both of what he sees, and of the fact that he is seeing it. Although more complex in certain respects, the situation with language is similar. The auditory signal undergoes a multistage train of processing which ultimately outputs a conceptual experience: the subject understands what is said. This is very much like (high-level) perception. If I am told that it is four o'clock, I hear that it is four o'clock, just as, when I look at my watch, I see that it is four o'clock. Like the visual experience, the locutionary experience possesses a dual character: we are aware both of what is said, and of the fact that the speaker is saying it.

In calling understanding an experience, like perception, I want to stress its conscious character. Understanding what is said involves entertaining a mental representation of the subject-matter of the utterance that is both determinate enough (truth-evaluable) and consciously available to the subject. This suggests a criterion, distinct from the Minimalist criterion, for demarcating what is said. Instead of looking at things from the linguistic side and equating 'what is said' with the minimal proposition one arrives at through saturation, we can take a more psychological stance and equate what is said with (the semantic content of)

---

21 For an implicit use of that procedure, see Kripke, Saul Naming and Necessity (Blackwell, 1980), p. 12.
the conscious output of the complex train of processing which underlies comprehension. To be sure, that output itself is subject to further processing through e.g. inferential exploitation. Consider, once again, vision. Seeing John's car, I can infer that he is around. Similarly, hearing that John has had breakfast, I can infer that he is not hungry and does not need to be fed. Just as what is seen corresponds to the primary conscious output of visual processing, not to what can be secondarily derived from it, 'what is said' corresponds to the primary truth-evaluable representation made available to the subject (at the personal level) as a result of processing the sentence. It is therefore minimal in a certain sense, though not (as we shall see) in the sense of Minimalism.

Accordingly, I distinguish between two sorts of pragmatic process. The contextual processes which, like saturation, are (subpersonally) involved in the determination of what is said I call primary pragmatic processes. In contrast, secondary pragmatic processes are ordinary inferential processes taking us from what is said, or rather from the speaker's saying of what is said, to something that (under standard assumptions of rationality and cooperativeness) follows from the fact that the speaker has said what she has said. To the extent that the speaker overtly intends the hearer to recognize such consequences as following from her speech act, they form an integral part of what the speaker means by her utterance. That is, roughly, Grice's theory of 'conversational implicature'. An essential aspect of that theory is that the hearer must be able to recognize what is said and to work out the inferential connection between what is said and what is implied by saying it. Again, it follows that what is said must be consciously available to the interpreter. It must satisfy what I call the Availability constraint.

In this framework we solve the difficulty raised in section 1.5. We no longer have two sorts of case of implicature — the prototypical cases where the interlocutors are aware of what is said, aware of what is implied, and aware of the inferential connection between them, and the cases in which there is no such awareness. Conscious awareness is now a built-in feature of both what is said and the implicatures. That is so because what is said is the conscious output of linguistic-cum-pragmatic processing, and the implicatures correspond to further conscious representations inferentially derived, at the personal rather than sub-personal level, from what is said (or, rather, from the speaker's saying what is said). The alleged cases in which the speech participants themselves are not distinctly aware of what is said and of what is implied are reclassified: they are no longer treated as cases of 'implicature', strictly speaking, but as cases in which a primary pragmatic process operates in the (sub-personal) determination of what is said.

1.7. 'Saying' as a pragmatic notion

So far I have followed Grice, who construes saying as a variety of meaning. But this

---

23 As Ian Rumfitt once put it, "what is said in the course of an utterance is nothing other than what somebody who understands the utterance understands to be said" ("Content and Context: the Paratactic Theory Revisited and Revised", in Mind 102 [1993], p. 439).


25 This is consonant with the approach taken by some semanticists who insist that e.g. scalar ‘implicatures’ « are not computed after truth-conditions of (root) sentences have been figured out ; they are computed phrase by phrase » (Chierchia, Gennaro ‘Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface’, forthcoming). In Chapter 2 I will stress the fact that primary pragmatic processes operate locally, in contrast to secondary pragmatic processes, which can only operate when the truth-conditions of the sentence have been figured out.
pragmatic approach to 'saying' is controversial. Most philosophers use the notion of 'what is said' (or 'the proposition expressed') in such a way that it is not a 'pragmatic' notion — having to do with what the speaker means or with what the hearer understands. What is said is supposed to be a property of the sentence (with respect to the context at hand) — a property which it has in virtue of the rules of the language.

Minimalism is closely associated with such a nonpragmatic way of looking at what is said. In the minimalist framework, saturation is the only contextual process allowed to affect 'what is said', because it alone is a bottom-up process, i.e. a process triggered (and made obligatory) by a linguistic expression in the sentence itself. All other contextual processes determine aspects of meaning external and additional to what is said. Take, for example, 'free enrichment' — the process responsible for making the interpretation of an utterance more specific than its literal interpretation (as when 'jumped' is contextually understood as 'jumped over the cliff'). That form of enrichment is 'free' in the sense of not being linguistically controlled. Thus what triggers the contextual provision of the relevant temporal restriction in example (1) ['I've had breakfast'] is not something in the sentence but simply the fact that the utterance is meant as an answer to a question about the speaker’s present state of hunger (which state can be causally affected only by a breakfast taken on the same day). While saturation is a bottom-up, linguistically controlled pragmatic process, free enrichment is a top-down, pragmatically controlled pragmatic process. Insofar as it is pragmatically rather than linguistically controlled, free enrichment is taken to be irrelevant to 'what is said', on the nonpragmatic construal of what is said.

I will discuss the nonpragmatic construal of what is said in Chapter 4. For the time being, I'm interested in the pragmatic construal, based on Grice's idea, and the reasons it provides for rejecting the minimalist constraint (§1.8). Before turning to that issue, however, I want to rebut a couple of objections to the pragmatic construal.

The first objection is this. If, following Grice, we construe saying as a variety of meaning, we will be prevented from acknowledging an important class of cases in which the speaker does not mean what he says. Irony is a good example of that class of cases. If I say 'John is a fine friend' ironically, in a context in which it is obvious to everybody that I think just the opposite, it is clear that I do not mean what I say: I mean the opposite. Still, I say that John is a fine friend. Grice's construal of saying as a variety of meaning prevents him from acknowledging that fact. According to Grice, when I say 'John is a fine friend' in the mentioned situation, I do not really say that John is a fine friend — I pretend to be saying it. The pragmatic construal of saying forces Grice to draw a distinction between 'saying' and 'making as if to say'.

As far as I am concerned, I find Grice's distinction (between genuine saying and making as if to say) perfectly legitimate, but I can understand the worries of those who feel that the notion of 'saying' he uses is too much on the pragmatic, illocutionary side. We certainly need a notion of 'what is said' which captures the objective content of an utterance irrespective of its pragmatic force as a serious assertion or as an ironical utterance. Still, I find the objection superficial, for it is quite easy to actually construct the desired notion within Grice's own framework. Grice uses 'say' in a strict sense. In that sense whatever is said must be meant. But we can easily define a broader sense for 'say':

$$S \text{ says that } p, \text{ in the broad sense, iff he either says that } p \text{ (in the strict sense) or makes}$$

---

26 As I pointed out in footnote 3, p. 00, that is true even when saturation consists in providing a so-called 'unarticulated constituent'.

27 « The verb ‘say’, as Grice uses it, does not mark a [locutionary] level distinct from that marked by such illocutionary verbs as ‘state’ and ‘tell’, but rather functions as a generic illocutionary verb » (Bach, ‘You Don’t Say ?’, p. 41).
as if to say that p (again, in the strict sense of ’say’).

I will henceforth use 'say' in that broad sense, which remains within the confines of the pragmatic construal.

Another objection to the pragmatic construal focusses on the loss of objectivity that allegedly goes with it. What is said is objective in the sense that it is possible both for the speaker to make a mistake and say something other than what he means, and for the hearer to misunderstand what the speaker is saying. Those mistakes are possible, the objector will argue, because what is said is an objective property of the sentence (in context). But on the pragmatic construal, it is not clear that this objectivity can be captured. Imagine the following situation: the speaker wants to say that Paul is tall, and, mistaking Tim for Paul, says 'He is tall' while pointing to Tim. The speaker thus inadvertently says that Tim is tall. Now imagine that the hearer also mistakes Tim for Paul. Thanks to this lucky mistake, he grasps what the speaker means, thinking that this is what he has said. The speaker and the hearer therefore converge on a certain interpretation, which is not objectively what was said, but which they both (mistakenly) think is what was said. How, in the framework I have sketched, will it be possible to dissociate what is actually said from the protagonists' mistaken apprehension of what is said? Have we not equated what is said with their understanding of what is said?

We have not. We have equated what is said with what a normal interpreter would understand as being said, in the context at hand. A normal interpreter knows which sentence was uttered, knows the meaning of that sentence, knows the relevant contextual facts (who is being pointed to, etc.). Ordinary users of the language are normal interpreters, in most situations. They know the relevant facts and have the relevant abilities. But there are situations (as in the above example) where the actual users make mistakes and are not normal interpreters. In such situations their interpretations do not fix what is said. To determine what is said, we need to look at the interpretation that a normal interpreter would give. This is objective enough, yet remains within the confines of the pragmatic construal.

1.8. Availability v. Minimalism

In the framework I have sketched, there is a basic constraint on what is said:

\[ Availability : \]
What is said must be intuitively accessible to the conversational participants (unless something goes wrong and they do not count as 'normal interpreters').

This constraint leads us to give up Minimalism. That is the price to pay if we want Availability to be satisfied.

The reason why Availability is incompatible with Minimalism is simple enough. The aspects of the meaning of (1)-(6) which the minimalist construes as conversational implicatures are, one may admit, contextual ingredients in the overall meaning of the utterance. They do not belong to the conventional meaning of the sentence. The minimalist claims that they do not belong to 'what is said' either, because they are optional: those contextual aspects of the meaning of the utterance are not necessary for the latter to express a complete proposition. But the Availability constraint pulls in the other direction. The very fact that the minimal propositions allegedly expressed are not consciously available shows that it would be a mistake to equate them to what is said; rather, the Availability constraint dictates

\[ 28 \] This is all tacit knowledge, not the sort of 'conscious awareness' I talk about in connection with secondary pragmatic processes.
that the aspects of meaning which Minimalism construes as external to what is said (e.g. the implicit reference to a place in (3), or to the cut in (2), or to a time-interval in (1)) are actually constitutive of what is said, because when we substract them from the intuitive meaning of the utterance the proposition which results is no longer something accessible to the participants in the speech situation. Thus we have two quite distinct phenomena: examples like 'I am French'/ 'I am a good cook' involve something which is said and whose saying implies something else; examples like (1)-(6), in contrast, do not involve the distinction between what is said and what is implied but a different distinction between the literal meaning of the sentence and contextual ingredients entering into the determination of what is said. If we maintain that those ingredients are indeed 'optional' rather than necessary for propositionality, this implies that we must give up the minimalist criterion according to which the context contributes to what is said only when this is necessary for some proposition to be expressed.

According to the view we arrive at, truth-conditional interpretation is pragmatic to a large extent. Various pragmatic processes come into play in the very determination of what is said; not merely saturation — the contextual assignment of values to indexicals and free variables in the logical form of the utterance — but also free enrichment and other processes which are not linguistically triggered but are pragmatic through and through. Figure 1.3 summarizes the contrast between the two conceptions (Minimalism, and the Availability-based approach).

[Figure 1.3 about here]

According to the Availability-based approach, the crucial distinction is not between mandatory and optional contextual processes, but between those that are 'primary' and those that are 'secondary'. Primary pragmatic processes include not only saturation, but also 'optional' processes such as free enrichment. Independent evidence for their inclusion in this category is provided by the fact that, in general, the notion of 'what is said' we need to capture the input to secondary, inferential processes already incorporates contextual elements of the optional variety. Consider examples (1)-(6) once again. In each case we may suppose that the speaker implies various things by saying what she does. Thus, by saying that she's had breakfast, the speaker implies that she is not hungry and does not want to be fed. By saying that the child is not going to die, the mother implies that the cut is not serious; and so forth. Now those implicatures can be worked out only if the speaker is recognized as expressing the (nonminimal) proposition that she's had breakfast that morning, or that the child won't die from that cut. Clearly, if the speaker had had breakfast twenty years ago (rather than that very morning), nothing would follow concerning the speaker's present state of hunger and her willingness or unwillingness to eat something. The implicature could not be derived, if what the speaker says was not given the richer, temporally restricted interpretation. If therefore we accept the Gricean picture, according to which 'what is said' serves as input to the secondary process of implicature-generation, we must, pace Grice himself, acknowledge the non-minimal character of what is said. This provides some support to the Availability-based approach, as against Minimalism.
Chapter 2

Primary pragmatic processes

2.1. Enrichment, loosening, and transfer

Secondary pragmatic processes are 'postpropositional'. They cannot take place unless some proposition \( p \) is considered as having been expressed, for they proceed by inferentially deriving some further proposition \( q \) (the implicature) from the fact that \( p \) has been expressed. In contrast, primary pragmatic processes are 'pre-propositional': they do not presuppose the prior identification of some proposition serving as input to the process.\(^{29}\) Another difference is the fact that secondary pragmatic processes are conscious in the sense that normal interpreters are aware both of what is said and of what is implied and are capable of working out the inferential connection between them. Primary pragmatic processes are not conscious in that sense. Normal interpreters need not be aware of the context-independent meanings of the expressions used, nor of the processes through which those meanings are enriched or otherwise adjusted to fit the situation of use. Unless they are linguists or would-be linguists, they are aware only of the output of the primary processes involved in contextual adjustment.

Saturation is a primary pragmatic process. If the uttered sentence is 'She is smaller than John's sister', then in order to figure out what is said I must (at least) determine whom the speaker refers to by the pronoun 'she' and what the relevant relation is between John and the mentioned sister. Were saturation a secondary pragmatic process, I would have to proceed in reverse order, that is, to identify what is said in order to determine those things.

Beside saturation, which is linguistically mandated (bottom-up), there are, I claim, other primary pragmatic processes that are optional and context-driven (top-down). The paradigm case is free enrichment, illustrated by example (1):

(1) Mary took out her key and opened the door

In virtue of a ‘bridging inference’, we naturally understand the second conjunct as meaning that Mary opened the door with the key mentioned in the first conjunct; yet this is not explicitly articulated in the sentence. Insofar as the bridging inference affects the intuitive truth-conditions of the utterance, it does so as a result of free enrichment.\(^{30}\)

In typical cases free enrichment consists in making the interpretation of some expression in the sentence contextually more specific. This process has sometimes been described in the literature as 'specification'. For example the mass term 'rabbit' will be preferentially interpreted as meaning \textit{rabbit fur} in the context of 'He wears rabbit' and as meaning \textit{rabbit meat} in the context of 'He eats rabbit'.\(^{31}\) This not a matter of selecting a


\(^{31}\) This example is discussed in Nunberg, Geoff and Zaenen, Annie ‘Systematic Polysemy in Lexicology and Lexicography’, in Tommola, Hannu, Varantola, Krista, Tolonen, Tarja, and Schopp, Jürgen (eds.), \textit{Proceedings of Euralex 2} (University of Tampere, 1992). A number of similar examples are discussed
particular value in a finite set; with a little imagination, one can think of dozens of possible interpretations for 'rabbit' by manipulating the stipulated context of utterance; and there is no limit to the number of interpretations one can imagine in such a way. Nor can the process of specification be construed as linguistically mandated, that is, as involving a hidden variable. Were it linguistically mandated (bottom up), it would be mandatory, but it is not: In some contexts the mass term 'rabbit' means nothing more than RABBIT STUFF ('after the accident, there was rabbit all over the highway').

Can free enrichment be equated with specification, or are there instances of free enrichment that are not cases of specification? The provision of (optional) unarticulated constituents is supposed to be a case of free enrichment in which it is not the interpretation of some expression in the sentence that is enriched, but more globally the interpretation of the sentence. In most cases, however, what can be done in terms of unarticulated constituents can also be done in terms of specification. We can construe the implicit instrument in the second conjunct or (1) either as an unarticulated constituent (corresponding to the implicit prepositional phrase 'with the key'), or as an aspect of the interpretation of the predicate 'open' resulting from specification (the concept contextually expressed by 'open' being the specific, ad hoc concept OPEN_WITH_KEY, rather than the generic concept OPEN). The same options are presumably available for dealing with the 'rabbit fur/meat' example. In such cases, I will assume that there is a single form of free enrichment, which can be handled in different frameworks — either in terms of specification (ad hoc concepts) or in terms of unarticulated constituents. Which framework we choose to handle such cases depends upon extraneous considerations. (For example, if we want to preserve the principle of compositionality, we'd better opt for the specification view which spares us the postulation of syntactically unarticulated constituents.) Still, there is a type of case for which I think we need the notion of unarticulated constituent and cannot make do with specification and ad hoc concepts: whenever the alleged unarticulated constituent is the intended 'circumstance of evaluation', we can't deal with it in terms of specification or ad hoc concept. That type of case will be discussed in chapter 8.


Another issue regarding enrichment is whether or not it can be described as 'strengthening' or logical enrichment, as I suggested in *Direct Reference* (p. 261). A predicate has conditions of application, and strengthening consists in restricting the application of a predicate by contextually providing further conditions that are not linguistically encoded. Thus 'table' has such and such conditions of application packed into the concept TABLE, and through contextual strengthening the further condition IN_THE_LIVING_ROOM is provided, which results in a restricted application. Thus construed enrichment can account for the (so-called) contextual restriction of quantifiers and for the interpretation of (so-called) 'incomplete' definite descriptions. ('All the books are on the table', where a particular set of books and a particular table are in question).33

The converse of enrichment is loosening.34 There is loosening whenever a condition of application packed into the concept literally expressed by a predicate is contextually dropped so that the application of the predicate is widened. An example is 'The ATM swallowed my credit card'. There can be no real swallowing on the part of an ATM, since ATMs are not living organisms with the right bodily equipment for swallowing. By relaxing the conditions of application for 'swallow', we construct an ad hoc concept with wider application.

A third type of primary pragmatic process that is not linguistically mandated (bottom up) but contextually driven is semantic transfer.35 In transfer the output is neither an enriched nor an impoverished version of the concept literally expressed by the input expression. It's a different concept altogether, bearing a systematic relation to it. Thus 'parked out back' denotes either the property a car has when it is parked out back, or a different property, namely the property a car-owner has whenever his or her car has the former property ('I am parked out back'). Arguably, 'parked out back' literally denotes the former property, and comes to denote the latter property as a result of transfer. Similarly, the expression 'ham sandwich' in 'The ham sandwich left without paying' arguably denotes, through transfer, the derived property HAM_SANDWICH_ORDERER rather than the linguistically encoded property HAM_SANDWICH.

A number of problems arise in connection with primary pragmatic processes. Do we really need the four categories I have mentioned? It may be argued that we need less categories, or that we need more. In particular, if we construe enrichment not as logical strengthening but as a quasi-syntactic process of 'expansion' through which a representation

33 Stephen Neale (*This, That and the Other* [forthcoming], chapter 1) objects to my notion of strengthening:
It is sometimes said that enrichment in Sperber and Wilson’s sense involves strengthening in that the post-enrichment statement entails the statement one might associate with the utterance prior to enrichment. That this cannot be true in general is made clear by cases involving (non-persistent) determiners like “every”, “no” and “some”. If, when discussing a dinner party I attended last night, I say, “everyone drank a lot of wine”, I am saying only that everyone who attended the dinner party drank a lot of wine, and this does not entail that everyone (unqualified) drank a lot of wine. But I had already disposed of that objection in *Direct Reference* (pp. 262-3): “Strengthening, it may be argued, can be understood as operating locally. For example, in the case of ‘Every boy came’, we might say that it is the predicate ‘boy’ that is strengthened into ‘boy in the class’, rather than the proposition ‘Every boy came’ into ‘Every boy in the class came’.” (See also ‘The Pragmatics of What is Said’, p. 307 n.)
(or a structured content) is made more complex by the addition of further constituents, then it may be that we don't need transfer, nor even loosening, but that we can handle everything in terms of that (little constrained) notion of expansion. Be that as it may, I will not deal with those problems, internal to the theory of primary pragmatic processes, in this chapter. What I am concerned with is a more basic set of issues pertaining to the very distinction between primary and secondary pragmatic processes.

2.2. Rejecting the Gricean picture

Even though he construed saying as a variety of nonnatural meaning, Grice espoused Minimalism. On his view, disambiguation and saturation suffice to give us the literal interpretation of the utterance — what is literally said. All other pragmatic processes involved in the interpretation of the utterance are secondary and presuppose the identification of what is said. Interpretation is construed as a two-step procedure: (i) The interpreter accesses the literal interpretations of all constituents in the sentence and uses them to compute the proposition literally expressed, with respect to the context at hand; (ii) on the basis of this proposition and general conversational principles he or she infers what the speaker means (which may be distinct from what is said, i.e. from the proposition literally expressed).

The picture I have presented also makes interpretation a two-step procedure, but there is a major difference. The primary pragmatic processes that are involved in determining what is said include not only saturation (and disambiguation) but also optional processes such as free enrichment, loosening and semantic transfer. Those processes take us from the literal meaning of some constituent (the meaning that is linguistically encoded, or that which results from saturating the linguistically encoded meaning) to a derived meaning which may be richer, poorer, or involve some kind of transfer. I hold that, for such processes to take place, there is no need to antecedently compute the proposition literally expressed. That is why I take those processes to be 'primary', like saturation. What I am disputing therefore is the claim that the process of semantic composition which consists in putting together the semantic values of the parts to determine the semantic value of the whole begins by paying attention only to literal semantic values (as delivered through disambiguation and saturation), and turns to derived values only after the literal semantic value of the whole (the proposition literally expressed) has been computed.

There is a simple argument purporting to show that the Gricean picture must be right. Enrichment, loosening and transfer all take as input the literal meaning of some expression; hence they cannot take place unless that literal meaning has been accessed. This seems to support the Gricean picture, according to which we process the literal interpretation first, and move on to the derived interpretation only when this is required to make sense of the speaker's utterance.

This argument is fallacious.\(^36\) I admit that the literal interpretation must come first, insofar as the derived interpretation is derived from it through enrichment, loosening, or transfer. But I want to resist the conclusion that the literal interpretation must be 'processed' first, in the sense that is relevant to the debate. Or, to put it another way: What I am rejecting is not the claim that the literal interpretation of the constituent is accessed before the derived interpretation — that I take to be obvious — but the claim that a similar priority holds at the level of the complete sentence; that is, I reject the claim that the process of semantic

composition begins by paying attention only to literal semantic values, and turns to derived values only after the literal semantic value of the whole (the proposition literally expressed) has been computed. It is this picture which I think is unwarranted.

If I am right, the asymmetric dependence of derived meaning upon literal meaning does not rule out an account according to which literal meaning and derived meaning are on equal footing as far as semantic composition is concerned. Thus we can imagine that the literal meaning and the derived meaning of a given expression are processed in parallel, in constructing an interpretation for the whole utterance. In the model I have in mind, the literal meaning of the expression is accessed first and triggers the activation of associatively related representations. That literal meaning is a natural candidate for the status of semantic value, but there are others: some of the representations activated by association contribute further candidates for the status of semantic value. All candidates, whether literal or derived, are processed in parallel and compete. When an interpretation which fits the broader context of discourse is found, it is selected (i.e., it undergoes semantic composition) and the other candidates are suppressed.

On this view, derived meanings still proceed (associatively) from literal meanings, which they indeed presuppose; but, although generated serially, they are processed in parallel. The literal meaning has no compositional privilege over derived meanings; they compete and it is possible for some derived meaning to be retained (if it fits the broader context of discourse) while the literal interpretation is suppressed. In other words, the derived interpretation is associatively derived from the literal interpretation, but it is not inferentially derived. Inferential derivation entails computation of the literal value of the global sentence, while associative derivation is a 'local' process which does not require prior computation of the proposition literally expressed.

Consider, as an example, Geoff Nunberg's famous ham sandwich. The waiter says 'The ham sandwich has left without paying'. On the Gricean picture the interpreter computes the proposition literally expressed by the sentence — namely the absurd proposition that the sandwich itself has left without paying — and from its absurdity infers that the speaker means something different from what she says. On the parallel model I have outlined the description 'the ham sandwich' first receives its literal interpretation, in such a way that a representation of a ham sandwich is activated; activation then spreads to related representations, including a representation of the man who ordered a ham sandwich. All these representations activated by the description 'the ham sandwich' contribute potential candidates for the status of semantic value of the expression; all of which are equally susceptible of going into the interpretation of the global utterance. Now the ham sandwich orderer is a better candidate than the ham sandwich itself for the status of argument for '... has left without paying'. It is therefore the derived, non-literal candidate which is retained, while the literal interpretation is discarded.

37 By ‘semantic value’ here I mean what the process of semantic composition operates on. My point is that the semantic value contributed by an expression need not be the literal meaning of that expression — it may be affected by pragmatic processes, including optional pragmatic processes. (Later in this book I will talk of ‘pragmatic value’ instead of ‘semantic value’ to stress the fact that such values result from the operation of such processes.)
38 See my Direct Reference, pp. 263-6.
39 As Ivan Sag (‘Formal Semantics and Extralinguistic Contexts’, in Cole, Peter [ed.], Radical Pragmatics [Academic Press, 1981], 273-94) and Geoff Nunberg (‘Transfers of Meaning’) pointed out, it is not the description as a whole, but the predicate ‘ham sandwich’ in the description which has a derived, non-literal value; the ham sandwich example is a case of property transfer, not a case of deferred reference (see footnote 47, p. 00). Since nothing hinges on this point, I will simplify matters in this and the next two sections by treating the description itself as a unit susceptible to both a literal and a derived interpretation.
An important difference between the Gricean model (according to which the literal interpretation is processed first) and the parallel model just outlined is this: on the parallel model it is possible for an utterance to receive a non-literal interpretation without the literal interpretation of that utterance being ever computed. The non-literal interpretation of the global sentence does not presuppose its literal interpretation, contrary to what happens at the constituent level. If the non-literal interpretation of some constituent fits the context especially well it may be retained (and the other interpretations suppressed) before the literal interpretation of the sentence has been computed. Whether or not this sort of thing actually happens, this is at least conceivable, on the parallel model.

2.3. Accessibility

When someone talks of 'wearing rabbit', the literal meaning of the mass term 'rabbit' (namely RABBIT STUFF) is accessed, but it has to compete with other candidates for semantic value. The more specific representation RABBIT FUR is also activated since it is associatively connected to the representations encoded by both 'rabbit' and 'wear'. As a result of this multiple activation, it is possible for the representation RABBIT FUR to be more active, in this context, than the less specific representation RABBIT STUFF which is linguistically encoded. Whatever we think of this particular example, it seems to me that the following situation can obtain: An expression linguistically encodes a certain representation; that representation becomes active when the expression is uttered, but another, associatively related representation is also activated (in part — but in part only — through the encoded representation) and turns out to be more active in that context than the original representation from which (in part) it derives.40 If we assume that the candidate for semantic value which is retained and undergoes semantic composition is that which is most accessible, i.e. that which corresponds to the most active representation when the interpretation process stabilizes, we explain how a derived meaning resulting from enrichment (or loosening, or transfer) can be selected as semantic value, in lieu of the literal meaning from which it is derived.41

The same phenomenon can happen dynamically. Some representation may be activated through its associative links to the representation linguistically encoded and become more active than the latter as a result of the coming into play of further linguistic material which raises its activation level. Before considering such a case, let us look at a simpler example of accessibility shift along the temporal dimension.

Consider sentence (2) from Direct Reference (p. 265):

(2) John was arrested by a policeman yesterday; he had just stolen a wallet.

In order to interpret the utterance one must assign a reference to the pronoun 'he' in the second sentence. The two persons who were mentioned in the first sentence, namely John and the policeman, are obvious candidates, and we may suppose that there is no other (sufficiently

40 I assume that representations can be activated because they are being processed (or have been recently), or because they receive activation from associatively related representations which are themselves activated. Frequency of processing is another important factor (among several others); it can be conceived of as lowering the activation threshold of a representation. See e.g. Barsalou, Lawrence and Billsman, Dorrit ‘Systematicity and Semantic Ambiguity’, in Gorfein, D. (ed.), Resolving Semantic Ambiguity (Springer, 1989), 146-203.

accessible) candidate. (There would be one if, for example, the speaker pointed to someone while uttering the pronoun.) We may suppose that when the pronoun 'he' is uttered one of the two candidates, John or the policeman, is more accessible than the other; John because he is foregrounded (qua subject of the sentence), or the policeman because he was mentioned last. For the sake of the argument, let us suppose that the second factor is more important than the first one, so that the policeman is slightly more accessible than John when the pronoun is uttered. This may well change when the other constituents of the sentence are processed. When the predicate 'had just stolen a wallet' is uttered, John becomes more accessible than the policeman as a candidate for the status of referent of 'he', even if the policeman was more accessible at an earlier stage in the processing of the utterance. For John is the subject of 'was arrested' and therefore occupies the role of the person being arrested; now that role is linked to the role of the person doing the stealing, in some relevant frame. Because of this link, the representation of the referent of 'he' as the person doing the stealing contributes some activation to the representation of the person being arrested and therefore raises the accessibility of John qua occupier of this role. John thus becomes the most accessible candidate.

If we turn to a case in which some candidate for semantic value is distinguished from the others by being literally encoded, we see that the same sort of temporal shift in accessibility can occur. When the words 'the ham sandwich' are uttered, we may consider that the representation of the ham sandwich is more active than other, related representations which are activated through their links to that representation. Thus we may suppose that the representation of the ham sandwich is more active than the 'derived' representation of the ham sandwich orderer. This is similar to the fact that the policeman is more accessible than John when the pronoun 'he' in (2) is uttered, except that the policeman is (perhaps) more accessible because he was mentioned last, while the representation of the ham sandwich is more active because it is linguistically encoded and has some form of priority over the ham sandwich orderer (derived value). In both cases, the initial ranking is reversed when further linguistic material comes into play. After the predicate in the sentence 'The ham sandwich has left without paying' has been processed, the ham sandwich is no longer a more accessible candidate than the ham sandwich orderer — the order of accessibility is reversed. The explanation, again, is very simple and does not appeal to inference on the hearer's part. The predicate 'has left without paying' demands a person as argument; this raises the accessibility of all candidates who are (represented as) persons. In this way the representation of the ham sandwich orderer gains some extra activation which makes him more accessible than the ham sandwich, after the predicate has been processed.

2.4. Objections and responses

In the framework I have sketched the interpretation which eventually emerges and incorporates the output of various pragmatic processes results from a blind, mechanical process, involving no reflection on the interpreter's part. The dynamics of accessibility does everything, and no 'inference' is required. In particular, there is no need to consider the speaker's beliefs and intentions.

Dan Sperber objected to me in conversation that accessibility can lead us astray: Sometimes the first interpretation that comes to mind (the most accessible one) turns out not

---

to be satisfactory and forces the hearer to backtrack. According to Sperber, the possibility of such garden-path effects shows that success, for a candidate semantic value, cannot be equated with sheer accessibility. This objection is misguided, I think. The most accessible interpretation at some stage $s$ in the interpretation process may well turn out to be unsatisfactory at some later stage $s'$, thereby resulting in a garden path effect and the need to backtrack. This does not show that interpretational success cannot be cashed out in terms of accessibility. At any given stage, the most accessible interpretation will be the winning one (at that stage). In garden path utterances we have two successive stages to consider. Some interpretation is the most accessible one, hence wins, at $s$, but that interpretation fails to fit some schema, hence loses, at a later stage $s'$. In an accessibility-based framework, this means that this interpretation's accessibility at $s'$ is no longer sufficient for it to be the winning candidate (at $s'$). Another candidate (which was less accessible at $s$, but turns out to be more accessible at $s'$) takes over, hence the garden path effect. The distinction between successive stages of interpretation, together with the notion of an accessibility shift, is sufficient to account for garden path effects within the accessibility-based framework.

The notion of an accessibility shift also provides immediate answers to the questions which, according to Kent Bach, my account of the ham sandwich metonymy and similar phenomena in terms of local pragmatic processes raises. Bach writes:

Recanati supposes that the process of metonymical transfer takes place without the intrusion of a thought of the absurd proposition associated with the literal meaning of (i) ['The ham sandwich is getting restless']. That is, the hearer does not have to compute that the speaker does not mean that a certain culinary item is getting restless in order to determine what the speaker does mean. But, I ask, how can the hearer go from the concept of ham sandwich to that of ham-sandwich-orderer without first entertaining the absurd minimal proposition? What triggers the 'local process' and, for that matter, keeps it from being triggered in a cases like (ii) ['The ham sandwich is getting eaten'], uttered in similar circumstances? Recanati's account predicts that the hearer would entertain the proposition that the ham-sandwich-orderer is getting eaten, since the local process it posits would get triggered before a full proposition were reached. And yet (ii) could be understood perfectly well.

But my account does not predict that the hearer would understand (ii) ('The ham sandwich is being eaten') as being about the ham sandwich orderer. There is something, in my account, which keeps the local process of transfer from being triggered in a case like (ii). What keeps the local process of transfer from being triggered in a case like (ii) is precisely what answers Bach's legitimate question: 'How can the hearer go from the concept of ham sandwich to that of ham-sandwich-orderer without first entertaining the absurd minimal [i.e. literal] proposition?' In my framework, the process of metonymical transfer is 'local', not global, yet it is sensitive to the linguistic (and extralinguistic) context in which the expression which receives the metonymical interpretation occurs. In particular, it is sensitive to the meaning of the predicate expression; this is why it is triggered in (i) but not in (ii).

For the literal proposition to be entertained, the literal semantic value of '(the) ham sandwich' must undergo semantic composition with the semantic value of the predicate; but as soon as the semantic value of the predicate is accessed, the metonymical interpretation of the subject-term becomes more accessible than its literal interpretation, which is therefore prevented from going into the proposition expressed. The activation of the concept which

---

44 Ibid., p. 158.
corresponds to the literal interpretation of the predicate is both what triggers the process of metonymical transfer in (i) and what prevents it from taking place in (ii). Yet the transfer's sensitivity to the interpretation of the other constituents in the sentence does not mean that the transfer is a global process involving the computation of the absurd, literal proposition. The interpreter does not go from the concept of ham sandwich to that of ham-sandwich-orderer after having entertained the absurd literal proposition; rather, it is because the interpreter goes from the concept of ham sandwich to that of ham-sandwich-orderer (as a result of an accessibility shift resulting from the interpretation of the predicate) that he or she does not entertain the absurd literal proposition.

In this account, the metonymical transfer from ham sandwich to ham-sandwich-orderer results from a shift in accessibility triggered by the interpretation of another constituent in the sentence. The literal interpretation of the subject-term was more accessible than the metonymical interpretation before the predicate came into the picture, but the metonymical interpretation becomes more accessible as a result of interpreting the predicate. This accessibility shift has two consequences. First, the literal value of the subject term in (i) does not undergo semantic composition with that of the predicate, precisely because the semantic value of the predicate makes the literal value of the subject term less accessible than some other, non-literal interpretation. Second, if the sentence contained another predicate which did not require a person, but rather a culinary item, as argument, then the accessibility shift would not occur and the description 'the ham sandwich' would be given the literal interpretation; this account for the difference between (i) and (ii).

2.5. Interactive processing

So far I have considered the interpretation of a given constituent as if the interpretations of the other constituents were fixed; but nothing is fixed. All constituents can be given derived as well as literal interpretations, and how we interpret a constituent cannot but affect how we interpret the others. Consider (3), for instance.

(3) The city is asleep.

A city is not the sort of thing that sleeps (in the normal sense of the term), hence the overall interpretation of (3) is likely to involve some process of non-literal interpretation. There are various possibilities, though. 'The city' can be interpreted metonymically as standing for 'the inhabitants of the city' (transfer); or 'asleep' can be interpreted metaphorically in the sense of 'quiet and showing little activity' (loosening). Given the likelihood of a non-literal interpretation for some constituent, if 'the city' literally applies to a city, 'asleep' will have to be taken non-literally; conversely, if 'asleep' is literal, 'the city' will not be.

The same sort of trade off applies to seemingly unproblematic examples such as the following:

(4) I finished the book.
(5) John heard the piano.

If 'finish' has its standard sense in (4), its object (what is said to be finished) must be a process, for only processes can start or finish in the standard sense. It follows that 'the book' must be interpreted non-literally — it must be interpreted as standing, not for a certain book, but for the process of, say, reading a certain book. Conversely, if 'the book' is interpreted literally and stands for a certain book, 'finish' must have a derived sense and mean something
like *finished reading*. Pustejovsky defends the former analysis, Langacker the latter.\(^{45}\)

(5) is another example (discussed by Langacker) which can be interpreted in two ways. We may construe 'the piano' as standing for the *sounds emitted by the piano* (metonymy), in accordance with the 'postulate' that only sounds can be heard; or we may consider 'hear' as polysemous. When it is said that 'only sounds can be heard', 'hear' is taken in its basic sense (\(\text{HEAR}_1\)), but there is another sense (\(\text{HEAR}_2\)), which can be defined as follows: An object is heard\(_2\) whenever the sound it emits is heard\(_1\). Langacker analyses (5) as saying that John heard\(_2\) the piano, rather than as saying that he heard\(_1\) the sounds emitted by the piano.

Though (3), (4) and (5) exhibit the same sort of trade off, there is an important difference between (3), on the one hand, and (4) and (5) on the other hand. Depending on which constituent receives a non-literal interpretation, the global interpretation of (3) itself will vary. In one case the speaker means that the inhabitants of the city are sleeping, in the other she means that the city itself is quiet and shows little activity; the truth-conditions are distinct. It follows that it is easy to say which constituent is responsible for the non-literal interpretation of the sentence: we have only to consult our intuitions concerning the truth-conditions of the utterance. Examples (4) and (5) are very different, however. Whether one constituent or the other bears the onus of non-literal interpretation, the overall interpretation of the utterance does not seem to change. Hence it is not easy to establish which analysis is right, and one must appeal to indirect evidence, such as, for example, the fact that we can say things like: 'I can both hear and touch the piano'. (This fact tends to support Langacker’s view: when we say that we hear the piano, we are referring to the piano. It is the verb, not the noun-phrase, which has a derived sense in this context.)\(^{46}\)

The trade off talked about in this section is a particular case of a more general and fairly complex phenomenon: the search for coherence in interpretation. The examples involve two constituents, A and B, each with (at least) two possible interpretations (\(A_i\) and \(A_j\), and \(B_i\) and \(B_j\)). Now there is a better 'fit' between \(A_i\) and \(B_j\) or between \(A_j\) and \(B_i\) than there is


\(^{46}\) A number of similar tests are offered by Geoff Nunberg in ‘Indexicality and Deixis’ (in *Linguistics and Philosophy* 16 [1993], 1-43) and ‘Transfers of Meaning’. For example, Nunberg argues that (a) and (b) must be analysed in fairly different ways, despite their superficial similarity :

(a) I am parked out back.
(b) This one <pointing to a car key> is parked out back.

Example (b) is a case of 'deferred ostension': an object (the key) is 'demonstrated', but the ultimate 'referent', as opposed to the 'demonstratum', is another object suitably related to the demonstratum, namely the car itself. In contrast, 'I' does not refer to the car in (a); it normally refers to the speaker. It's the *predicate* 'parked out back' that bears the onus of non-literality. That the referent is the car rather than the (demonstrated) key in (b) is established by the fact that semantic features of the referring expression such as number or gender do NOT correspond to the key, but to the car; thus we would say 'this one' (in the singular) even if we were exhibiting a whole set of keys for the car in question. Also, we can say things like:

(c) This <pointing to a car key>, which is an old Chevrolet, is parked out back.
The relative clause clearly qualifies the car, not the key. These tests provide evidence that the referent is the car, even if it is the key which is 'demonstrated'. When we turn to (a), the situation is totally different. Gender and number correspond to the speaker, not to the car. We would not say 'We are parked out back' if there was a single car owner and several cars (whereas we can say 'Those <pointing to a key> are parked out back' if there is a single key and several cars). And we can't say 'I, who am an old Chevrolet, am parked out back'.

26
between either \( A_i \) and \( B_i \) or \( A_j \) and \( B_j \), hence the 'trade-off' effect: the two global interpretations \(<A_i, B_j>\) and \(<A_j, B_i>\) are favoured, in such a way that the choice of either \( A_j \) or \( A_i \) as the correct interpretation for \( A \) 'coerces' a particular interpretation \((B_i \text{ or } B_j)\) for \( B \). The tendency to prefer coherent interpretations (with a high degree of fit between the various semantic values) is what we must now try to account for. But first, we must introduce the notion of a schema.\(^{47}\)

2.6. The role of schemata

To say that two semantic values \( A \) and \( B \) 'fit together' is to say that there is an abstract schema which \(<, >\) instantiates. The particular case discussed above, 'coercion', can be analysed as follows. An expression \( E \) activates an abstract schema in which there is a slot for a value of a certain type; as a result, the semantic value of \( E \) will preferably enter into composition with a semantic value of the relevant type. Thus in a sentence like 'The city is asleep', if we give to 'asleep' its literal value (thereby activating the SLEEP schema), the value of 'the city' will have to be of the relevant type (e.g. human or animal), hence non-literal.

The role of schemata in interpretation is best seen in connection with more complex examples. Consider example (2) once again:

(2) John was arrested by a policeman yesterday; he had just stolen a wallet.

As I said earlier, John is selected as the referent for 'he' in the second clause because (i) the referent of 'he' is said to have stolen and John is known to have been arrested, and (ii) there is a frame or schema in which the two roles (stealing, and being arrested) are linked. This schema is jointly activated by the predicates 'was arrested' and 'had stolen'. An interpretation in which the same person steals and is arrested (and in which he is arrested because he has stolen — see below) satisfies the schema, and is more likely to be selected than one which violates it.

Given the link we have established between coherence (or fit) and schemata, the question we must answer becomes: Why are schema-instantiating interpretations more successful than others? In line with what has been said so far, the following answer suggests itself. Interpretational success — what brings a 'candidate' or potential semantic value into the actual interpretation of the utterance — to a large extent depends on the candidate's accessibility or degree of activation. Now a schema is activated by, or accessed through, an expression whose semantic value corresponds to an aspect of the schema. The schema thus activated in turn raises the accessibility of whatever possible semantic values for other constituents of the sentence happen to fit the schema. The schema itself gains extra activation from the fact that some other constituent of the sentence has a possible interpretation which fits the schema. In such a case all 'candidates' or potential semantic values which fit the schema evoked by some of them mutually reinforce their accessibility and therefore increase

the likelihood that they will be globally selected as part of the interpretation of the utterance. Coherent, schema-instantiating interpretations therefore tend to be selected and preferred over non-integrated or 'loose' interpretations. As a result, schemata drive the interpretation process.

The role played by schemata explains why the process of utterance interpretation is to such a large extent top down and driven by world knowledge. The interpreter unconsciously enriches the situation described by the utterance with many details which do not correspond to any aspect of the uttered sentence but are contributed in order to fit an evoked schema. Thus, in the policeman example above, not only is the reference of 'he' determined by the STEAL (x) - IS ARRESTED (x) schema, but (among other things) the causal interpretation of the relation between the two events mentioned — the fact that the referent was arrested because he had stolen — is also determined by the schema.

48 Think of an example like (6):
(6) John hates the piano.
Contrary to (4)-(5), this example cannot be described in terms of 'coercion'. I mentioned earlier various contentions to the effect that only sounds can be heard and only processes finished. By virtue of these principles, some process of non-literal interpretation must take place in order to make sense of phrases such as 'finish the book' or 'hear the piano'. But a piano is certainly an object that can be hated, however strictly one construes the predicate 'hate'. Still, some contextual enrichment is in order, because to hate the piano is to hate it under some aspect or dimension. One may hate the sounds emitted by the piano, or one can hate playing the piano, or one can hate the piano as a piece of furniture, etc. The relevant dimension is contextually provided through the process of enrichment.