The *how and why of polysemy: A pragmatic account*\(^1\)

Ingrid Lossius Falkum, University of Oslo


\(^1\)This research was supported by a postdoctoral fellowship from the Research Council of Norway (project no. 205513). I would like to thank Robyn Carston for many valuable and inspiring discussions of the topic of the present paper, and three anonymous reviewers for their thoughtful and constructive comments which helped improve the paper.
The *how and why* of polysemy: A pragmatic account

*Abstract*

A large number of word forms in natural language are polysemous, that is, associated with several related senses (e.g., *line*, *run*, *tight*, etc.). While such polysemy appears to cause little difficulty in verbal communication, it poses a range of theoretical and descriptive problems. One concerns its very existence: What is it about our language systems that make them so susceptible to polysemy? In this paper I discuss two approaches to polysemy with different answers to this question: (i) A code-based approach that treats polysemy in terms of the operation of lexicon-internal generative rules, and (ii) an inference-based approach that takes polysemy to be governed by pragmatic inferential processes applying at the level of individual words. After evaluating how each of these accounts fares with respect to some empirical data, I look more broadly at their implications for the emergence and development of polysemy; I conclude that, overall, the pragmatic approach provides the most promising basis for a unified account of the role of polysemy in several domains, and for explaining what motivates its proliferation natural language.

*Keywords:*

Polysemy, pragmatics, rule-based accounts, lexical underdeterminacy, Relevance Theory
1. Introduction

A large number of natural language word forms are polysemous, that is, associated with several related senses. Here are some illustrations:

1. Kate began a book.
2. Jane Austen wrote good books.
3. a. There was rabbit all over the highway.
   b. Steven had rabbit for dinner.
   c. The model wore rabbit on the catwalk.
4. John is a lion.
5. The ham sandwich is getting impatient.

In (1), the VP began a book is compatible with several readings, each involving a different event (e.g., reading/writing/mending/dusting/ripping up... etc.). It is common to refer to constructions of this type as ‘logical metonymy’ (Pustejovsky, 1995), where the argument of the verb in syntax is different from that argument in ‘logical form’: In (1), the idea would be that the NP a book denotes part of an event, which is used to stand for the event as a whole (e.g., the VP ‘reading a book’). The interpretation of the adjective good in (2) requires a specification of its conceptual content (e.g., ‘good reads’), which would be different from the one it has in, e.g., good knife/football player/weather/child, and so on (Katz, 1964). The alternations between the different senses of the noun rabbit in (3) – ‘rabbit remains’, rabbit meat’, ‘rabbit fur’ – is standardly analysed as a form of systematic (or regular) polysemy (Apresjan, 1974) where the related senses of a word are predictable on the basis of the ontological category of its denotation (cf. other animal terms; crocodile, seal, mink, lamb, etc.). Finally, in (4) and (5), we have a metaphorical use of the noun lion (‘strong, courageous, takes risks’, etc.), derived on the basis of properties associated with the animal denotation (e.g., Glucksberg, 2001), and a metonymic use of the NP The ham sandwich (‘the person who ordered the ham sandwich’), based on a contextually salient association between the ham sandwich and the person who ordered it (Nunberg, 1979). Such metaphorical and metonymical extensions are usually taken to be prime sources of polysemy in language.
Already Bréal (1924 [1897]) noted that when talking to each other we rarely get confused by the multiplicity of meanings that a word can have. Sometimes, of course, speakers may exploit the polyseous potential of a lexical item to create confusion or a humorous effect (e.g., by use of a pun), but generally, polysemy causes little difficulty for users of a language; it is something that we handle effortlessly and unconsciously, most of the time. As speakers, we can trust hearers to quickly and reliably figure out the lexical meanings we intend to communicate on a given occasion.

In contrast, polysemy raises a host of theoretical and descriptive problems. A first issue concerns the delimitation of the polysemy phenomenon. How – if at all – should it be distinguished from the accidental multiple encoding we find in homonymy (e.g., \textit{bank, coach}), on the one hand, and contextually modulated senses (e.g., \textit{good weather/good student/good book}), derived from a single encoded meaning, on the other hand? Several tests for distinguishing between ambiguity (homonymy/polysemy) and non-specificity (monosemy) have been proposed (e.g., Goddard, 2000; Kempson, 1977; Quine, 1960), but as Geeraerts (1993) has meticulously shown, different tests may not always agree with each other and by manipulating the context, they can be made to yield inconsistent results.

A second taxing issue is how polyseous lexical items are represented in the mental lexicon. According to ‘sense enumeration lexicons’, all the different senses of a lexical item are stored under a single entry, and comprehension involves selection of the contextually appropriate sense among the list of candidates (e.g., Brugman & Lakoff, 1988; Katz, 1972). Given the proliferation of polysemy, a problem for such fully encoding lexicons is that they would have to store indefinitely many semantic distinctions for each lexical item. So-called

---

2 This discrepancy has led some scholars to talk of a ‘polysemy paradox’ (Ravin & Leacock, 2000; Taylor, 2003).
3 In lexicography, this translates into a methodological issue of determining which cases of multiple encodings that should be listed as distinct entries (homonymy) and which should be listed under a single entry (polysemy). For any single entry, there are also decisions to be made concerning which senses are established (hence should be listed) and which are not (hence should not be listed).
4 By ‘mental lexicon’, I refer to individual speakers’ stable mental representations of words, which include information regarding their semantic properties (in the form of ‘meanings’ or ‘senses’), as well as phonological and syntactic properties, which are accessed when a word is encountered in discourse.
'core meaning approaches' take polysemy to be represented in terms of a single, maximally general meaning, from which the appropriate one is contextually derived (e.g., Caramazza & Grober, 1976; Ruhl, 1989). A challenge for this kind of approach is to determine the appropriate level of abstraction for defining the core meaning of a lexical item, in order to capture what is common to all its possible, sometimes radically different uses. In between these two extreme views we find several 'middle-ground approaches', assuming only a limited number of stored senses of a lexical item and the rest to be derived in context (e.g., Carston, 2002; Tyler & Evans, 2003), sometimes combined with a proposal for a principled set of criteria for distinguishing between those senses that are stored in the lexicon and those that are constructed on-line in language use (Tyler & Evans, 2003). This kind of approach, involving a differential representation of polysemous senses (some may be stored, some may be contextually derived) is supported by experimental evidence (e.g., Klepousniotou, 2007; Klepousniotou, Titone, & Romero, 2008; Pylkkänen, Llinás, & Murphy, 2006), although the results are to some extent conflicting (cf. Klein & Murphy, 2001).

A third issue is how lexical meanings get extended into several different meanings. Assuming that some (possibly many) senses of polysemous lexical items are derived during on-line processing, what is the nature of the processes or mechanisms involved? In other words, what kind of phenomenon is polysemy? Is it mainly a result of the operation of lexical rules for sense extension (e.g., Copestake & Briscoe, 1995; Ostler & Atkins, 1992; Pustejovsky, 1995)? Is it a direct reflection of how our cognitive categories are structured more generally (e.g., Brugman, 1988; Brugman & Lakoff, 1988; Lakoff, 1987)? Or does it arise through pragmatic processes operating over underspecified lexical meanings and contextual knowledge (e.g., Carston, 2002; Sperber & Wilson, 1998)?

A fourth issue that the proliferation of polysemy in natural language raises is the fundamental question of why it exists. Why are word meanings extended in this way? What is it about our language systems – specifically their lexical component – that make them so susceptible to polysemy?
The focus of this paper will be these two last issues: the how and why of polysemy. I will consider two opposing views regarding the nature of the polysemy phenomenon – rule-based vs. ‘radical’ pragmatic approaches – with very different answers to the questions of how linguistic polysemy arises, and what its underlying motivation may be. After evaluating how each of the accounts fares with respect to explaining the polysemy data exemplified in (1) through (5), I discuss their implications for the emergence and development of polysemy. I conclude that, overall, the radical pragmatic account provides the most promising basis for a unified account of the role of polysemy in several domains, and for explaining what motivates its proliferation in natural language. As regards the first two issues concerning the delimitation of polysemy and its representation in the mental lexicon, I will assume, based on extant experimental evidence, a differential representation of polysemy (where, depending on their degree of conventionalisation, some senses may be stored in our mental lexicons, some may be contextually derived), and operate with an intuitive distinction between conventional (i.e., ‘encoded’ or ‘semantic’) polysemy and contextually-derived polysemy, acknowledging that there may be no clear-cut way of drawing this distinction.

2. Two approaches to polysemy

A fundamental difference between rule-based and pragmatic approaches to polysemy lies in their radically different conceptions of what a language is. Underlying rule-based approaches is the view that language provides an information-rich code that enables speakers and hearers to encode and decode their thoughts in much detail, with pragmatics as a useful add-on to this

---

5 The cognitive linguistic hypothesis about underlying conceptual mappings being the source of linguistic polysemy (e.g., Lakoff, 1987) will not be considered further in this paper. Although the hypothesis may indeed be correct for some, even many, cases of conventional polysemy, it leaves open the question of how new polysemous senses are constructed as a result of communicative interactions between speakers of a language, which is the main concern of this paper.

6 One reviewer pointed out that this seems to imply that there is a continuum rather than a proper distinction between conventional and contextual polysemy, and that this would present a problem for the view that there is a distinction between encoded word meanings and unencoded (inferred) senses. One solution might be to abandon the assumption that words encode concepts and instead see them as encoding underspecified (non-conceptual, abstract schematic) meanings, a position which is currently being pursued by Carston (2012, 2013) within the relevance-theoretic framework. While I am generally sympathetic to this view, it would exceed the scope this paper consider it further here.
linguistic capacity, operating primarily when some interpretation other than the linguistic default was intended. By contrast, radical pragmatic accounts see the role of the linguistic system as being that of providing a minimal input or clue – a ‘sketch’, or ‘blueprint’ of the speaker’s meaning – which the pragmatic inferential system uses as evidence to yield hypotheses about occasion-specific, speaker-intended meanings. In this sense, we may call the first a code-based approach, and the second an inference-based approach. In what follows, I will consider each of them in turn.

2.1 Code-based approaches: polysemy as lexical rules

Before Grice ([1967] 1989), virtually all theories of communication were based on the so-called code model. The guiding assumption of this model is that communication is a matter of encoding and decoding of messages in the form of signals (e.g., Peirce, 1955; Saussure, 1974; Shannon & Weaver, 1949; Vygotsky, 1986, and many others). On this approach, linguistic communication proceeds by a speaker encoding a thought into a sentence of a language – where a language is seen as a code that pairs phonetic and semantic representations of sentences – and by the hearer decoding the uttered sentence into an identical thought.

A well-known problem with the code-model of communication is that linguistic utterances typically contain context-sensitive and/or linguistically ambiguous expressions, as illustrated by (6) and (7):

(6) I lost my bat yesterday.
(7) John is a fine colleague.

An understanding of (6) requires assigning the appropriate referents to the indexical expressions (I, my, yesterday) and disambiguating the homonymous noun bat (‘wooden instrument’ vs. ‘flying rodent’) and the conventionally polysemous verb lost (‘deprived of’ vs. ‘deprived of through death’). In (7), the hearer must, in addition to assigning the appropriate referent to the proper name John, form a hypothesis about the meaning of fine in the NP fine colleague (‘considerate’, ‘diligent’, ‘hard-working’, etc.). In cases such as these it is widely agreed that the hearer must consider information beyond that which is
linguistically encoded in order to derive the speaker-intended meaning, as the
context-sensitive expressions make reference to the discourse situation in which
they were uttered. So, code-models of communication need some way to account
for how hearers assign contextually appropriate meanings to the expressions
used (see Sperber & Wilson, 1986/1995, for a more extensive critique of such
communication models).

Rule-based approaches to polysemy can be seen as modern, sophisticated
versions of the code-model of communication as described above, in which a
considerable amount of the context-sensitivity of lexical meanings is built into
the linguistic system. More specifically, polysemy is seen as being generated by a
set of lexicon-internal generative rules, which operate over information-rich
semantic representations to yield default interpretations (Asher, 2011; Asher &
Lascarides, 2003; Copestake & Briscoe, 1992, 1995; Lascarides & Copestake,
1998; Ostler & Atkins, 1992; Pustejovsky, 1995). To illustrate, consider the
‘logical metonymy’ in (1), repeated below as (8), analysed by rule-based
approaches in terms of a lexicon-internal mechanism forcing a non-conventional
reading of the complement (Pustejovsky, 1995)\(^7\).

(8) Kate began a book.

Constructions of this kind are seen as involving a verb that subcategorises for an
NP or a progressive VP syntactically, but which semantically requires a
complement with an event interpretation. In cases where this requirement is not
satisfied by the surface syntactic structure, a coercion mechanism changes the
denotation of the NP from an entity into an event consistent with eventive
information stored as part of the lexical representation of the noun. In (8), the
appropriate event would be provided by the so-called \textit{telic} role of the noun:

\(^7\) More explicitly, Pustejovsky’s (1995: 111) suggestion is that each expression \(\alpha\) may have a set
of shifting operators available to it, which he calls \(\Sigma\alpha\), which may operate over an expression to
change its type and denotation:

FUNCTION APPLICATION WITH COERCION (FAC): If \(\alpha\) is of type \(c\), and \(\beta\) is of type \(<a, b>\), then,

(i) if type \(c = a\), then \(\beta(\alpha)\) is of type \(b\).
(ii) if there is a \(\sigma \in \Sigma\alpha\) such that \(\sigma(\alpha)\) results in an expression of type \(a\), then \(\beta(\sigma(\alpha))\) is
of type \(b\).
(iii) otherwise a type error is produced.
‘books are for reading’ (or alternatively, by its agentive role: ‘books come about as a result of a process of writing’). In this way, the interpretation ‘Kate began reading a book’ (or ‘Kate began writing a book) is generated by the linguistic system by default.

This type of approach was originally proposed by Pustejovsky (1991, 1995), whose main aim was to provide a more explanatory account of polysemy than a mere listing of senses in the lexicon (cf. Katz, 1972). A more promising approach, he argued, which captures how word senses may partially overlap and be logically related to each other, is a lexicon where items are decomposed into information-rich templates, so-called qualia structures, combined with a set of generative mechanisms for the composition of lexical meanings.

An advantage of this rule-based theory of the processing of logical metonymy is that it accounts for clear interpretive tendencies in uninformative contexts (in the absence of any further contextual clues the preferred or ‘default’ interpretation of (8) would be that ‘Kate began reading a book’ (or writing it), and not that she, e.g., began ripping it up). The availability of such preferred readings is often taken as evidence of a linguistic-semantic process. The claim is that if the lexicon does not propose such a sense it is unclear how it can arise since it is not otherwise indicated by the context (Asher, 2011; Copestake & Briscoe, 1995).

Notwithstanding its intuitive appeal, the problems with this approach are many (Asher, 2011; Blutner, 2002; Bosch, 2007; de Almeida, 2004; de Almeida & Dwivedi, 2008; Falkum, 2007; Fodor & Lepore, 2002). First, there seem to be many cases where a verb makes a demand on a complement that its lexical entry does not satisfy. In the generative lexicon account, such cases are seen as semantically ill-formed: Pustejovsky (1998) gives as an example the VP enjoy the rock, which does not have a default interpretation due to the lack of a telic role defined for the noun rock. Exceptions arise, according to Pustejovsky, when the object is construed relative to a specific activity, as in The climber enjoyed that rock, where rock acquires telicity on the basis of the semantics of the subject NP. However, consider the arguably well-formed utterances in (9) and (10).

(9) Peter enjoyed the nice weather.
Karen enjoyed the children.

Assuming that the intended interpretations here are that ‘Peter enjoyed being outside in the nice weather’ and ‘Karen enjoyed playing with the children’, it is unclear how they could be generated when there seem to be no telic information in the lexical representations of the nouns weather and children that the coercion mechanism could take as input to the compositional process. It is also difficult to see how they could acquire telicity on the basis of the of the subject proper nouns. Thus, it seems that the generative lexicon theory would either make no interpretive predictions for cases such as (9) and (10), or wrongly predict that they are ill-formed. Second, it is difficult to see how the rule-based account can avoid making wrong predictions about many compositional interpretations. For instance, the VPs begin a car and begin a thermometer should be interpreted as ‘begin driving a car’ and ‘begin measuring the temperature’, due to the telic roles encoded by the complement nouns (cars are for driving; thermometers are for measuring temperatures) (Fodor & Lepore, 2002). It is unclear what would prevent such clearly infelicitous interpretations from being constructed as ‘default’. Third, by modelling the processing of logical metonymy entirely in terms of a lexicon-internal process, the rule-based account is unable to account for the interpretive flexibility that is arguably involved in these constructions.

Although preferred readings in uninformative contexts clearly exist (e.g., the tendency to interpret the VP begin a book as ‘begin reading a book’ or ‘begin writing a book’), more specific contextual information can easily point the hearer toward a ‘non-default’ interpretation, which would have to override the ‘default’ reading (e.g., ‘begin dusting a book’, ‘begin mending a book’, ‘begin designing a book’, ‘begin ripping up a book’, etc.). As a consequence, the rule-based account must allow the compositional interpretations generated by the linguistic system to be defeasible, but if this is so, some justification has to be given for why such defeasible semantic rules are necessary in lexical interpretation, when defeasibility is widely agreed to be one of the hallmarks of our pragmatic capacity.

Many rule-based approaches give a similar analysis of the adjectival specification exemplified by (2) above, repeated here as (11).
Jane Austen wrote good books.

On the Pustejovskyan (1995) approach, the meaning of an evaluative adjective like good is generated in linguistic context by a process of ‘selective binding’, which enables an adjective to make available a selective interpretation of an event expression contained in the lexical representation (or the ‘qualia structure’) for the head noun. In (11), the idea would be that good selectively modifies the event description given by its telic role (‘books are for reading’), giving rise to the interpretation ‘good reads’. This analysis runs into the same sort of problems as that for ‘logical metonymy’ above. It has little to say about uses where there is arguably no telic or eventive information for the adjective to selectively modify but where the compositional process seems to proceed as usual (e.g., good children, good weather). It must also appeal to pragmatics for an explanation of non-default interpretations, for instance, the number of other context-dependent ways in which a book could be good (e.g., ‘entertaining’, ‘easy to read’, ‘intellectually challenging’, ‘beautifully designed’, ‘useful to kill flies with’, etc.)

A seemingly stronger case for a lexical rule-based analysis is the type of polysemy that patterns with the syntactic count-mass distinction, often referred to as ‘systematic polysemy’ (cf. Apresjan, 1974). Here the related senses of a word can be predicted from the ontological category of its denotation, and are linguistically marked by the count or mass syntax of the NP in which it occurs. Consider again the examples in (3) above, repeated here as (12).

There was rabbit all over the highway. (‘rabbit stuff’)

Steven had rabbit for dinner. (‘rabbit meat’)

The model wore rabbit on the catwalk. (‘rabbit fur’)

Computational semantic approaches have influentially argued that systematic polysemy is generated by an inventory of lexical inference rules, where the effect of the rules is to change the value of a [+COUNT] or [+MASS] feature in the lexical
representation of the noun, thereby altering its denotation accordingly (e.g.,
Copestake & Briscoe, 1992, 1995; Ostler & Atkins, 1992). One such lexical
inference rule is the UNIVERSAL GRINDER (originally proposed by Pelletier, 1975),
which creates from a count noun denoting a physical object a mass noun with
properties for an unindividuated substance, yielding the ‘rabbit stuff’ sense in
(12)a. above. In addition, the lexicon is thought to contain a set of
conventionalised sub-cases of this rule, including a specialised rule of MEAT-
GRINDING that forms food-denoting mass nouns from animal-denoting count
nouns, and a rule of FUR-GRINDING that forms fur-denoting mass nouns from
animal-denoting count nouns, yielding the ‘meat’ and ‘fur’ senses of rabbit in
(12)b. and (12)c. above. The idea is that this wholly linguistic account avoids a
listing of predictable senses in the lexicon and provides an explanation of how
such sense alternations can be extended productively to any new members of a
category targeted by the rules.

Again, one of the main problems with this rule-based approach is its lack
of interpretive flexibility. Even this kind of ‘systematic’ polysemy seems to
exhibit a considerable degree of context-sensitivity. Consider the following
eamples, where the linguistic alternation between count and mass uses of
nouns yields senses that go beyond those that are normally taken to be
generated by linguistic rules.

(12)  d. Will a hamster bite if it senses rabbit on my hands? (‘rabbit odour’)
e. [Biology teacher]: Rabbit is smaller than hare. (‘rabbit faeces’)
f. [Hunter]: This time of year I prefer using rabbit (‘electronic rabbit
calls’).
g. Last winter, we discovered rabbit, moose and fox in our garden. (‘rabbit
tracks’)

The contextually appropriate interpretation of each of the uses of rabbit in
(12)d.-g. should be easily inferable from the situation of utterance, but their one-
off character makes it seem unlikely that any of them can be generated by a
lexical rule. So it seems that lexical rules, even if they could be shown to be real,
would only be able to account for a subset of the interpretations that the
alternation between count and mass uses of nouns may give rise to. But if a
considerable number of senses are derived pragmatically, we may question the
motivation for the sense extension rules in the interpretive system: what do they
add to a theory of polysemy comprehension by way of explanation?\footnote{Two
reviewers pointed out that an analysis in terms of ellipsis of the NP head might also
be possible for (12d–g) – where, for example, rabbit in (12g) might be an ellipsis for
‘rabbit tracks’, in a case where the prior discourse makes this interpretation
available – with the consequence that such cases of specialised interpretations should
not be seen as equivalent to the conventional interpretations in (12a–c). However, an
ellipsis analysis might in principle also be possible for the conventional polysemy
patterns in (12a–c), where, for instance, rabbit in the utterance Steven had rabbit
for dinner might be seen as an ellipsis for ‘rabbit meat’. While I do not think that
this is the correct way to account for these examples, the fact that an ellipsis
analysis might be possible for all the examples in (12), depending on the information
given by the prior discourse, suggests to me that we have to do with a single
phenomenon here.}

Another problem is that the rules inevitably overgenerate. For instance,
the utterance Sam enjoyed but later regretted the rabbit (Copestake & Briscoe,
1995: 42), whose ‘default’ interpretation seems to be that Sam enjoyed but later
regretted eating the rabbit, would, given the universal grinder and the rules of
animal meat-grinding and animal fur-grinding, be three-ways ambiguous, and it
is not clear how hearers determine when one rule has prevalence over the
others. If we have to appeal to some sort of pragmatic mechanism to do this,
which indeed seems likely, it leaves us again with the question of what role the
lexical rules are playing in the interpretive process.

Turning to the examples of metaphorical and metonymic sense
extensions in (4) and (5) above, repeated here as (13) and (14), it is widely
agreed that consideration of discourse context is required for their
interpretation (e.g., Gibbs, 1994; Glucksberg, 2001; Nunberg, 1979; Sperber &
Wilson, 2008; Wilson & Carston, 2006).

(13) John is a lion.
(14) The ham sandwich is getting impatient.

However, rule-based accounts have suggested that sense extension rules may be
involved in some metonymic and metaphorical processes as well. For instance, it
has been proposed that since the metaphorical extension from animals to
humans with some particular characteristic(s), exemplified in (13), appears to be
productive \textit{(John is a lion/pig/lamb, etc.)}, it can be (partly) expressed in terms of a lexical rule, although the properties ascribed to the human by use of the animal term would not be encoded in its lexical representation (Briscoe & Copestake, 1991; Copestake & Briscoe, 1995).\footnote{This kind of regularity might also be captured in a conceptual metaphor approach (Lakoff & Johnson, 1980), in which (13) could be seen as a linguistic instantiation of the underlying conceptual metaphor HUMAN PERSONALITY TRAITS ARE ANIMALS.} Similarly, metonymic extensions such as that from the ham sandwich to the person who ordered the ham sandwich in (14) are analysed as involving a basic sense extension rule \textsc{physical object} \rightarrow \textsc{human} (Copestake & Briscoe, 1995) or, as in Asher’s (2011) recent proposal, a lexical semantic process of coercion, where a type conflict requires an adjustment of predication in order to satisfy a type presupposition (e.g., the predicate \textit{is getting impatient} presupposes an external argument of the type \textit{agent}).

Given that consideration of pragmatic factors is clearly required in order to derive the contextually appropriate meaning in both these cases, it is unclear what is gained by introducing lexical rules here. For one thing, in view of the considerable context-dependence of the processes of metaphor and metonymy, the rules would only be able to account for a (very) small subset of cases. For another, the pragmatic mechanism(s) that allow(s) us to construct the range of metaphorical and metonymic meanings that are clearly \textit{not} rule-governed, should also enable us to derive the senses in (13) and (14).

In this section, I have discussed rule-based approaches to polysemy. On the basis of a set of standard examples, I have argued that in spite of making accurate predictions in a number of ‘default’ situations, rule-based accounts leave much work for the pragmatic system to do, both in overriding ‘default’ interpretations in contexts where another non-default interpretation was clearly intended, and in constructing unpredictable (non-rule-governed) interpretations, for instance, in cases of metaphor and metonymy. But if pragmatics can do this work, it seems likely that it can also do the part of the interpretative work that rule-based accounts do adequately. In the next section, I consider how the polysemy data can be analysed within a wholly pragmatic-inferential account of utterance interpretation.
2.2 An inference-based approach: polysemy as pragmatics

The challenge for code-based accounts of interpretation, even of the highly sophisticated kind proposed by Pustejovsky and others, is that what is conveyed by linguistic communication – both at the implicit (‘what is implicated’) and the explicit (‘what is asserted’) levels – generally goes well beyond what can be coded, and does so in a highly flexible way. Most pragmatic contextualist accounts of verbal utterance understanding therefore follow Grice ([1967] 1989) in his view that communication is first and foremost an inferential process, and involves a kind of mind-reading: By using a verbal utterance, a speaker provides evidence of her intention to communicate something to the hearer, and the hearer recovers this intention by an inferential process using the evidence provided. A particularly influential pragmatic theory that builds on this insight is relevance theory (Carston, 2002; Sperber & Wilson, 1986/1995; Wilson & Sperber, 2004, 2012). I will now discuss an alternative approach to polysemy based on this framework.

According to relevance theory, human information processing “tends to be geared to the maximisation of relevance” (Sperber & Wilson, 1986/1995: 260), where relevance is seen a potential property of inputs to cognitive processes (e.g., verbal utterances, gestures, facial expressions, etc.), and is assessed in terms of the amount of effort used to process the input and the ‘positive cognitive effects’ the individual may derive from it (where a positive cognitive effect can be described broadly as a ‘worthwhile difference to the individual’s representation of the world’). Other things being equal, the more cognitive effects an input yields to an individual and the less effort it takes to process it, the more relevant it is to the individual at that particular time.

Further, relevance theory takes verbal utterances to constitute a special kind of input, communicating “a presumption of [their] own optimal relevance” (ibid.). By requesting the addressee’s attention, the communicator conveys that her utterance is more relevant than alternative stimuli competing for his attention at the time. An optimally relevant utterance is one that achieves enough cognitive effects to make the utterance worth processing, while avoiding causing the hearer any unnecessary effort in achieving those effects. The hearer’s goal in
communication is to find an interpretation of the speaker’s utterance that meets
the expectations of relevance raised by the utterance itself.

In this framework, the distinction between linguistic semantics and
pragmatics is seen as corresponding to different processes involved in utterance
comprehension: (i) decoding of the linguistic material into a ‘logical form’, and
(ii) pragmatic inference. A logical form is seen as a structured set of concepts: a
‘template’ or ‘schema’ for a range of possible propositions, which contain slots
that have to be filled – a process that requires pragmatic inference (Carston,
2002). In this way, the relevance-theoretic approach to verbal understanding
distinguishes itself sharply from rule-based accounts in that most of the
interpretive work is performed not by lexicon-internal generative mechanisms
but by pragmatic processes operating over underspecified semantic
representations.

On this account, then, the speaker-intended event associated with the VP
in instances of ‘logical metonymy’ would be derived entirely by means of a
pragmatic process. Consider again the example in (1), repeated here as (15).

(15) Kate began a book.

Let us assume that (15) has the following logical form (cf. de Almeida & Dwivedi,
2008):

(16) KATE BEGAN [VP [v^0 e] [NP a BOOK]]

As shown by (16), the syntactic structure of sentences containing a logical
metonymy can be seen as containing an extra VP with an empty verbal head. The
verbal gap that remains in the logical form of such constructions will have to be
saturated using information from the discourse context. Although the process
itself will be linguistically mandated and consist in supplying a missing
constituent to the proposition expressed, the relevant event associated with the
VP will be supplied by a wholly pragmatic process. Imagine the following
context:
Kate, John and Sue work as book conservators at the British Museum. At the moment they are working on restoring a collection of medieval books, all of which are in a poor condition after having been stored on the shelves for many years. Because they are completely covered in dust, each book has to be carefully dusted before being rebound. One day, after hours of hard work, John asks if they should all take a break and go for coffee. Sue has just finished her pile and is ready to follow John to the coffee bar in the Great Court, when Kate utters: ‘Hang on a minute! I’ve just begun a huge old book.’

The most relevant interpretation of the last part of Kate’s utterance is clearly that she has just ‘begun dusting a huge old book’, not reading it, which would be the default interpretation predicted by the Pustejovskyan account. This would be the one that satisfies the hearer’s context-specific expectations of relevance: it is the one that requires the least processing effort and offers satisfactory effects, in the form of an adequate explanation for the content of Kate’s previous utterance of ‘Hang on a minute!’ The prediction is that in (17) the hearer will go straight for this interpretation without the prior computation – and subsequent cancellation – of a ‘default’ interpretation.¹⁰

Now consider again VPs such as begin a car and begin a thermometer, for which the rule-based, Pustejovskyan approach discussed in 2.1 above made wrong interpretive predictions (‘begin driving a car’, ‘begin measuring the temperature’). The pragmatic account proposed here would, of course, come with no such interpretive predictions, but it would equally require that an event be supplied when the VPs are embedded within an utterance in a context. Imagine the context of a garage where Bill is employed as a mechanic. Here it is easy to imagine the most relevant interpretation (i.e. the least effort demanding, yielding the expected sort of cognitive effect(s)) of an utterance of Bill began a car being that ‘Bill began repairing a car’. However, a speaker using the VP begin

¹⁰Pustejovsky (1995) acknowledges that in some cases coercion can lead to different eventive interpretations depending on the aspect of the qualia structure that is modified. For instance, begin a book has two possible default interpretations: ‘begin reading a book’, derived from the telic quale of book, and ‘begin writing a book’, derived from the agentive quale.
a car to describe a situation in which someone began driving a car would (in most cases, at least) not be optimally relevant, as the choice of this expression instead of the more conventional start a car, would, due to the extra effort of processing it would induce, send the hearer off searching for additional effects, which would not be part of the speaker's intended meaning.

On the other hand, rule-based accounts are no doubt correct in their assumption that some interpretations come more readily to mind in uninformative contexts, and could therefore be said to have a ‘default’ character. However, the claim that this is evidence of a linguistic semantic process considerably underestimates the fact that hearers rarely come to the interpretation process ‘empty handed’, as it were; utterances are not understood in a vacuum. If, instead of being objective and linguistically given, context is seen as a psychological construct – a subset of the hearer’s assumptions about the world – which may include assumptions derived from the observation of the physical environment, encyclopaedic knowledge, memories and beliefs as well as the preceding linguistic context (cf. Sperber & Wilson, 1986/1995), there would not be any entirely context-free interpretations. When the assumptions that may be derived from the discourse context are scarce, for instance, if (15) above were to be interpreted in isolation, the hearer will have to rely more on information stored in his long-term memory in interpreting the utterance. Given this, any interpretive preferences observed for logical metonymies in the absence of further context might stem not from lexically stored information but from highly accessible real-world knowledge about the denotations of the lexical concepts in the utterance. For instance, a person reading a book may be regarded as a stereotypical event, which may be stored in encyclopaedic memory as a chunk and accessed as a single unit of information. Retrieving this information from encyclopaedic memory during the interpretation of (15), as a result of the decoding and activation of the lexical concept book, would require little processing effort compared to other possible interpretations (e.g., dusting, designing, mending, ripping up, etc.) which would involve accessing several units of information and thus be more costly in processing terms. In this way, we may account for why certain interpretations are often favoured over others without
being committed to the view that these are always computed as a result of default inferences generated by the lexicon.\textsuperscript{11}

In relevance theory, lexical interpretation is seen as typically involving the construction of \textit{ad hoc} concepts – occasion-specific senses – which may be narrower or broader than the linguistically encoded senses (Carston, 2002; Wilson & Carston, 2006, 2007; Wilson & Sperber, 2012). A mentally-represented concept, a constituent of the 'language of thought' (Fodor, 1975, 2008), is seen as an address or entry in memory that may give access to different kinds of information, including (i) \textit{lexical} information connected with the linguistic form that encodes the concept (i.e. its phonological and syntactic properties), and (ii) a set of assumptions, or \textit{encyclopaedic} information, about the denotation of the concept, that is, conceptually represented assumptions and beliefs, including stereotypes and culture-specific information, and also, in many cases, imagistic and/or sensory-perceptual representations (Sperber & Wilson, 1986/1995: 86). For example, the concept \textit{BOOK} may give access to assumptions such as ‘Books can be read, they are physical objects, are often entertaining, can be intellectually challenging, …’ and so on. Lexical interpretation involves taking the encoded concept and its associated encyclopaedic information, together with a set of contextual assumptions, as input to the inferential process of constructing a hypothesis about the speaker's intended meaning. Consider again the adjectival specification in (2) above, repeated here as (18).

(18) Jane Austen wrote \textit{good} books.

On this account, one pragmatic sub-task for the hearer in interpreting the utterance in (18) would be a specification or narrowing of the concept linguistically encoded by \textit{good} in the NP \textit{good books} (e.g., ‘good reads’) by a process of \textit{ad hoc} concept construction, taking as input encyclopaedic information associated with the other lexical concepts in the utterance. The

\textsuperscript{11}The different predictions about the processing of logical metonymy made by rule-based and pragmatic theories have been subject to some experimental testing (see, for instance, de Almeida, 2004; de Almeida & Dwivedi, 2008; McElree, Frisson, & Pickering, 2006; McElree et al., 2001; Pickering, McElree, & Traxler, 2005; Traxler, Pickering, & McElree, 2002). However, the results are inconclusive and to some extent conflicting, so more research would be needed to settle this debate.
adjective *good* can be seen as encoding a very general concept, which, on most occasions of use, will have to be pragmatically adjusted in order for the hearer to arrive at the speaker-intended interpretation. Often, as in (18), the encyclopaedic information associated with the head noun will play a key role in this process. The pragmatic process of *ad hoc* concept construction is considerably more flexible than the rule-based generative mechanism discussed in 2.1 above, in that not only linguistically-specified information but *any* activated encyclopaedic or situation-specific assumption can be used in deriving the communicated concept, as long as the hearer’s occasion-specific expectations of relevance are satisfied. In this way, the relevance-theoretic account predicts that *good book* may communicate different occasion-specific senses (e.g., a good book could be one that is ‘entertaining’, ‘easy to read’, ‘intellectually challenging’, ‘beautifully designed’, ‘useful to kill flies with’, etc.). It also provides an account of the problem cases for the rule-based approach, namely, examples such as *good children* (e.g., ‘well-behaved’, ‘diligent’, ‘responsible’, ‘caring’, etc.), *good weather* (e.g., ‘sunny’, but could be used to mean ‘overcast’ if in a fishing context, etc.), *good time* (e.g., ‘enjoyable’, ‘fun’, ‘relaxing’, etc.), and so on (with no linguistically-specified purpose encoded by the head noun that could serve as input to lexical rules), all of which would involve the construction of a different *ad hoc* concept on the basis of activated encyclopaedic and/or situation-specific information, constrained by the hearer’s expectations of relevance.

Similarly, the uses of *rabbit* in (3) above, repeated here as (19), can be analysed in terms of pragmatic narrowing where the concept communicated has a more specific denotation than the concept linguistically encoded.

(19)  

a. There was *rabbit* all over the highway. (‘rabbit remains’)  
b. Steven had *rabbit* for dinner. (‘rabbit meat’)  
c. The model wore *rabbit* on the catwalk. (‘rabbit fur’)  

Let us assume that when encountering a mass use of the noun *rabbit*, the output of linguistic decoding – due to the presence of mass syntax – will be a concept
that is constrained to unindividuated entities (RABBIT STUFF). This is in line with the so-called Cognitive Individuation Hypothesis (Wisniewski, Lamb, & Middleton, 2003), according to which the speaker’s use of a count or mass expression leads the hearer to construe the entity referred to as individuated or unindividuated respectively. This provides a highly underspecified input to pragmatic processing. Then, the pragmatic system will construct a narrower ad hoc concept (‘rabbit remains’, ‘rabbit meat’, ‘rabbit fur’) on the basis of the decoded concept, highly activated encyclopaedic information associated with it (e.g., rabbits are animate creatures of flesh and blood, are edible, have fur, etc.), often in combination with other contextual information derived from the utterance situation (e.g. the knowledge that a convoy of trucks has just passed on the highway, that Steven is a gourmet chef, etc.).

There are several advantages to this pragmatic account compared with a standard rule-based account of systematic polysemy. First, it provides the necessary interpretive flexibility for these constructions, allowing for a different ad hoc concept to be constructed in each of (19)a.-c., in the same way as in the more creative examples in (19)d.-g. below.

(19)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>d.</td>
<td>Will a hamster bite if it senses rabbit on my hands? (‘rabbit odour’)</td>
</tr>
<tr>
<td>e.</td>
<td>[Biology teacher]: Rabbit is smaller than hare. (‘rabbit faeces’)</td>
</tr>
<tr>
<td>f.</td>
<td>[Hunter]: This time of year I prefer using rabbit (‘electronic rabbit calls’).</td>
</tr>
</tbody>
</table>

---

12 Here I take an approach to the grammatical count-mass distinction on which there is no lexical or semantic difference between count and mass expressions, and that all differences follow from the syntactic structure in which they occur (cf., e.g., Allan, 1980; Borer, 2005; Bunt, 1985; Pelletier, 2012). On this view, it is NPs, not nouns as such, that are the bearers of (syntactic and semantic) count and mass properties. Nouns themselves (and by extension the concepts they encode) are underspecified with respect to their count and mass properties. Further, assuming that the conceptual distinction between individuals and unindividuated entities is independent of count-mass syntax, our intuitions about the count or mass properties of many concepts (e.g., that the concept DOG is count but WATER is mass) may arise from mentally stored encyclopaedic (or real-world) knowledge about their denotations. This could also explain, at least in part, why some mass syntax embeddings seem less acceptable, or more ‘marked’, than others (e.g., ‘I don’t want book in this room’ vs. ‘I don’t want books in this room’).

13 There is some experimental evidence that collective mass expressions such as furniture are perceived as denoting individuals (e.g., Bale & Barner, 2009; Barner & Snedeker, 2005, but cf.; Wisniewski, Imai, & Casey, 1996). This suggests that the grammatical distinction may not be a direct or perfect reflection of the conceptual distinction: sometimes there may be a conflict between linguistic conventions and real-world properties (e.g., furniture), or the real-world properties of a concept’s denotation may allow for more than one possible construal (cf. French meubles).
Last winter, we discovered rabbit, moose and fox in our garden (‘rabbit tracks’)

Second, while on the rule-based account an utterance such as *Sam enjoyed but later regretted the rabbit* came out as being three-ways ambiguous as a result of the operation of the universal grinder, the rules of animal meat-grinding and animal fur-grinding (Copestake & Briscoe, 1995), such overgeneration does not arise on the pragmatic account, where only interpretations that are consistent with the hearer’s expectations of relevance (i.e. achieves enough implications, at a low enough processing cost) will be computed.

There is no doubt considerable regularity involved in polysemy of the kind exemplified in (19)a.-c., where the related senses of the noun can be predicted from the ontological, or ‘real world’, category of its denotation (e.g., animals), and the polysemy extends productively to any new members of that category. This is also one of the main arguments in favour of a rule-based account, where the lexical rules are seen as capturing language users’ knowledge of these regularities. But if such sense alternations are not generated by an inventory of lexical rules, as assumed by the pragmatic account, how can we explain this regularity?

There is clearly a tight connection between ‘regular’ sense alternations and real-world regularities, to the extent that it seems reasonable to assume that the sense alternations (whether they are regarded as linguistic or not) have their origin in a number of highly regular and predictable states of affairs in the world (Fodor & Lepore, 2002; Rabagliati, Markus, & Pylkkänen, 2011). Our general knowledge of the world tells us, for instance, that there is an inherent relation between an animal and its meat (or fur), and we can easily infer, upon encountering a new kind of animal, that the relation also applies to this instance. It seems likely that the same sort of inference would be made easily accessible to us when we encounter an animal term with mass syntax (e.g., ’John loves rabbit’), or when it occurs without a specification of its count or mass properties (e.g., ‘John regretted the rabbit’), as a result of the activation of encyclopaedic

\[14\] Though which real-world regularities form the basis for conventional sense alternation patterns in a given language or language community may be subject to some arbitrariness.
knowledge associated with the concept in question (e.g., the concept RABBIT), making the meat (or fur) sense easy to access or construct. Further, what may start out as an ad hoc concept in its initial uses (e.g., the narrowing of a mass occurrence of an animal-denoting noun into the meat sense) may become stabilized or conventional over time within a language community as a result of frequent adjustment of the lexical meaning of the word in a specific direction. In such a case, the construction of the ad hoc concept may become progressively more routinized, and a 'pragmatic routine' or inferential shortcut develop (cf. Vega-Moreno, 2007), which is triggered by the activation of the concept in the appropriate context (e.g., MASS OCCURRENCE OF ANIMAL TERM --> ANIMAL MEAT). Such routinized inference patterns might be useful procedures in comprehension, by increasing the accessibility of certain interpretations and thereby contributing to a reduction of hearers' processing effort and thus to the overall relevance of the utterance.\footnote{Though the current account makes no predictions regarding exactly which sense alternation patterns that end up being conventional in a given language or language community.} However, rather than being part of the linguistic system, these inferential short-cuts have a pragmatic basis and can easily be cancelled out by contextual information (linguistic or otherwise) pointing to a different interpretation. Frequent activation of these inferential routines might lead to further conventionalisation of senses, and finally, in some cases, to lexicalisation. An example of this may be the mass occurrence of the noun chicken in English, whose meat sense seems conventional to the extent that it may have acquired a conceptual address of its own. Thus, in this case, it is possible that we have to with two linguistically encoded senses of the noun, where one has developed as a result of frequent pragmatic adjustment of the other in a specific direction. Notice that this is quite different from claiming that certain groups of nouns are associated with lexical rules for sense extension. On this account, where the development of a pragmatic routine might be one step on the way towards a new lexically-stored sense, the conventional nature of many sense alternations is given a wholly pragmatic explanation.

Finally, turning to the metaphorical and metonymical extensions in (4) and (5) above, repeated here as (20) and (21), these are prime examples of pragmatic processes on the relevance-theoretic account:
(20)  John is a lion.

(21)  The ham sandwich is getting impatient.

Relevance theory analyses the metaphorical use of lion in (20) as an instance of *ad hoc* concept construction, but where the outcome would be a *broader* concept than the one linguistically encoded (cf. Sperber & Wilson, 2008; Wilson & Carston, 2006). The decoding of the noun *lion*, resulting in the activation of the concept LION, will cause the hearer to access encyclopaedic information stored about its denotation (e.g., a lion is a large cat, is tawny-coloured, is a skilled hunter, is strong, courageous, takes risks, etc.). Suppose that the most contextually obvious referent for *John* is the speaker’s colleague (JOHNx), who is manifestly not a lion in any literal sense, but known to be a high-risk climber.

The encyclopaedic assumptions associated with the concept LION that are likely to be added to the context in the interpretation of (20), then, would be a subset of those that can be applied equally to humans (e.g., lions are strong, courageous, take risks, etc.) and which contribute to the relevance of the interpretation. So the hearer may broaden the concept encoded by *lion* to an *ad hoc* concept LION* (paraphrasable as ‘strong, courageous, takes risks, etc.’), which would denote actual lions as well as those humans who possess these properties. This interpretation would be a result of the hearer’s mutually adjusting tentative hypotheses about explicit content (JOHNx IS A LION*), implicated premises (A LION* IS STRONG, COURAGEOUS, TAKES RISKS, etc.), and implicated conclusions (JOHNx IS STRONG, COURAGEOUS, TAKES RISKS, etc.), which are incrementally modified against the background of the hearer’s context-specific expectations of relevance. (For discussion of the mutual adjustment process in lexical interpretation, see, e.g., Carston, 2002; Sperber & Wilson, 1998; Wilson & Carston, 2006, 2007).

The metonymic use of the ham sandwich to refer to ‘the person who ordered a ham sandwich’ in (21) can, from a relevance-theoretic perspective, be seen as an instance of reference substitution based on a highly accessible contextual assumption activated by the utterance situation, constrained by the hearer’s occasion-specific expectations of relevance. Imagine (21) being uttered at a café by Jane the waitress to Sam the waiter during lunchtime, a very busy
time of the day. The waiters are running around trying to serve customers their correct orders in time. Against this background, an anticipated conclusion of Jane's utterance would be that whoever among the customers is getting impatient should be served his or her food as quickly as possible. The linguistically specified concept HAM SANDWICH would provide additional activation to an already highly accessible contextual assumption about ham sandwiches being possible orders at this café, and by a process of spreading activation, about customers having ordered ham sandwiches. Let us say that at the time of utterance of (21) there is only one customer waiting for his order of a ham sandwich. The encoded meaning of ham sandwich would then activate the contextual assumption 'customer a has ordered a ham sandwich'. The interpretation of the ham sandwich as communicating 'the person who ordered a ham sandwich' allows Sam to identify customer a as the referent of the expression, and warrants the implicated conclusion (implicature) that customer a should be served his food as quickly as possible. The overall inferential process leading to the derivation of this implicature severely constrains the range of possible associative relations that the encoded concept HAM SANDWICH may enter into, and which may form the basis for the metonymic reference substitution in (21).

The regularity associated with many metonymic uses (cf. Apresjan, 1974), for instance, the development of a convention of referring to customers via their food orders among the employees of a café, provides an important motivation for many rule-based analyses of the phenomenon. From a relevance-theoretic pragmatic point of view, this can be seen as cases where a repeated use of a linguistic metonymy that links different concepts together has set up a pattern of conceptual activation, or a 'pragmatic routine', which gives rise to a sense of regularity (other examples may be PRODUCT FOR PRODUCER, BUILDING FOR INSTITUTION, DIAGNOSIS FOR PATIENT, etc.).

---

16 Such pragmatic routines have similar characteristics as the structures that cognitive linguists call 'conceptual metonymies' (first discussed by Lakoff & Johnson, 1980). An important difference between the two approaches, however, is that the pragmatic account takes the systematic conceptual correspondences to arise for communicative purposes, rather than as surface reflections of underlying conceptual metonymies. See Wilson (2009) for a discussion of this issue in connection with metaphor.
In this section, I have discussed an inferential, relevance-theoretic approach to polysemy. On the basis of the same set of examples used to evaluate rule-based accounts in section 2.1, I have argued that the pragmatic account seems capable not only of handling the most context-dependent cases of polysemy where rule-based accounts must appeal to pragmatics, but also that part of the interpretive work that they do adequately. It also avoids many of the problems associated with rule-based accounts, in particular with respect to overgeneration and interpretive inflexibility. Given this, I think it remains for proponents of rule-based accounts of polysemy to explain what makes the rules necessary, and what is to be gained by deriving some senses in one way (via lexical rules) and others in a distinct way (via pragmatics). At least, it seems that considerations of theoretical economy would favour a unitary pragmatic approach.

3. Why polysemy?

So far I have discussed two different approaches to the question of what the nature of the mechanisms or processes involved in the derivation of polysemy may be, and argued, on the basis of a set of standard polysemy examples, that the pragmatic-inferential account seems to provide the most promising alternative for a unified treatment of the data. But why do we find such proliferation of polysemy in our languages in the first place? What is it about our language systems, specifically their lexical component, that makes them so susceptible to polysemy? In this section, I address this issue, and show how the rule-based and pragmatic-inferential accounts may come up with different answers to the question of what the underlying motivation for polysemy in natural language may be.

As we have seen in the two previous sections, rule-based and pragmatic-inferential accounts propose different solutions to the problem of linguistic underdeterminacy: how addressees bridge the gap between (surface) linguistic meanings (i.e., underspecified meanings) and speaker meanings (i.e., contextually enriched meanings). While rule-based accounts build a lot of context-sensitivity into the lexicon, postulating a set of generative lexical mechanisms that operates over information-rich lexical entries to yield default
interpretations, the pragmatic-inferential approach takes linguistic utterances to provide no more than clues to speaker-intended meanings, which have to be inferred on the basis of contextual evidence. It is possible to see these different solutions as reflecting two distinct views on the nature of linguistic underdeterminacy discussed by Carston (2002): one in which it is a form of ‘convenient abbreviation’ and another that takes it to be essential. I will discuss these two positions and their implications for the question of polysemy motivation in what follows.

On the first, ‘convenient abbreviation’ view, linguistic underdeterminacy is a matter of effort-saving convenience for the hearer, possibly resulting from a convention of linguistic usage or a natural drive towards communicative efficiency (Carston, 2002: 29). Although sentence meaning more often than not underdetermines the proposition expressed by it, a sentence that fully encodes the speaker's meaning could in principle be supplied. Consider the utterance in (6) again, repeated here as (22)a., and a suggestion for a fully encoding counterpart in (22)b.

(22) a. She lost her bat yesterday.

b. Susan Thompson caused Susan Thompson to be deprived of the wooden implement with a handle and solid surface used for hitting the ball in baseball that belonged to Susan Thompson, between 2pm and 4pm on 14 October 2013 somewhere in the Islington area, London, UK.

In order to save herself the effort of having to express a long, complex sentence such as that in (22)b., the speaker can choose to use a sentence that does not fully encode her intended meaning, and rely on the hearer using his pragmatic capacity to turn it into a fully propositional representation. On this approach, our pragmatic ability would be a useful add-on to our language capacity but would not be strictly essential in enabling us to express ourselves and communicate the

---

17 This lies close to the view held by Quine (1960) and Katz (1972), who argued that every proposition expressed by a natural language sentence was describable in terms of a context-independent ‘eternal’ sentence. Katz termed this the principle of effability.
way we do. Some version of this view of linguistic underdeterminacy is likely to
underlie many rule-based approaches to natural language processing.

The second, essentialist view takes linguistic underdeterminacy to be an
essential feature of the relation between linguistic meanings and speaker
meanings (ibid.). Given the complexity and fine-grainedness of the thoughts that
speakers can entertain and communicate to each other (e.g., their containing
private references to time, space, people events and so on), they generally do not
lend themselves to a full encoding by natural language sentences. So although
the sentence in (22)b. comes closer to encoding the speaker-intended meaning
than the one in (22)a., any attempt to be fully explicit is bound to fail (for
extensive discussion of the rationale behind this position, and a defense of the
‘ineffability of thought’, see Carston, 2002: Chapter 1). On this approach, our
ability to make pragmatic inferences about speaker-intended meanings would
provide the essential foundation for our expressive and communicative abilities.
This is the view of linguistic underdeterminacy that underlies the pragmatic,
relevance-theoretic account of verbal understanding discussed in Section 2.2
above.

Both these positions provide a basis for polysemy. On the convenient
abbreviation view, polysemy could be motivated by a goal of economy of
expression, representing an effort-saving strategy for the speaker and
contributing to communicative efficiency. Instead of going through the laborious
task of fully encoding the lexical sense she has in mind – which she could do if
she wanted to – the speaker will often choose a more economical form of
expression, trusting the hearer to pragmatically infer her intended lexical
meaning. Although rule-based approaches do not explicitly adhere to this
position, it provides them with a plausible explanation for why polysemy is such
a pervasive phenomenon in natural language, given their postulation of
information-rich codes and the comparatively restricted role allocated to
pragmatic reasoning. However, with respect to polysemy, rule-based approaches
make a stronger claim about effability in that they do see the language as being
fully encoding in the canonical case, with all the information – as well the
procedures for manipulating it – necessary for the hearer to arrive at the
speaker-intended meaning being built into the lexicon, though reserving a role
for pragmatic reasoning in cases where an interpretation other than the default one was intended. In this way, lexical sense extension rules could be seen as contributing to communicative efficiency by minimising (surface) linguistic complexity, requiring the hearer to use his linguistic capacity to generate default compositional interpretations. In those cases where a non-default interpretation is intended, it will be derived pragmatically on the basis of discourse context, through a process of cancellation and substitution, but the main interpretive work is done linguistically on the basis of lexical rules.

On the essentialist view of underdeterminacy, however, where the linguistic codes that speakers make use of are not capable of fully encoding their thoughts (i.e. speaker meanings) and must be supplemented by pragmatic inference, polysemy would follow as a natural consequence. If the vocabularies of our languages are not capable of encoding the range of concepts we can entertain and communicate, polysemy – understood as the ability of words to express different meanings in different contexts – would be a necessity. If this view of linguistic underdeterminacy is correct, providing a full account of polysemy in terms of the workings of the linguistic system should, in principle, not be feasible. At least, polysemy would not have to be entirely linguistically generated if communicators possess a powerful enough pragmatic-inferential capacity. However, although the essential nature of underdeterminacy would be the ultimate motivation for polysemy on this position, it is still compatible with the idea that polysemy often represents an effort-saving strategy for communicators. The proliferation of polysemy in natural languages suggests that language users may find it easier to take an already existing word and extend it to a new sense than to invent an entirely new word. One reason may be that the stabilisation of a new word in a language is a relatively slow process that has to be coordinated over a large group of individuals over time. But the typically pairwise coordination involved in any given communicative act is a less elaborate affair (for discussion, see Sperber & Wilson, 1998). Given our

---

18 It should be noted that on an account such as Pustejovsky’s (1995), polysemy is an essential aspect of language, which arises as a result of generative mechanisms operating over underspecified, albeit informationally-rich, lexical representations (in the form of qualia structures). However, this is quite a different position from that of the inferential account, in which polysemy is taken to be an essential aspect of communication.
pragmatic ability to form hypotheses about speaker meanings on the basis of linguistic utterances and contextual information, there would, in most cases, be no need for a new word to describe something that may just as well be described by using an already existing word with an extended meaning.

Although we find a basis for the existence of polysemy in both accounts discussed here, it seems to me that it has a stronger motivation on the pragmatic-inferential account, where it arises as a natural consequence of lexical concepts being unable to fully encode speaker-intended concepts rather than optionally as part of communicators’ striving toward brevity. In the final section of this paper, I will consider of some further implications of a fully pragmatic approach to polysemy.

4. A fully pragmatic approach: perspectives and implications

In modern pragmatic theory, and relevance theory in particular, the capacity to infer speaker meanings on the basis of the evidence provided is taken to be reliant on the more general theory of mind capacity, that is, the ability to infer and attribute contentful mental states to others (e.g., Baron-Cohen, 1995; Premack & Woodruff, 1978; Wimmer & Perner, 1983). The claim is that the theory of mind capacity provides the foundation for the kind of ostensive-inferential communication that humans engage in (Sperber & Wilson, 2002); it is what enables communicators to bridge the gap between characteristically underspecified linguistic meanings and intended speaker meanings, including inferring speaker-intended concepts from lexically-encoded concepts. As I have argued in the previous section, this view provides a strong basis for polysemy, with our pragmatic inferential ability playing a fundamental role in its development and proliferation in verbal communication. Below I consider some of its implications for acquisition, diachrony and non-verbal forms of communication.

4.1 Polysemy in acquisition

It is widely agreed that young children’s word learning requires an early capacity for intention reading (e.g., Akhtar, Carpenter, & Tomasello, 1996; Baron-Cohen, Baldwin, & Crowson, 1997; Bloom, 2000; Clark, 1997). Children also rely early
on their pragmatic ability in interpreting polysemy-related phenomena such as
class extensions, where a word is used in a novel lexical category (e.g., *Can you
yellow the circle?*) (Bushnell & Maratsos, 1984; Clark, 1982), metaphor (e.g.,
Deamer, 2013; Özçaliskan, 2005) and metonymy (Author et al., revised and
resubmitted)\(^{19}\). One study that specifically investigated children’s ability to cope
with systematic polysemy found an early emerging facility for using pragmatics
in sense resolution (Rabagliati, Marcus, & Pylkkänen, 2010). In this study, young
children showed an adult-like ability to make ‘licensed’ sense alternations, for
instance, to correctly interpret the conventional alternation between the movie
sense and the physical object sense of the word *DVD*, but were also more willing
than adults to accept ‘unlicensed’ senses, for instance, the physical object sense
of *movie* in *The movie is shiny.*\(^{20}\) This suggests that in the early stages of language
learning, children may actually be more ‘pragmatic’ than adults in the sense that
they show an even greater degree of flexibility in interpretation, accepting
senses for words that go beyond those that adults find appropriate (or relevant).
(A weakness of Rabagliati et al.’s study is that the sentences containing the
systematic polysemy were presented to the participants in isolation, without a
supporting context. It is possible that if given contexts that clearly biased the
‘unlicensed’ senses, adults too might have been more likely to accept these
senses).

Rabagliati et al. (2010) suggest that a rule-based account could explain
these results in terms of children initially possessing a broad range lexical rules,
generating both licensed and unlicensed senses (cf. Copestake & Briscoe, 1995),
some of which are ‘unlearned’ over time as a result of increasing exposure to

---

\(^{19}\) It should be mentioned here, though, that several studies have shown that children have
difficulties deriving so-called ‘scalar implicatures’ (e.g., Noveck, 2001), and typically do not pass
standard false-belief tasks until around the age of four (e.g., Doherty, 2000); this has led many
scholars to take both pragmatics and theory of mind to be relatively late-emerging abilities.
Recent studies, however, have suggested that children’s apparent pragmatic difficulty with scalar
implicature may have more to do with the complexity of the tasks used or with the acquisition of
various other cognitive skills than with the development of the pragmatic capacity *per se* (Barner,
Brooks, & Bale, 2011; Papafragou & Tantalou, 2004; Pouscoulous, Noveck, Politzer, & Bastide,
2007). Also, recent work with non-verbal versions of the false-belief tasks (e.g., Baillargeon, Scott,
& He, 2010) suggests that infants are already able to attribute false beliefs, and may have an early
theory of mind ability.

\(^{20}\) However, whether this is actually an instance of an ‘unlicensed’ sense seems questionable. For
instance, as one reviewer pointed out, sentences such as *If a redbox movie is scratched, can you
replace it for free?,* evoking a physical object sense of *movie*, are readily retrieved from the
internet.
their language, a process which would involve ruling out infrequent senses from their lexical repertoire. For instance, if children never hear the word *movie* used in a physical object sense, they might discount the probability of there being a rule that creates physical objects from abstract entities in their language. I find this hypothesis very implausible, given the range of other polysemy-related phenomena that children are able to cope with from a very early age, which are unlikely to require the presence of any lexical rules for interpretation. This early pragmatic competence should enable children to cope with cases of systematic polysemy too, although it may sometimes overgenerate. Furthermore, one would have to explain how such initially too broad rules develop quite independently of the child’s conceptual knowledge, given the tight connection that clearly exists between systematic sense alternations and real-world regularities in the language of adults. Another possible rule-based explanation could be that children start out by being radically pragmatic, but acquire the lexical rules at a later stage in development as a result of exposure both to their language and to the real-world relations on the basis of which the lexical rules arise. While this hypothesis seems more plausible, one would have to offer a reason for why the ‘rules’ must be part of the linguistic system and not rather an artefact of conceptual organisation, reflecting highly predictable and regular states of affairs in the world, where children’s developing ability to handle systematic polysemy could emerge as a by-product of acquiring a more adult-like conceptual organisation and pragmatic competence.

### 4.2 Polysemy in diachrony

If the fully pragmatic account is on the right track, all synchronic instances of polysemy should, in principle, be traceable back to the operation of a pragmatic process. A prevalent hypothesis about semantic change is that its main driving force is pragmatic, being motivated by speaker-hearer interactions and communicative strategies (e.g., Traugott & Dasher, 2002). It is also widely thought that semantic change must go through a stage of polysemy, in which related meanings of a word that emerged at historically different periods coexist over time in a language, both in individual speakers and in language communities (Hopper, 1991). While rule-based accounts generally have little to
say about the historical development of polysemy, these hypotheses about semantic change are quite compatible with the pragmatic-inferential account of polysemy proposed here. For instance, what may start out as an *ad hoc* concept in one stage may become stabilised or conventional over time for individual speakers or within a language community, as a result of frequent adjustment of a lexical meaning in a specific direction (e.g., a narrowing of the mass sense of *chicken* into a meat sense, cf. Section 2.2). In such a case the construction of the *ad hoc* concept may become progressively more routinized until an inferential shortcut (or ‘pragmatic routine’) develops (cf. Vega-Moreno, 2007). A possible development from there might be a lexicalisation of the new sense, and potentially, that the original *ad hoc* concept takes over from the originally encoded concept.  

At the synchronic level, then, individual speakers may differ with regard to which senses of a word they have stored in their mental lexicons. For instance, for some speakers of English, the broadened, metaphorical sense of *lion* (discussed in Section 2.2) may be conventional and thus lexically stored. Recognising this concept as the one intended by a speaker on a given occasion of use would then be a matter of disambiguation rather than *ad hoc* concept construction. For other speakers, even within the same speech community, *lion* may have only one encoded sense (the animal sense), and the broadened, metaphorical sense would be derived pragmatically through *ad hoc* concept construction. So the construction of a particular *ad hoc* concept may be an occasional, or even first time, affair for one communicator and a routine pattern for another. In this way, pragmatic inference serves an important function in compensating for such differences among members of a language community.

---

21 One reviewer objected that if all polysemy is a result of a process of meaning extension, this would predict that there should always be a gap in history between when different senses of a word enter a language, but that this is likely not to be the case. For instance, in the case of the noun *door*, which has both a physical object sense (‘He painted the door’) and an aperture sense (‘He walked through the door’), the different senses may have arisen at precisely the same times. While it is an empirical question exactly when specific senses arise in a language, I agree that cases such as *door* (and similar cases such as a *window*, *entrance*, *jalousie*, *portal*, etc.) are special in the sense that it is difficult to see how one sense could have been derived from the other. However, I still think that these (apparent) sense alternations can be given a pragmatic explanation, in terms of increased activation of certain aspects of encyclopaedic knowledge associated with the denotations (cf. Langacker’s 1984 notion of ‘active zones’). If something like this is true, there may be more than one route by which senses of a word may arise.
enabling them to end up with the same lexical senses but in many cases via distinct routes (see Wilson & Carston, 2007, for further discussion).

4.3 Non-verbal polysemy?

Another implication of treating polysemy as a fundamentally communicative phenomenon in this way is that, in principle, we should expect not just words but any simple ostensive stimulus, for instance, manual and facial gestures, to be susceptible to polysemy. One example might be the use that car drivers make of a single flash of their headlights to another driver, which could (in England), mean at least the following: (a) ‘go ahead in front of me’; (b) ‘thanks for giving way to me’; (c) ‘watch out – there is traffic police up ahead’. The general meaning of this signal might be something like ‘friendliness indication’, with the more specific meanings derivable on the basis of context (e.g., positioning of the cars on the road, whether or not one of the drivers has already flashed his/her lights, etc.).

Another example might be smiles, which, depending on the context, can communicate a range of related feelings: amusement, affection, sympathy, and so on (cf. Wharton, 2009). An example from preverbal communication might be a toddler extending his or her hands upwards, which, depending on the toddler’s position, could convey several different meanings: [on floor] ‘I want you to pick me up’; [in high chair] ‘Help me down’; [under object of interest] ‘Get me to that object up there’, and so on. The gesture itself could be seen as having a general meaning along the lines of ‘take me from one place to another’, with the more specific actions intended by the toddler being contextually-determined.

5. Final remarks

The central topic of this paper has been whether the aspects of meaning that are involved in the construction of polysemy have a primarily linguistic or non-linguistic basis, and the extent to which its proliferation and development in natural language can be explained given each of these views. More specifically, the question has been whether polysemy results primarily from the operation of lexicon-internal processes or from pragmatic-inferential processes applying at the level of individual words. Some people may argue that the difference

---

22 Thanks to Robyn Carston for this example.
between these two accounts is one of degree only, and simply concerns how
large a role pragmatics should play in the theory: While rule-based approaches
maintain that a considerable amount of linguistic knowledge is involved in the
generation of polysemy, with pragmatics taking over when an interpretation
other than the linguistic default meaning is intended, the wholly pragmatic
approach downplays the linguistic aspect and suggests that polysemy arises
mainly as a result of the operation of pragmatic processes over underspecified
lexical meanings, taking contextual information and encyclopaedic assumptions
about conceptual denotations as input to the inferential process. However, I
think that the difference between these two approaches is much more
fundamental than this, and involves two radically different conceptions of what a
language is, and the role it plays in the communication process. As we have seen,
rule-based approaches treat the language as providing a rich code that enables
communicators to encode and decode their thoughts in much detail, and
pragmatics merely as a useful add-on to this capacity. By contrast, the pragmatic
account takes the role of the linguistic system to be that of providing a minimal
input or clue, which the inferential system – seen as the essential foundation for
our expressive and communicative abilities – uses as evidence to yield
hypotheses about occasion-specific, speaker-intended senses. I have argued
throughout this paper that the assumption that a large part of the interpretive
work involved in the processing of polysemy should be attributed to the
linguistic system itself requires independent justification, given that we have an
independently motivated pragmatic interpretation system capable of rapidly
generating new senses in contexts. The pragmatic-inferential approach predicts
that lexical items are used to express a variety of occasion-specific senses, which
include but go far beyond the default senses predicted by rule-based approaches.

Finally, I have suggested that the proliferation of polysemy appears to
have a stronger motivation on the pragmatic-inferential account, where it arises
as a natural consequence of lexical meanings not being able to fully encode
speaker-intended senses. Viewing polysemy as a fundamentally communicative
phenomenon in this way allows us to provide a unified account of its role in
several domains, including acquisition, diachrony and non-verbal forms of
communication.
References


Author et al. (revised and resubmitted). 'The moustache sits down first': On the acquisition of metonymy.


Pelletier, F. J. (2012). Lexical nouns are both +MASS and +COUNT, but they are neither +MASS nor +COUNT. In D. Massam (Ed.), *Count and Mass Across Languages* (pp. 9-26). Oxford: Oxford University Press.


