Ambiguity and Coherence

Alex Lascarides,
Centre for Cognitive Science,
University of Edinburgh,
2, Buccleuch Place,
Edinburgh, EH8 9LW,
Scotland, UK
alex@cogsci.ed.ac.uk

Ann Copestake
CSLI
Stanford University,
Cordura Hall,
Stanford, CA 94305,
USA
aac@csli.stanford.edu

Ted Briscoe
IRCS, University of Pennsylvania,
Suite 400C, 3401 Walnut Street,
Philadelphia,
PA 19104–6228,
USA
ejb@linc.cis.upenn.edu

August 15, 1995

Abstract

Several recent theories of linguistic representation treat the lexicon as a highly structured object, incorporating fairly detailed semantic information, utilising various types of relationships between entries, especially inheritance, and allowing multiple aspects of meaning to be represented in a single entry (e.g. Pustejovsky, 1991; Copestake, 1992; Copestake and Briscoe, 1995). One consequence of these approaches is that word senses cannot be thought of as discrete units which are in one-to-one correspondence with lexical entries. This has many advantages in allowing an account of systematic polysemy and of graded distinctions between senses (Cruse, 1986, 1995), but leaves the problem of accounting for effects such as zeugma and the absence of crossed readings, which have traditionally been explained in terms of multiple lexical entries, but which can also arise in examples where other criteria demand that a single lexical entry be involved. Copestake and Briscoe (1995) claimed that these cases could be explained in terms of discourse coherence, but did not describe how this might work. We remedy this here, by developing a general account of zeugma and the non-availability of crossed readings.

1 Introduction

This paper is part of a general attempt to explicitly and formally investigate interactions between the lexicon and pragmatics. We assume that lexical and linguistic processing is informationally encapsulated and utilises relatively simple 'taxonomic' lexical semantic knowledge, in contrast with pragmatic reasoning which is open-ended and involves arbitrary knowledge and interactions. Previous papers have discussed how pragmatics affects lexical processes such as metonymy (Lascarides, 1995; Lascarides and Copestake, 1995), disambiguation (Asher and Lascarides 1995a, Asher and Sablayrolles 1995), metaphor (Asher and Lascarides, 1995b) and blocking (Briscoe *et al*, 1995). Here we extend that work by formalising a general principle of discourse interpretation which captures the intuition that hearers prefer not to infer things that have an adverse effect on the quality of coherence in the discourse. In particular, this principle formulates how discourse coherence can constrain how words in the discourse are interpreted. One effect of this is to give an account of some rhetorical effects, which have frequently been used as tests for lexical ambiguity (e.g. Zwicky and Sadock, 1975).

The traditional picture of lexical organisation is that distinct word senses should be represented as discrete units (e.g. Kempson, 1977:81f). This is taken to account for the markedness of (1) which demonstrates the figure of speech traditionally known as zeugma, and also for the impossibility of crossed readings: (2) and (3), for example, cannot have readings where John has a tool designed for shaping and Bill has a folder of documents.[1]

(1) He was wearing a pair of women's stockings and a look of considerable embarrassment. (adapted from Cruse, 1986)

(2) John and Bill each have a file.

(3) John has a file and Bill does too.

Copestake and Briscoe (1995) argue that an account of polysemy based on distinct lexical entries alone is inadequate because it does not adequately capture the systematic nature of much polysemy and does not allow for gradations in distinctions (see e.g. Cruse, 1986, 1995). Furthermore, in many cases it seems that multiple aspects of a word's meaning should be encoded in a single lexical entry — for example, *book* can be treated as having a dual nature as both a physical object and an abstract representation, which are both encountered in examples such as (4a,b).

(4) The books on the top shelf are about syntax.

(5) That thesis has thousands of pages and is quite unreadable.

Such examples (which are are classified as cases of *constructional polysemy* by Copestake and Briscoe) cannot be easily accounted for if we assume multiple lexical entries for *book* and *thesis*. Some recent theories of the lexicon attempt to account for constructional polysemy and other cases of systematic polysemy by treating the lexicon as a highly structured object, incorporating fairly detailed semantic information, utilising various types of relationships between entries, especially inheritance, and allowing multiple aspects of meaning to be represented in a single entry (e.g. Pustejovsky, 1991, 1995; Copestake, 1992; Copestake and Briscoe, 1995). However, if the lexical entries are not distinct, we are left with no formal explanation of the oddness of sentences such as (6) and the absence of crossed readings in (7), which cannot refer to a situation where Kim sold two works to a publisher and Sandy sold two spare paperbacks to a book dealer.

(6) Kim's thesis is orange and unreadable.

(7) Kim and Sandy both sold two books.

Copestake and Briscoe claimed that such cases could be explained in terms of discourse coherence, but did not describe how this might work. In this paper, we will provide an account of zeugma and the absence of crossed readings which will cover examples such as those shown. We will claim that although the existence of a multiple lexical entry (MLE) is part of the explanation for some cases of zeugma and non-crossed readings, inferences about pragmatic interpretation and discourse coherence can cause this effect, even when there is a *single* lexical entry. Our account will thus complement and expand the traditional one.

We will first discuss zeugma and crossed readings and their use as a test for ambiguity, briefly discuss the traditional MLE approach and illustrate its shortcomings as an account of these phenomena. We will then advocate a complementary approach, which involves recourse to pragmatic reasoning, constraints on discourse coherence and psychological models of incremental interpretation. This revised account will extend the MLE approach by explaining how zeugma and non-crossed readings can occur in the absence of distinct lexical entries. We describe this informally in §3 by means of some examples. By studying how information in the discourse leads to zeugma and non-crossed readings, we will come to some general conclusions about the way in which lexical organisation should be linked to other components of the linguistic interpretation model. The theories of discourse structure and pragmatics are introduced in the next section, followed by a formal account of these examples in §5.

[1]We use *sense* pretheoretically, to correspond to something like the distinctions which might be made by lexicographers. We use zeugma to refer to a construction in which a single word is applied to two (or more) others in such a way that its sense(s) are appropriate to each in different ways.

2 Zeugma and crossed readings

For the purposes of this paper, we will assume that zeugma is a word pun or a figure of speech, in which two terms are inappropriately linked together. Sentences (8) to (12) are all examples of zeugma, and they indicate that it can occur with a wide variety of syntactic constructions.

(8) Some dam busters blew up banks and so did some bank robbers.

(9) I tried to take the plane to Chicago, but it was too heavy.
Groucho Marx

(10) Heseltine left with a smile, a wave and his wife.
BBC news

(11) Mr. Pickwick took his hat and his leave.
Dickens, Pickwick Papers

(12) That whole committees of people, whole corridors and floors of professionals thought it worth making defies the licence fee and belief.
A. A. Gill, The Sunday Times, January 29th, 1995

In all of these sentences, a single occurrence of a word is used in incompatible ways. In the attributed examples, this is being done deliberately as a pun. In the MLE account, the zeugma in all of these cases would have arisen from the use of an ambiguous lexical item: *bank* in (8), *take* in (9) and (11), *with* in (10) and *defy* in (12). In (11) and (12) one of the usages is part of an idiom (*take one's leave*, *defy belief*).

It is important to be careful about which contexts are used to test for zeugma. It is clearly inappropriate to use contexts where contradictory statements are made of the same entity. Thus sentence (13) is inconsistent, but not zeugmatic:

(13) The teacher was pregnant and had a beard.

Rather, zeugma must arise from linking two mutually satisfiable assertions. For example, it's possible to bank money and then bank a plane, but these assertions are inappropriately linked in (14), resulting in zeugma.

(14) John banked the money and then the plane.

Zeugma and the absence of crossed readings are often taken to be the most reliable tests for lexical ambiguity as opposed to vagueness or generality of meaning (e.g., Cruse 1986, Zwicky and Sadock 1975). So, for example, the fact that we don't interpret (2) and (3) as John having a dossier and Bill a tool (even though they might possess different physical objects) indicates that *file* is ambiguous between the sense meaning *dossier* and that meaning *tool*.

(2) John and Bill each have a file.

(3) John has a file and Bill does too.

So on the assumption that lexical ambiguity implies MLEs, (2) and (3) indicate MLEs for *file*. Similarly, (15) is zeugmatic and (16) doesn't permit a non-crossed reading. So the word *bank* has MLEs; one for its mound of earth sense, and one for its financial institution sense.

(15) Some dam busters and bank robbers blew up banks.

(16) Texas and Alabama have preservation orders on their most beautiful banks.

On the other hand, *teacher* is neutral or general between male and female *teachers* in (17b) can refer to a mixture of male and female teachers.

(17) a. Teachers may take maternity or paternity leave.
b. John and Bill both like their teachers.

And because *brush* in (18) can refer to a mixture of paint brushes, cleaning brushes, tooth brushes and so on, there must be a lexical entry for *brush*, which is neutral with respect to the type of brush being referred to.

(18) Nylon bristles and plastic handles make cheap durable brushes.

Zeugma is often taken to be symptomatic of lexical ambiguity as opposed to lexical vagueness, but the concept of vagueness is heterogeneous. In a multiple inheritance hierarchy, multiple aspects of meaning can be combined in a single entry and since lexical entries themselves are not simply listed but are hierarchically arranged, a general entry can be given for a word usage which subsumes more specific usages. The neutral or general senses discussed above are treated in this way. Their underspecified senses can be specialised by linguistic context, and therefore they are termed constructionally polysemous in Copestake and Briscoe (1995). Below we argue that this (sub)type of lexical vagueness can lead to zeugma. This contrasts with the MLE view of zeugma, as being caused by distinct senses of a word that lack a more general subsuming sense (e.g., *bank*, *file*). Below we give some detailed arguments for this, but first we review the MLE explanation.

2.1 The MLE account of zeugma and the absence of crossed readings

The MLE explanation of zeugma and absence of crossed readings goes roughly like this. First, there are at least two entries for *bank* in the lexicon: call these *bank*$_{\text{ground}}$ (the mound sense) and *bank*$_{\text{org}}$ (the financial organisation sense). Moreover, there is no entry for *bank* which is general between these two senses. So there are two possible logical forms, for both (15) and (16), which in each case are identical except for the sense of *bank* involved.

We give these below in Discourse Representation Theory (DRT) (Kamp and Reyle, 1993), since our account of zeugma will be formally explicated using DRT. Throughout this paper we'll assume that logical forms are built via a constraint-based grammar such as Head Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1994), with a semantic component encoded in DRT (cf. Frank and Reyle, 1995), and that the process by which the DRS is built is monotonic.[2]

(15′)

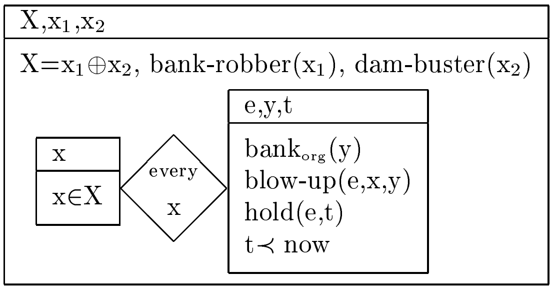
X, x_1, x_2

$X = x_1 \oplus x_2$, bank-robber(x_1), dam-buster(x_2)

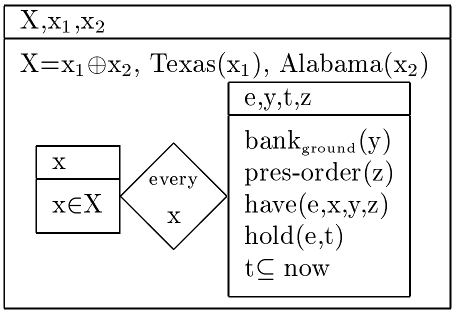
[x | $x \in X$] ⟨every x⟩ [e, y, t | bank$_{\text{ground}}$(y), blow-up(e, x, y), hold(e, t), $t \prec$ now]

[2] It should be stressed at this stage that the way we represent multi-sentence discourse will go beyond DRT, in that we will use rhetorical relations such as *Parallel* and *Contrast* to connect DRSs together. We do this so that coherence constraints on these relations can predict zeugma. Furthermore, we assume that the rhetorical relations that connect DRSs are computed via the semantic content of those DRSs. This means that the semantic representation of a clause must be evaluated against a model before we reason about how it connects to the semantic representation of the text so far. This is in further contrast to traditional DRT, where evaluation against a model doesn't occur until the semantic representation of the whole discourse is built.

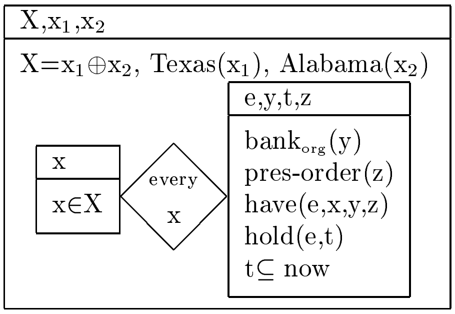
(15″)



(16′)



(16″)



We now immediately see why *banks* in (16) doesn't refer to a mixture of earth mounds and financial institutions: neither logical form would be satisfied in this case. There is no possibility of a single occurrence of *bank* in a sentence having both readings simultaneously, because there is no lexical entry for *bank* which is general between the senses. Similarly the logical forms of (15) lead to zeugma. Intuitively, the explanation runs along the following lines. As with lexical ambiguity in general, the interpreter uses pragmatic information in an attempt to decide whether (15′) or (15″) corresponds to the natural interpretation of (15). Knowledge about bank robbers favours interpreting *bank* as $bank_{\mathrm{org}}$, so this favours the logical form (15″) over (15′). But, on the other hand, knowledge about dam busters favours interpreting *bank* as $bank_{\mathrm{ground}}$, and so this favours (15′) over (15″). The interpreter cannot decide which of these two pieces of knowledge should take precedence and this irresolvable conflict in pragmatics produces a word pun effect: the conflict means the semantic ambiguity can't be resolved. MLEs alone do not provide a formal framework in which this pragmatic line of reasoning can be modeled. To do this it is necessary to link the compositional semantics to a logic for pragmatic reasoning, which predicts irresolvable conflict in appropriate cases. We will describe such a logic in §4.2.

In contrast to (15), the predicate for *teacher* in logical form will be non-committal as to whether the teacher referred to is male or female. So there will be single logical forms for (17a) and (17b) respectively. Therefore (17a) won't be interpreted as zeugmatic and (17b) will permit the domain of teachers that is quantified over to include a mixture of both male and female teachers.

If the connection with pragmatics can be formalised, the MLE account of zeugma and non-crossed readings appears to be adequate for homonymous words such as *bank*, where the financial institution and mound (of earth) senses are unrelated, and it is a coincidence that they share the same word form. The results of the ambiguity tests seem quite clear cut in these cases. But although examples like this are often taken as standard cases of ambiguity, they are actually quite infrequent compared to the cases where senses are related. For example, *bank* also has senses which correspond to the building used by a financial institution, the financial institution itself, the act of depositing money and the stock of tokens or money used in some games. At least the first two of these senses are regular and predictable, from the financial organisation sense, in that similar polysemy is found for other organisation denoting nouns or noun compounds such as *company*, *building society*, *credit union* and so on (Copestake and Briscoe 1995).

Unfortunately, zeugma and non-crossed readings often do not provide reliable ambiguity tests when applied to related senses; nor do MLEs provide satisfactory explanations of zeugma and non-crossed readings for polysemous words. We will demonstrate this by discussing some problems with the MLE approach.

2.2 Problems with tests

As is well known, ambiguity tests are difficult to apply consistently. Some of the problems point to inadequacies in the MLE account of zeugma. The zeugma and crossed readings tests can conflict with other tests for MLEs. Consider the sentences in (19):

(19) a. Nylon bristles and plastic handles are used to make cheap brushes.
b. Rembrandt used a brush and so did our janitor.
c. Rembrandt and our janitor used a brush.
d. The grandparents gave the two children brushes for Christmas.

The word *brush* is being used quite generally in (19a), indicating that even if there are separate entries in the lexicon for *paint brush*, *hair brush*, and so on, there must also be a lexical entry that is general between these items. But in contrast, according to the zeugma and crossed readings tests, examples (19b,c,d) indicate no such entry for *brush* which is general in this way. For example, (19b) and (19c) are both zeugmatic. But if there were a single lexical entry for *brush* they would each have a single logical form. So the MLE explanation of zeugma we outlined above would not apply in this case. Similarly, (19d) does not have a mixed reading, but there is nothing to prevent this under the MLE account. So the results of the zeugma and crossed reading tests for ambiguity conflict with the evidence of (19a).

As another example, consider *thigh*, which by default might be taken to refer to part of the human body. But it is possible to use it in a more general way, as illustrated by (20a). In contrast, while discussing a barbecue, *thigh* might be understood to refer to a piece of chicken meat, and in this context (20b) involves a pun.

(20) a. Thighs are the most muscular part of the vertebrate hind limb.
b. Bill likes mustard on his thighs but John prefers sun tan lotion.

Again, the evidence for ambiguity provided by the zeugma test conflicts with the other evidence.

It appears that the zeugma in (19b,c) and (20b) must be explained with more complex machinery than the MLEs alone. (19a) and (20a) are evidence that there are general entries for *brush* and for *thigh*, which if used to represent the logical forms of these sentences would predict that there was no zeugma. So, there must be some reason why *brush* and *thigh* are not simply left underspecified in the representation of the sentences (19b,c) and (20b). This idea can be fleshed out when one considers how real world knowledge and constraints on discourse coherence affect interpretation. If Rembrandt is using a brush, then by default, he is using a paint brush, and if a janitor is using a brush, then by default it is a cleaning brush. If one models this specialisation of *brush* to *paint brush* in sentences like

Rembrandt used a brush, then one can explain why (19b) is zeugmatic: there are conflicting messages about specialisation. Note that this account of zeugma does not appeal to distinct lexical entries for *paint brush* and *cleaning brush*; it appeals to *pragmatic* knowledge resources, and their effect on the interpretation of words like *brush* in a discourse context. The general idea is that even though there is a single lexical entry for *brush*, and so sentences like (19c) are not semantically ambiguous, one still gets irresolvable conflict in the pragmatic component—namely, conflict about how the sense of *brush* is specialised—making (19c) *pragmatically* ambiguous. Conflicting clues about specialisation will also explain the pun effect in (20b).

Another slightly different set of examples involve contexts where two related senses of a polysemous word can either yield zeugma or not, depending on the particular assertions being made, and how those assertions are presented. For example, words such as *thesis* and *book* systematically have usages which refer to the physical object, but can also be used to refer to the abstract contents (as illustrated in (4)). This gives rise to zeugma in some sentences and not in others: (21a) and (21b) involve predication over both the physical object (*has thousands of pages* and *is orange*) and the content (*is unreadable*). And yet (21a) and (21c) are better than (21b):

(21) a. That thesis has thousands of pages and is unreadable.
b. That thesis is orange and unreadable.
c. That orange thesis is unreadable.

So, according to the zeugma test, (21a) indicates that there is a lexical entry for *thesis* that is general between physical object and content, whereas (21b) indicates that there isn't any such lexical entry. Under standard assumptions about coordination and constraint based processing, it is impossible to account for (21a) if there are two different lexical entries for the different usages of *thesis*. Pustejovsky (e.g. 1991) and Copestake and Briscoe (1995) have argued that a single lexical entry can be used to represent both aspects of the word. Briefly, these accounts claim that lexical entries for nouns like *thesis* contain distinct attributes for form and constituency, which in the case of *thesis* will reflect the physical object / abstract content duality. Different predicates select for different attributes — for example, colour-denoting predicates such as *orange* select for the value of the form attribute, while *unreadable* selects for the contents.

Given this, one possible explanation of the contrast between (21a,b) is that even though both involve copredication—i.e., conjunction of predicates—on different aspects of *thesis*, the size of books is clearly relevant to their readability, and the colour of their covers is not. In other words, whether coordinating constituents is zeugmatic or not will depend not only on whether the lexical structure permits such co-predication, but also on how the assertions rhetorically link together, to contribute to the coherence of discourse. In this case, a close rhetorical connection can be found in (21a) but not in (21b), because in (21a), the causal connection between a book's size and its readability is inferrable. In contrast, (21c) doesn't involve co-predication, so there is no need to find a rhetorical relation and coherence constraints on rhetorical relations don't influence interpretation. We will discuss this approach in more detail in §3.1 and formalise it in §5.

2.3 Word Repetition

The MLE approach also fails to explain why zeugma and the absence of crossed readings tend to remain when the word form is repeated, as in (22) and (23) (see e.g. van Deemter 1990):

(22) John banked the money and then banked the plane.

(23) John had a file and Bill had a file.

The word pun effect in (22) is not as pronounced as in (15), but there is still a pun effect which the MLE account fails to predict. According to the MLE approach, (22) has four possible logical forms since each of the two occurrences of *bank* could correspond to either $bank_{\text{turn}}$ or $bank_{\text{deposit}}$. Similarly, (23) also has four logical forms. In the case of (22) the knowledge that you $bank_{\text{deposit}}$ money rather than $bank_{\text{turn}}$ it and that you $bank_{\text{turn}}$ planes rather than $bank_{\text{deposit}}$ them favours the logical form where the predicate which takes the money as an argument is $bank_{\text{deposit}}$ and the predicate which takes planes as an argument is $bank_{\text{turn}}$. So, unlike (15), there is no irresolvable conflict in pragmatics induced by the constraints imposed by logical form, and indeed, pragmatics successfully disambiguates the sentence. So, contrary to intuitions, the MLE account predicts that (22) isn't zeugmatic. Furthermore, two of the candidate logical forms for (23) invoke crossed readings, but intuitively these seem to be excluded, or, at least, strongly dispreferred.

The fact that (22) is zeugmatic and the interpretation of (23) isn't crossed must be explained by something other than the distinct lexical entries for the various senses of *bank* and *file*. We will claim that the natural interpretations of (22) and (23) essentially involve constraints on *discourse coherence* imposed by the way the phrases are rhetorically linked together. And it is these discourse coherence constraints that ultimately contribute to the zeugma in (22) and the absence of crossed readings in (23). For example in (23), the rhetorical link between the two clauses is closer under the non-crossed interpretation than the crossed one. One can see this intuitively by considering the topic of the discourse. If both John and Bill have a dossier sense of *file*, then the topic is *humans have a dossier*. If John has a dossier and Bill a tool, the topic is much more general: *humans have a physical artefact*, for instance. One prefers interpretations where discourse constituents are as closely connected as possible—this is a manifestation of Grice's Maxim of Relevance (Grice 1975, Sperber and Wilson, 1986), which can roughly be paraphrased as: the current constituent is as relevant as possible to the discourse context, subject to the constraints imposed by the cost of processing and inference. And so this Relevance heuristic of discourse interpretation explains why the non-crossed reading of (23) is preferred.

In (22), the two bits of real world knowledge concerning banking money and banking planes don't work independently, because there are constraints on how the two constituents (*John banked the money* and *John banked the plane*) are rhetorically linked together. There is a closer rhetorical connection between the two constituents if the senses of *bank* are the same for the two word forms. Again, this can be seen by considering the specificity of the topic of the discourse: it is *John banking things* if the senses of *bank* are the same, compared with *John doing things* if they're different. The Relevance heuristic of aiming for as close connections as possible between the clauses creates a feeling of irresolvable conflict among the various knowledge resources recruited, as in (15). To obtain an adequate rhetorical connection one must drop the pragmatic implicature that the money is $\text{bank}_{\text{deposit}}$ed, or the plane is $\text{bank}_{\text{turn}}$ed, so as to ensure that both senses of *bank* are the same. But one doesn't know which implicature to drop. So even though there is word repetition, thereby yielding a logical form for (22) which permits crossed readings, this is blocked by pragmatic considerations concerning discourse coherence. So, whereas (15) doesn't permit a mixed interpretation of *bank* on semantic grounds, (22) rules it out on pragmatic grounds. And both (15) and (22) yield irresolvable ambiguity. We will give a detailed, formal account of how this discourse machinery can complement the MLE approach, so as to predict the correct interpretations of (22) and (23) in §5.

3 Introducing a Pragmatic Approach

We have noted several problems with the MLE account of zeugma and non-crossed readings. We have claimed that it is necessary to complement the MLE account with a model of how pragmatic knowledge resources and discourse structure contribute to the overall interpretation. We have also shown that zeugma can occur even in the absence of MLEs. In this section we will illustrate three examples of zeugma in the absence of MLEs and illustrate informally how they arise in slightly different ways from the interaction of lexical semantics and pragmatics.

3.1 Weak coherence

In (21a), *has thousands of pages* is true of the physical thesis, whereas *unreadable* predicates of the content of the thesis.

(21) a. That thesis has thousands of pages and is unreadable.
b. That thesis is orange and unreadable.
c. That orange thesis is unreadable.

This can be achieved with a single lexical entry that has conventionalised aspects of an appropriate type, as outlined in §2.2. But we suggested above that this is not by itself a sufficient condition for acceptability. We must also find a rhetorical connection between the coordinated constituents in the VP. Intuitively, this is a way of checking that the Gricean Maxim of Relevance is being adhered to.

In (21a), real world knowledge can be used to infer a causal link between the number of pages of the thesis and its readability. This causal link permits us to link the constituents *the thesis has thousands of pages* and *the thesis is unreadable* with the rhetorical relation *Result*. They are also linked with *Parallel*, because of the structural and semantic relationship between the conjuncts (Scha and Polanyi 1988, Thompson and Mann 1987, Asher 1993a).

Inferring *Parallel* for (21b) is also possible, since the structural and semantic relationship is similar. But one cannot infer *Result* anymore, because our real world knowledge does not support a causal relationship between the constituents.

There are constraints that must be met for a *Parallel* relationship to be coherent (Asher, 1993a). One is that a common theme between the constituents must be found. This theme must be a proposition which: (a) is entailed by those that are connected together with *Parallel*, and (b) is simple in the sense that it is described by a single predicate argument structure (e.g., *x has physical properties* is simple whereas *x has an orange cover and x has thousands of pages* is not). One can compute the theme by generalising the predicates and arguments of the propositions being connected by *Parallel* (cf. Grover *et al*, 1994). In this case, the common theme is very general relative to the semantic content of the constituents being connected. This is because the only things that the predicates have in common are *properties of the thesis*. Compare this with a more specific theme such as *properties of the physical aspects of the thesis*, that would accompany a sentence like *That thesis is orange and has thousands of pages*. This general common theme makes the rhetorical connection in (21b) weak. The connection with (21a) was stronger, because although the *Parallel* relation had the same general common theme, there was also the *Result* relation which linked the constituents. So (21a) is better than (21b) because the rhetorical relation is stronger.

In contrast, we don't need a rhetorical connection for (21c) at all, since the VP does not contain coordinated constituents. The adjective *orange* in the NP can simply be used to ensure that the NP successfully refers, rather than requiring this property of being orange to have some relevant connection to the property in the VP. And therefore coherence constraints on rhetorical relations play no part in predicting whether (21c) is acceptable or not. Rather, having a single lexical entry for *thesis* that has conventionalised aspects of an appropriate type will suffice.

It is important to stress that this informal explanation of the difference in acceptability between the sentences in (21) uses a single lexical entry for *thesis* combined with pragmatic reasoning techniques. We will formalise this line of reasoning in §5.

3.2 Irresolvable conflict

(19a) is evidence that there is an entry for *brush* which refers to brushes in general, and covers paint brushes, cleaning brushes, hair brushes and so on. We must explain how even so, zeugma occurs in (19c).

(19) a. Nylon bristles make cheap durable brushes.

c. Rembrandt and our janitor used a brush.

Assume that the general entry for *brush* is used to build the logical form for (19c). So in contrast to (15), (19c) has a single logical form where reference to different kinds of brushes used by the agents is permitted. As in (21a,b), we must calculate the rhetorical connection between the propositions associated with the items linked by the coordination, in order to ensure that Grice's Maxim of Relevance is satisfied. In the case of (19c), this entails finding a rhetorical connection between *Rembrandt used a brush* and *our janitor used a brush* (since the reading being considered here is distributive). The only candidate relation is *Parallel* because of the semantic content and structural properties of these constituents, and this rhetorical relation comes with a coherence constraint: there must be a common theme.

Now, these two propositions verify two default laws from real world knowledge: the first is that Rembrandt, if he's using a brush, is normally using a paint brush; and the second is that our janitor, if he's using a brush, is normally using a cleaning brush. We will assume that the common theme is determined by the *pragmatic* interpretation of the constituents as opposed to only their truth conditional content, since themes are essentially a pragmatic property of the discourse. So suppose these defaults from real world knowledge were to fire. Then we would be attempting to compute a common theme between the propositions *Rembrandt used a paint brush* and *Our janitor used a cleaning brush.*

Using generalisation on the propositions, the common theme is *humans use a painting or cleaning brush* (to the exclusion of all other kinds of brushes). This is a bad theme, because it violates Grice's maxim of Quantity. Quantity stipulates that one must use referring expressions that are neither too general nor too specific for the purposes of the discourse. But in this discourse context, we cannot calculate why the discourse is about painting or cleaning brushes, but not any other kind of brushes (i.e., not brushes in general), and not about painting brushes alone, or cleaning brushes alone. So it seems that a referring expression has been used that is either not specific enough, or not general enough. In other words, this theme involves an interpretation of *brush* that violates Quantity.

The discourse is coherent, because we could find a common theme and so satisfy the coherence constraints on the rhetorical relation. But if the above world knowledge contributes to the interpretation of the sentence, then the coherence is weak, because the resulting theme is a bad one due to Grice's Maxim of Quantity. In such cases, one attempts to maintain a stronger coherence, by computing an interpretation of the constituents which avoids the weakness. In this case, changing the pragmatic interpretation of the constituents will change the common theme. Coherence would be improved by dropping at least one of the default implicatures about the kinds of brushes that are being used. If one did this, then the range of denotations for *brush* in the theme would change, and so the theme wouldn't violate Quantity anymore. But one can't decide which pragmatic implicature to drop: should one drop the implicature that Rembrandt is using a paint brush, or that the janitor is using a cleaning brush, or both? This *pragmatic ambiguity* is irresolvable, leading to the observed word pun in (19c).

Thus, the difference between (15) and (19c) is that the former constitutes irresolvable semantic ambiguity, whereas the latter, while semantically unambiguous, constitutes irresolvable pragmatic ambiguity, in that one cannot decide which implicatures apply.

3.3 Reinterpretation

The previous examples exploited both coherence constraints on discourse and lexical structure to predict zeugmatic effects. Here, we consider an example where psychological mechanisms of sentence processing also play a part. Consider again sentence (19c).

(19) b. Rembrandt used a brush and so did our janitor.

We assume that the interpreter calculates meaning incrementally and dynamically (e.g. Marslen-Wilson and Welsh 1978, Crain and Steedman 1985, Altmann and Steedman 1988). In other words, in (19b), the interpreter will calculate the pragmatic interpretation of *Rembrandt used a brush*, before going on to process the rest of the sentence. So, using real world knowledge, one specialises *brush* to *paint brush* before interpreting the elided clause, and attempts to attach this to the preceding one with a rhetorical relation.

Upon attempting to attach this elided clause, one must use the rhetorical relation *Parallel* (because of the structural properties of the sentence) and must use this relation to resolve the ellipsis (Asher 1993a, Kehler 1994). The only possibility is to interpret the elided clause as *our janitor used a brush.* Now one must calculate the pragmatic implicature of this whole sentence. This in part involves checking the pragmatic implicatures that were inferred for the first phrase (i.e., the pragmatic implicature that Rembrandt uses a paint brush). If one did not check this, then it would never be cancelled by subsequent context, and so the pragmatic inference could never be nonmonotonic.

By carrying out this check, one finds that there are clues from real world knowledge that the janitor uses a cleaning brush. But if Rembrandt uses a paint brush, then as in (19c), the common theme for the *Parallel* relation is bad, because Quantity is violated. Consequently, to maintain coherence one aims for a different pragmatic interpretation. And as in (19c), it is not possible to decide which implicature to drop: Rembrandt using a paint brush, or the janitor using a cleaning brush, or both. We assume that the need for reinterpretation or reanalysis can create a 'garden path' effect (e.g. Frazier 1979), and that such effects can result from incremental pragmatic as well as syntactic or semantic processing decisions (e.g. Crain and Steedman, 1985). So the zeugmatic effects in (19b) are a consequence of both reinterpretation and indecision. First, reinterpretation takes place (the interpreter cannot rely on Rembrandt's brush being a paint brush anymore). And second, there is now indecision about how to interpret *brush.* In other words, as in (19c), although there is no semantic ambiguity, there is irresolvable pragmatic ambiguity.

Reinterpretation can also account for the word pun effects in (20b) and (9).

(20) b. Bill likes mustard on his thighs, but John prefers suntan lotion.

(9) I tried to take the plane to Chicago but it was too heavy.

In these cases, the most plausible senses available at the end of the initial clause for *thighs* and *take* must be revised when the second clause has been processed in order to infer the *Parallel* relation and resolve the ellipsis and anaphor, respectively. However, the zeugmatic effect derives from the retraction of the initial plausible lexical senses rather than from the impossibility of arriving at a consistent (though bizarre) reading.

3.4 The Three Components of Zeugma

In the three informal examples we gave above, we described three distinct ways in which zeugmatic effects can be created:

1. The rhetorical connection between constituents in the discourse is weak, and this weak coherence makes the discourse sound odd.
2. Constraints on discourse coherence can produce irresolvable conflict in the pragmatic interpretation of constituents in the discourse, and this irresolvable pragmatic ambiguity gives a word pun effect.
3. Incremental processing in pragmatics and discourse coherence force a reinterpretation.

Furthermore, these can be intermixed, so for instance some examples involve both an effect of reinterpretation and irresolvable conflict. This is a heterogeneous account of zeugma, but there are common threads. Pragmatic reasoning is central to zeugma, regardless of the presence of MLEs. And in all cases, the account of zeugma rests on the interaction between three components: lexical structure, (incremental) pragmatic reasoning and the rhetorical structure of discourse.

Formalising the above informal accounts therefore rests on providing general axioms that guide the information flow between lexical structure, pragmatic reasoning and constraints on discourse coherence. To model this information flow, we must link a theory of lexical structure to a theory of the rhetorical structure of discourse and the coherence constraints it imposes, and a theory of how pragmatics affects the interpretation of discourse. The rest of this paper is concerned with integrating these three components, in order to demonstrate how this integration leads to an account of zeugma and non-crossed readings for all the above examples. In the next section, we'll review the rhetorical structure and pragmatic frameworks that we'll use and then give the formal account of the examples we described in this section in §5.

4 Discourse Structure and Pragmatic Reasoning

This section first describes the background theory of discourse structure assumed here and the logic which is used to represent real world knowledge and also to compute rhetorical relations. We then go on to introduce a new discourse constraint on word meaning, which generalises the one assumed by Asher and Lascarides (1995a) in order to account for sense disambiguation in a discourse context. We discuss this in detail here because the new form of the rule will contribute to the explanation of zeugma involving weak coherence.

4.1 Discourse Relations

We need a theory of discourse structure which specifies discourse coherence constraints, and we use these to explain zeugmatic effects and the preference for non-crossed readings as suggested in §3. We will use Asher's (1993a) Segmented Discourse Representation Theory (SDRT) for this purpose. SDRT is an extension of DRT (Kamp and Reyle, 1993) that assumes richer discourse representations. In SDRT, an NL text is represented by a segmented DRS (SDRS), which is a pair of sets containing the DRSs or SDRSs (representing respectively sentences and text segments) and discourse relations between them. Discourse relations, which link constituents together, are modeled after those proposed by Hobbs (1985), Polanyi (1985) and Thompson and Mann (1987). We will mention four: *Narration*, *Result*, *Contrast* and *Parallel*. Examples of these are given below:

(24) a. Max stood up. John greeted him.
Narration (because the textual order of events matches temporal order)

b. John pushed Max. Max fell.
Result (because the pushing caused the falling).

c. Max has black hair, but Mary has brown hair.
Contrast

d. Max has black hair, and Mary has black hair too.
Parallel

These discourse relations place constraints on discourse coherence, which are defined in terms of the structures and semantic contents of the constituents being attached together. Asher (1993a) and Lascarides and Asher (1993) discuss and motivate the coherence constraints for each of the above relations in detail. For example, *Narration* entails that the textual order of events must match temporal order. There must also be a distinct common topic, that summarises what the narrative describes. In contrast, *Result* doesn't require a distinct common topic, but it does require a causal relation between the events described in the constituents. These constraints are formally specified below: in what follows, e_α stands for the main eventuality described in the constituent α and $\gamma \Downarrow \alpha$ means γ is a topic or summary of α:

- `Axiom on Narration`
  $\Box(Narration(\alpha, \beta) \rightarrow e_\alpha \prec e_\beta)$
- `Distinct Common Topic`
  $\Box(Narration(\alpha, \beta) \rightarrow \exists\gamma(\gamma \Downarrow \alpha \wedge \gamma \Downarrow \beta \wedge \neg(\alpha \Downarrow \beta) \wedge \neg(\beta \Downarrow \alpha)))$
- `Axiom on Result:`
  $\Box(Result(\alpha, \beta) \rightarrow cause(e_\alpha, e_\beta))$

Contrast and *Parallel* both require a partial structural isomorphism between the constituents. In addition, *Contrast* requires a contrasting theme, and *Parallel* a common theme. Asher (1993a) defines precisely how to calculate the structure of each constituent and also specifies which mappings between the structures are sufficiently close to bijections that the structural constraint on *Parallel* and *Contrast* is met. We will explicate these definitions in §5, when we use them in our formal analysis of zeugma. We also need a theory for computing themes, so that we can check that the constraints on themes are satisfied too — an approach similar to that of Grover *et al.* (1994). They use constraint generalisation on the constituents, as defined by the semantic type hierarchy that accompanies the lexicon. We refine this slightly: the common theme will be the generalisation of the propositions corresponding to the *pragmatic* interpretations (with respect to the discourse context) of the constituents being connected together. So for example, the common theme of (25) will be the generalisation of *Rembrandt used a paint brush* and *Rubens used a paint brush*, as opposed to the generalisation of *Rembrandt used a brush* and *Rubens used a brush.*

(25) Rembrandt and Rubens used a brush.

This is because the implicature that the brushes are paint brushes survives in this context. And so the common theme for (25) is *artists used a paint brush.*

Asher shows how one can resolve VP ellipsis as a byproduct of checking the coherence constraints on *Contrast* and *Parallel.* For example in (26), the elided clause is resolved to *and Rubens used a brush* because of the structural coherence constraints on *Parallel.*

(26) Rembrandt used a brush and so did Rubens.

The details of how the coherence constraints contribute to the interpretation of VP ellipsis will be given in §5, with respect to the zeugmatic sentence (19b):

(19) b. Rembrandt used a brush and so did our janitor.

4.2 Pragmatic Reasoning

We need to represent default real world knowledge, such as *If Rembrandt uses a brush, then normally it's a paint brush.* We also need a theory which computes the rhetorical connections between meanings of segments of text via the interpreter's background knowledge. For example, we want to be able to predict that the constituents in (21a) are connected together with *Result* and *Parallel*, whereas in (21b) the relation is *Parallel* alone.

(21) a. That thesis has thousands of pages and is unreadable.

b. That thesis is orange and unreadable.

The informal lines of reasoning about zeugma given in §3 cast light on the kind of intuitively compelling reasoning patterns we need to validate in the logic that underlies pragmatic reasoning. We must use a logic which automatically calculates the interaction between various knowledge resources such as lexical knowledge, real world knowledge and Gricean style pragmatic maxims which takes place during interpretation. In particular, it must detect conflict, and predict when it's resolvable and when it's not. This is because, on occasion, the various knowledge resources yield conflicting messages about how to interpret a word, and the logic must predict 'automatically' if and how this conflict is resolved (cf. our discussion of (15)).

We will exploit state-of-the-art techniques in commonsense reasoning from AI to reason about pragmatic interpretation. The particular theory we'll use is DICE (Discourse In Commonsense Entailment) (Lascarides and Asher 1991, 1993). DICE represents as default rules the interpreter's background knowledge about how syntax, compositional semantics, Gricean style pragmatic maxims and real world knowledge affect overall interpretation. These are used to compute both the pragmatic interpretation of constituents and the rhetorical relations between them. It is the theory of discourse attachment by which NL text is translated into discourse representations in SDRT.

The logic that underlies DICE is Asher and Morreau's (1991) commonsense entailment (CE). This is a conditional logic in which a formula of the form $A > B$ is supplied with a model theoretic and proof theoretic interpretation, which roughly means "If A, then normally B". Default knowledge in DICE is represented as conditionals of this form.

CE exploits two notions of validity, a monotonic one ($\models$) and a nonmonotonic one ($\approx\!\!\!\!|$). There are at least six attractive properties of $\models\!\!\!\!\approx$.

1. All the consequents of $\models$ are computed before one starts to calculate the consequents of $\mathrel{|\!\approx}$.

2. $\mathrel{|\!\approx}$ validates Defeasible Modus Ponens (DMP): when the default laws whose antecedents are verified all have consequents that are consistent with the knowledge base (KB) and with each other, then all the consequents are nonmonotonically inferred.

3. $\mathrel{|\!\approx}$ validates Specificity: in cases of conflicts among defaults where one has a more specific antecedent than the others, $\mathrel{|\!\approx}$ will verify the conclusions of the more specific default.

4. $\mathrel{|\!\approx}$ validates Skepticality: in cases of conflicts among defaults where there is no default that has a most specific antecedent relative to the others, $\mathrel{|\!\approx}$ won't verify any conclusions of the defaults. Because of Specificity and Skepticality, CE automatically predicts when knowledge conflict can be resolved and when it can't.

5. $\mathrel{|\!\approx}$ is robust in that if $\Gamma \mathrel{|\!\approx} \phi$ then ϕ will survive as a consequence of the premises Γ augmented with logically independent information.

6. For each deduction $A \mathrel{|\!\approx} B$ there is a corresponding embedded default in the object language (that is, a formula in which one occurrence of the connective $>$ occurs within the scope of another) which links boolean combinations of the formulae A and B, and which is verified to be true. We gloss this embedded default formula as $\sqsupset(A, B)$. So $\sqsupset(A, B)$ means $A \mathrel{|\!\approx} B$. This amounts to a weak deduction theorem. The object language formula $\sqsupset(A, B)$ means that A nonmonotonically yields B in the metalanguage.

We will also on occasion use the notation $\sqsupset_{KB}(A, B)$ to mean $\sqsupset(KB \wedge A, B) \wedge \neg\sqsupset(KB, B)$. In other words, $\sqsupset_{KB}(A, B)$ means that B nonmonotonically follows from the KB augmented with A, but not from the KB alone. The details of the logic can be found in (Asher and Morreau 1991, Asher 1993b).

We can use CE to represent the knowledge recruited to compute the pragmatic interpretation of discourse. This knowledge is of many different types: knowledge of the real world; knowledge of Gricean-style pragmatic maxims; knowledge of the communicative intentions of the author, and so on. We have examined in detail elsewhere how to represent these different pieces of knowledge as formulas of CE in DICE (Lascarides and Asher 1991, 1993; Lascarides and Oberlander, 1993; Asher and Lascarides, 1994, 1995a).

Real world knowledge such as *When Rembrandt uses a brush, it's normally a paint brush* can be represented via the schemata below:

- `Brush Purpose`:
  $\Box(brush(y) \rightarrow purpose(y, brush\text{-}purpose(e)))$

- `Rembrandt:`
  $agent(e, x) \wedge rembrandt(x) \wedge instrument(e, y) \wedge purpose(y, \sqsupseteq paint(e'))$
  $\quad > purpose(y, paint(e'))$
  where $purpose(y, \sqsupseteq paint(e'))$ is an abbreviation for: $purpose(y, P(e')) \wedge (paint(e') \rightarrow P(e'))$

That is, when Rembrandt is the agent of an event where the instrument is something whose purpose is more general than painting, then normally, the purpose of that instrument in that particular event is painting. Information about purposes of brushes—as given in Brush Purpose—has to be part of common sense knowledge for independent reasons. For our current concerns it is irrelevant whether or not this information is also accessible to the lexicon.

There is also a similar rule, which links the purpose of agents to the purposes of their instruments: Specialisation will predict that when a janitor (whose 'purpose'/occupation is to clean) uses a brush, then normally it is a cleaning brush:

- `Specialisation`:
  $agent(e, x) \wedge purpose(x, P(e')) \wedge instrument(e, y) \wedge purpose(y, \sqsupseteq P(e'))$
  $\quad > purpose(y, P(e'))$

These laws enable CE to predict that *Rembrandt used a brush* is pragmatically interpreted as *Rembrandt used a paint brush*, and similarly for *janitor* and cleaning brush. The pattern of inference is DMP on Rembrandt and Specialisation respectively.[3] We would need to extend these representations of real world knowledge to handle quantified sentences such as *No artist used a brush*, but for the sake of simplicity and brevity, we ignore this issue here.[4]

In addition, we can use CE to infer rhetorical relations—such as *Contrast*, *Parallel*, *Result* and *Elaboration*, among others—that connect the meanings of segments of text together. DICE uses the interpreter's background knowledge to build SDRSs from NL text (Lascarides and Asher 1991, 1993). The way this is done is as follows. First, a constraint-based grammar with a DRT semantic component (cf. Frank and Reyle, 1995) is used to build up the DRS for the clause that is currently being processed. Thus DRSs for clauses are built in a monotonic fashion via information from syntax, much as in traditional DRT. However, in contrast to traditional DRT, one then reasons in DICE about how this DRS attaches to the preceding discourse structure with a rhetorical relation. And in contrast to the monotonic construction of the DRS representing the clause, the inferences in DICE underlying the choice of rhetorical relation for attachment is nonmonotonic, because rhetorical relations aren't always linguistically marked. Rather, they are inferred on the basis of clues from real world knowledge, the discourse context, communicative intentions and Gricean style pragmatic maxims. A further contrast to traditional DRT is that once the semantic representation of the

[3] We don't rely on Specialisation to hanlde the inference about Rembrandt, because we wish to remain agnostic as to wether Rembrandt's 'purpose' is to paint.

[4] It would be relatively straightforward to extend the rules to handle these cases; one need only use DRS duplex conditions in the antecedents of these rules, where the quantifier is left underspecified.

clause currently being processed is built by the grammar, it is evaluated against the model. This is because DICE uses the semantic content of the clause as a clue about which rhetorical relation to use to attach it to the representation of the discourse built so far. In traditional DRT, semantic interpretation doesn't occur until the representation of the whole discourse is built.

As well as real world knowledge such as Rembrandt and Specialisation, there are default rules like Result in DICE for computing rhetorical relations: if β is to be attached to α with a rhetorical relation where α is part of the discourse structure τ already, and the event described in α caused that described in β, then normally, the rhetorical relation is *Result*.

- `Result:`
  $\langle \tau, \alpha, \beta \rangle \wedge \mathit{cause}(e_\alpha, e_\beta) > \mathit{Result}(\alpha, \beta)$

Result uses information about the constituents α and β to come to nonmonotonic conclusions about how they're connected together. In general, axioms in DICE that are about computing rhetorical relations are of this general form: $\langle \tau, \alpha, \beta \rangle \wedge \phi > R(\alpha, \beta)$, where ϕ is information about α, β and the discourse context τ, and R is a particular discourse relation.

4.3 A Discourse Constraint on Word Interpretation

DICE is sufficiently rich to encode how the rhetorical structure of the discourse affects the way words are interpreted. Indeed, we'll show that the rule in DICE which encodes how rhetorical structure affects lexical sense disambiguation can also create zeugma. To explicate how rhetorical structure affects word meaning—and in particular, disambiguation—we give a detailed analysis of some cases which involve the rhetorical relation *Elaboration*. As described in Asher and Lascarides (1995a), $\mathit{Elaboration}(\alpha, \beta)$ can be inferred by reasoning that the event condition in β is a subtype of that in α.

- `Elaboration`
  $\langle \tau, \alpha, \beta \rangle \wedge \mathit{Subtype}(\alpha, \beta) > \mathit{Elaboration}(\alpha, \beta)$

Elaboration states that if β is to be attached to α with a discourse relation, where α is part of the SDRS τ built already, and β is a subtype of α, then normally, the discourse relation is *Elaboration*. For the sake of brevity, we don't repeat the rules by which one infers $\mathit{Subtype}(\alpha, \beta)$, since the details of this aren't central to our main concerns here. These rules used lexical information, and are powerful enough to infer that $\mathit{Subtype}(27\text{a}, 27\text{b})$ holds for the text (27), because learning to use a paint brush is a subtype of learning to paint, and the janitor is plausibly a member of school staff.

(27) a. All the school staff learned how to paint.
     b. Even the janitor learned to use a paint brush.

So Specificity in DICE between Narration given below (from Lascarides and Asher, 1991) and Elaboration predicts that (27a) and (27b) are linked by *Elaboration*.

- `Narration`: $\langle \tau, \alpha, \beta \rangle > \mathit{Narration}(\alpha, \beta)$

Inferring that (27) was an elaboration crucially uses the information that the brush in (27b) is explicitly stipulated to be a paint brush. Whereas learning to use a paint brush is a subtype of learning to paint according to the rules in Asher and Lascarides (1995a), learning to use a (general) brush isn't—the brush in question may be a hair brush and learning to use a hair brush is not a subtype of learning to paint. And so without the information that the brush is a piant brush, we can't infer *Subtype*(α, β) so that the rule Elaboration applies. But what about the discourse (28), where there is no such explicit stipulation about the brush, and indeed, where real world knowledge about janitors favours cleaning as opposed to paint brushes?

(28) a. All the school staff learned how to paint.
b. The janitor learned to use a brush.

Intuitively, this text is still an elaboration, and one *infers* that the brush used is a paint brush, because of the discourse context. We explain here how the discourse context wins over the real world knowledge about the brushes janitors use, because the rules in DICE that encode these discourse effects on word meaning will also create zeugma in the cases we've discussed.

Asher and Lascarides (1995a) provided rules in DICE which encoded interactions between words and discourse, in service to sense disambiguation in a discourse context. They paid special attention to cases where information in the discourse context conflicts with information within the sentence as to how a word should be interpreted. Asher and Lascarides argued in favour of a rule they called Lexical Impotence. This rule ensured that, if disambiguating a word according to the information from within a sentence led ultimately to discourse incoherence, this disambiguation was blocked and, instead, the word was disambiguated via information in the preceding discourse context.

This was used to model the lexical disambiguation in (29).

(29) a. The judge asked where the defendant was.
b. His lawyer said he was drinking across the street.
c. The court bailiff found him slumped underneath the bar.

Suppose one disambiguates *bar* in (29c) according to the clues from within (29c) itself (so the bar is the courtroom bar). Then one cannot connect this to (29b) with a rhetorical relation. This is because the only candidate relation is *Narration*, but if the defendant is at the courtroom bar in (29c), then a spatial constraint on *Narration* is violated because he's across the street in (29b). Lexical Impotence ensured that the discourse information and need for coherence wins out in this case, and *bar* is disambiguated to the pub sense, so that *Narration*'s spatial constraints aren't violated and coherence is established.

But Lexical Impotence cannot explain the interaction between words and discourse in (28), because if the brush is a cleaning brush (as the real world knowledge in (28b) favours), then a narrative, list or parallel relation can be inferred. So a rhetorical connection can be found when the sentential information drives the interpretation of the word *brush*; this sentential information doesn't lead to incoherence. So Lexical Impotence wouldn't apply when interpreting (29), and we must model why the preference triggered by the real world knowledge about janitors is blocked in some other way. We'll generalise Lexical Impotence here, to increase the empirical coverage and further explicate how rhetorical structure influences how words are interpreted. The rule for interaction given here will handle both (29) and (28). And it will capture the following intuition: we prefer not to infer things that have an adverse effect on the quality of coherence in the discourse.

Consider (28) once again. Suppose that Specialisation fired, so that the brush in (28b) is interpreted as a cleaning brush. Then the event (28b) isn't a subtype of that in (28a), because the janitor is learning to use a cleaning brush (i.e., learning to clean) in (28b), rather than learning to paint. So the only rule for computing the rhetorical connection that would be verified is Narration. Narration has coherence constraints: one must find a distinct common topic. This topic must summarise both α and β. The only candidate here is *school staff learning to paint or clean*. But one could argue that, as in (19b), although this is coherent, it is only weakly so, because the topic is a disjunction of the individual sentential topics. It is not clear what the discourse is 'about'.

Assume that $\mathit{weak}(\tau)$ means that the SDRS τ is at best only weakly coherent. Assume furthermore that a weak theme or topic yields only weak discourse coherence. Then, intuitively, one wants to avoid discourse interpretations that yield an SDRS τ where $\mathit{weak}(\tau)$ is true.[5] In principle, the interpretations of constituents that yield weak discourse coherence can be blocked, if (a) the weak coherence hinges essentially on a particular pragmatic inference on one of the constituents, and (b) we block those pragmatic inferences in those cases, so that we can attempt to improve overall discourse coherence. In other words, a pragmatic implicature of a particular constituent may be blocked, if that pragmatic implicature has an adverse effect on overall discourse coherence.

We can capture this intuitively compelling interpretation strategy as a rule in DICE. We will gloss all monotonic information about the constituents α and β as $\mathit{Info}(\alpha)$ and $\mathit{Info}(\beta)$ (sometimes abbreviated to $\mathit{Info}(\alpha, \beta)$).[6] Let *KB* be the interpreter's knowledge base: this includes the update function $\langle\tau, \alpha, \beta\rangle$, the semantic content of τ, α and β, and the rules of DICE. Then the Interpretation Constraint below captures the intuition that people try not to infer propositions that lead to weakly coherent or incoherent discourse. It states: if (a) β is to be connected to α with a discourse relation, and β and α are both true, and (b) if the KB that includes not only the update task of β to α, but also the information β', nonmonotonically leads to a discourse of only weak coherence or no coherence at all, then normally (c) β' doesn't hold.

- `Interpretation Constraint`
  (a) $\langle\tau, \alpha, \beta\rangle \wedge \mathit{Info}(\alpha, \beta) \wedge$
  (b) $\sqsupset_{KB}(\beta', \mathit{weak}(\tau \cup \beta)))$
  (c) $> \neg\beta'$

This rule captures a dependence between pragmatic inferences from information within the sentence and information from the discourse context. Consider the case when β' is a pragmatic implicature of β in isolation of the discourse context (e.g., $\mathit{Info}'(\beta) > \beta'$ is a rule in the KB, where $\mathit{Info}'(\beta)$ is a gloss for some relevant information pertaining to the syntagmatic representation of β). If assuming β' yields only weak discourse coherence when one attempts to attach the constituents together, then the Interpretation Constraint blocks this pragmatic inference, because its consequent will conflict with that of the default rule $\mathit{Info}'(\beta) > \beta'$, and it is also more specific than this rule (because the antecedent of the Interpretation Constraint entails $\mathit{Info}'(\beta)$ since it entails *all* monotonic information about β). So β' would follow via DMP from a KB that included just β in isolation of discourse τ. But in the context of τ, β' does not follow because the Interpretation Constraint blocks it.

The Interpretation Constraint can now replace Lexical Impotence because it predicts that sentential clues about disambiguation are blocked when they lead to discourse incoherence. To see this, consider (29) again. (29c) on its own is pragmatically interpreted as involving the courtroom bar. For the sake of simplicity, assume this is encoded in DICE as $\mathit{Info}'(\gamma) > \gamma'$, where γ is the semantic representation of (29c), and γ' asserts that the *bar* is the courtroom bar.[7] Now, we must attach this γ to the semantic representation β of (29b). The only candidate is *Narration* (via the axiom Narration). But *Narration* has spatial constraints. These were formally spelled out in Asher and Lascarides (1995a); informally, they capture the intuition that objects stay where they are in a narrative, unless one can infer otherwise. If the bar is the courtroom bar—i.e., the KB contains γ'—then *Narration*'s spatial

[5] Providing formal truth conditions for $\mathit{weak}(\tau)$ would involve complex reasoning about the semantic content of τ and the cognitive states of the participants in the discourse. This goes beyond the purposes of this paper, and so we rely in this paper on intuitive predictions of when $\mathit{weak}(\tau)$ is true, such as the intuition that the discourse theme violates a Gricean maxim.

[6] We assume that in cases where α is, for example, *Rembrandt used a brush*, $\mathit{Info}(\alpha)$ will contain the information $\mathit{brush}(y) \wedge \mathit{purpose}(y, \mathit{brush\text{-}purpose}(e'))$, because this is monotonically inferrable from the truth conditional content of α. However, $\mathit{Info}(\alpha)$ will *not* contain the information $\mathit{purpose}(y, \mathit{paint}(e'))$ since this is only nonmonotonically inferrable.

[7] In Asher and Lascarides (1995a), this rule was represented as $FS_{bar} > FS_{bar_1}$, where FS_{bar} is all the information in the feature structure that represents sentence (29c), and FS_{bar_1} is that feature structure, with *bar* replaced by its courtroom sense. We have changed it here for the sake of simplicity, to the more general notation above; nothing hinges on this difference.

constraint is violated, so that no discourse relation can be inferred. Using this same knowledge here, this means that $\sqsupset_{KB}(\gamma', \textit{weak}(\tau \cup \gamma))$ holds—$\tau \cup \gamma$ is incoherent when γ' is part of the KB. So clauses (a), (b) and (c) of the Interpretation Constraint are verified, with β, γ and γ' substituted for α, β and β' respectively in the schema. So normally, $\neg\gamma'$ holds. This conflicts with the rule $\textit{Info}'(\gamma) > \gamma'$ which also forms part of the KB. But the Interpretation Constraint is more specific, and so $\neg\gamma'$ is inferred. That is, the bar is not the courtroom bar. This was exactly the inference that arose from Lexical Impotence, and so the Interpretation Constraint can replace this rule.

The Interpretation Constraint can also explain how the rhetorical structure of (28a,b) affects the interpretation of *brush*. It leads to an inference that the brush in question is a *paint brush*, thereby overriding the real world knowledge about the kinds of brushes that janitors normally use, given in Specialisation. Let the DRSs representing the compositional semantic content of (28a,b) be α and β respectively. Then let β' be the condition that the brush is a cleaning brush. Specialisation can be glossed as $\textit{Info}'(\beta) > \beta'$, where $\textit{Info}'(\beta)$ is the relevant information about β that ensures the antecedent of Specialisation is verified. So $\textit{Info}'(\beta) > \beta'$ applies.

But the condition (b) of the Interpretation Constraint is verified with this β', because as we've stated, interpreting the brush as a cleaning brush yields a nonmonotonic inference that the discourse relation is *Narration*, with a weak common topic and so $\textit{weak}(\tau \cup \beta)$ is inferred with respect to this information. That is, $\sqsupset_{KB}(\beta', \textit{weak}(\tau \cup \beta))$ holds. So, the Interpretation Constraint applies, and its consequent is $\neg\beta'$; that is, the brush is not a cleaning brush. This conflicts with Specialisation. But the Interpretation Constraint is more specific, because its antecedent contains as a conjunct all the information about β that's in the antecedent of Specialisation. So by Specificity, $\neg\beta'$ is inferred. That is, the brush is not a cleaning brush. Informally, because the pragmatic interpretation of (28b) in isolation of (28a) has an adverse effect on discourse coherence (although not making it totally incoherent) when one attaches it to (28a), this pragmatic interpretation is blocked by the Interpretation Constraint, so as to improve overall coherence.

Indeed, the antecedent of the Interpretation Constraint applies for each β' where the brush is a tooth brush, a hair brush and so on (hence we *don't* remain neutral about what kind of brush the janitor used). So for these β', DMP on the Interpretation Constraint yields the conclusion that the brush is not a tooth brush, a hair brush and so on. But the Interpretation Constraint does *not* apply for the β' where the brush is a paint brush. When this β' is added to the KB, then one can infer $\textit{Subtype}(\alpha, \beta)$ holds, and thus Elaboration applies and by Specificity on this and Narration, $\textit{Elaboration}(\alpha, \beta)$ is inferred. Thus adding this β' to the KB does not yield a weak discourse, and so the antecedent of the Interpretation Constraint doesn't apply for this β', and so one doesn't conclude $\neg\beta'$ in this case. So the Interpretation Constraint ultimately restricts the brush to being a paint brush, because all other possible interpretations of *brush* are ruled out by the Interpretation Constraint, on the basis that they yield a discourse that is only weakly coherent. So DICE, with the aid of the Interpretation Constraint, correctly predicts that (28) is an elaboration, and the brush in question is a paint brush. We'll see in §5 that the Interpretation Constraint not only disambiguates words and causes sense shifts via rhetorical structure, but it also plays a crucial part in creating zeugma in the absence of multiple lexical entries in the lexicon.

5 Integrating the Components

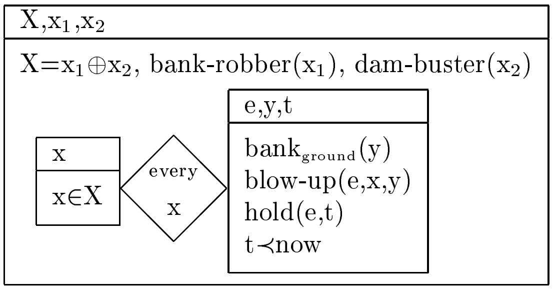
In the previous section we saw how rules at the discourse level can have effects on interpretations; for example the Interpretation Constraint causes *brush* to be interpreted in a particular way in (28), which conflicts with the clues about how it should be interpreted given our knowledge about the kinds of brushes janitors use. We'll now explain further how lexical information, compositional semantics, pragmatics and discourse coherence interact, in order to explain how zeugma and the absence of crossed readings occur.

5.1 Homonyms

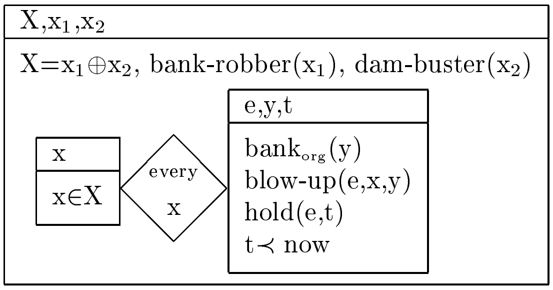
We start with an example that's zeugmatic due to homonymous ambiguity. We assume that unrelated homonyms are encoded by completely distinct lexical entries. For example *bank* has two completely unrelated lexical entries corresponding to the financial institution and raised ground senses. Because the structures are distinct, and there is no lexical entry which corresponds to their generalisation, there is no possibility of a single use of *bank* in a sentence having both readings simultaneously. So, consider sentence (15) and its logical forms once more.[8]

(15) Some dam busters and bank robbers blew up banks.

(15′)



(15″)



To check the coherence of this sentence, we must do two things: (a) resolve the semantic ambiguity; and (b) check the coherence of linking the coordinated constituents together. We concentrate here on the former task, because the latter depends on it being accomplished.

We must consider the pragmatic inferences that accompany (15). Two pieces of knowledge are relevant. First, if a bank robber is blowing up an object, which is either a financial institution or a mound of earth but we do not know which, then in the absence of further information to the contrary, we jump to the conclusion that the object in question is a financial institution. Conversely, for dam busters, the default is that the object is a mound of earth. These two rules are represented in CE as follows:

- `Bank Robber`:
  $\textit{bank-robber}(x) \wedge \textit{blowup}(e,x,y) \wedge (\textit{bank}_{\text{ground}}(y) \vee \textit{bank}_{\text{org}}(y)) > \textit{bank}_{\text{org}}(y)$
- `Dam Buster`:
  $\textit{dam-buster}(x) \wedge \textit{blowup}(e,x,y) \wedge (\textit{bank}_{\text{ground}}(y) \vee \textit{bank}_{\text{org}}(y)) > \textit{bank}_{\text{ground}}(y)$

Bank Robber and Dam Buster are intended for expository purposes in this paper. They will do for the sentences involving *bank* that we consider, but would have to extended to deal with quantified sentences, such as *No modern bank robber blows up banks.* The rules above could be derived from

[8] For the sake of brevity, we ignore the collective readings and concentrate on the distributive ones. The two logical forms thus arise because of the two lexical entries for *bank*. The notation ⊕ denotes an operator which forms plural individuals (Link 1983).

more general rules about the kinds of things that dam busters and bank robbers blow up (see Asher and Lascarides (1995a) for an account of how this is done). But we gloss over this for now since it's not relevant to the account of zeugma. For our purposes the only crucial point concerns the logical relationship between the rules; we must ensure that the interaction between this and the compositional semantics of (15) has the desired effect. These rules ensure that although *A bank robber blew up a bank* and *A dam buster blew up a bank* are ambiguous sentences at the semantic level, they are disambiguated at the pragmatic level (under the assumption that the interpreter's KB contains the knowledge that one or other of the logical forms is true).

How do Bank Robber and Dam Buster affect the pragmatic interpretation of (15)? We assume the KB contains the information that (15′) or (15″) is true, but not both. So the antecedents of Bank Robber and Dam Buster are verified. Now, we are assuming that (15) is distributive, so the embedding functions that verify the DRSs (15′) and (15″) permit bank robbers and dam busters to be agents to different events, involving different banks. But even so, neither logical forms permit a *mixture* of banks to be blown up; $15' \vee 15'' \wedge \neg(15' \wedge 15'')$ thus entails that either all the banks involved are financial, or they're all earth banks. So even though the antecedents of Bank Robber and Dam Buster could be verified by different agents, different banks and different events, the rules are nevertheless dependent. Because the logical forms don't permit bank robbers to blow up financial institutions while the dam busters blow up earth banks, the above real world knowledge conflicts. Moreover, it is irresolvable conflict, because the antecedents of Bank Robber and Dam Buster are logically unrelated. Consequently by Skepticality in CE, the interpreter comes to no conclusions about whether *bank* should be interpreted as $bank_{\text{org}}$ or $bank_{\text{ground}}$. So even after having brought pragmatic reasoning to bear, it is not possible to decide how to resolve the lexical ambiguity. In other words, irresolvable conflict in pragmatics produces a word pun effect in (15). The semantic ambiguity of (15) is irresolvable, because of Skepticality in Pragmatics. The interpreter does not even come to the point of checking how the coordinated constituents rhetorically link together, because there was no resolution of the truth conditional content of the constituents.

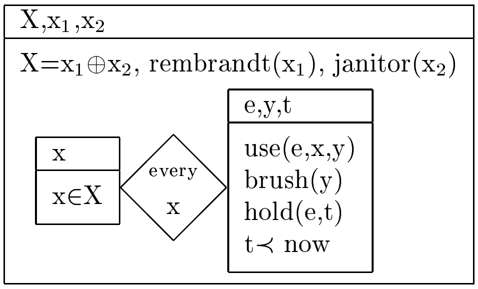
5.2 Syntax and Pragmatics: A Single Lexical Entry

Now consider an example where there is a single lexical entry, but nevertheless there is a feeling that the sentence involves a word pun:

(19) c. Rembrandt and our janitor used a brush.

We assume that there is a single lexical entry for *brush*, which refers to paint brushes, cleaning brushes, tooth brushes and so on. These subclasses of brush can also be represented in the lexicon, but they would be specialisations of the more general entry. Therefore there is only one logical form for (19c):

(19) c′.



However, pragmatics still yields a word pun effect. This is how it occurs.

We must check the coherence of coordinating the constituents in (19c) together. SDRT assumes that rhetorical relations relate propositions rather than properties. Extending SDRT to incorporate rhetorical connections between abstracted DRSs is straightforward, but it is a technical exercise that doesn't cast light on the points we want to make about zeugma. So for the sake of clarity, we make the simplifying assumption that in cases like (19c), the task is to compute the rhetorical connection between the propositions *Rembrandt used a brush* and *our janitor used a brush*. Let these propositions be α and β respectively. Using the mechanisms available in SDRT, one infers in the monotonic component of CE that the relation is *Parallel*. We must check that the coherence constraints on *Parallel* are verified: that is, we must find a common theme.

Now, suppose that we stipulate that β' is the DRS conditions that the brush that Rembrandt is using is a paint brush, and the brush that the janitor is using is a cleaning brush. Then, using the method for computing the common theme that we discussed earlier, we infer that the common theme for the *Parallel* relation is *humans used a paint or cleaning* brush. As we argued in §3.2, this violates Grice's Maxim of Quantity, at least in this discourse context, making it a weak theme. Grice's Maxim of Quantity stipulates that one should not use a word which has an overspecific or underspecific denotation for whatever it is the speaker is trying to communicate. For example, in most contexts, *Give me my thing* cannot be used to mean *Give me my pen*. A detailed formal account of when a theme produces good quality coherence and when it doesn't would involve complex reasoning involving speech acts and the communicative intentions of the participants, because for a theme to be of good quality it must be possible to infer why the speaker is talking about this theme. This intentional reasoning is beyond the scope of this paper, and so we rely here on intuitive predictions about when a theme is bad, and when it violates Quantity.

What's overspecific or underspecific obviously depends on context. Here, the general brush purpose is replaced with the purpose of painting or cleaning. Consequently, hair brushing and all the other possible purposes that are more specific than brush-purpose have been ruled out in pragmatics. Contexts where the speaker intends to convey *paint brush or cleaning brush* and nothing more are rare. We cannot compute why painting or cleaning is the topic of the discourse, to the exclusion of all else in this context. It should be stressed, however, that this violation of Quantity is context sensitive: referring to *paint brush or cleaning brush* in the context below is fine, because we can work out why the theme involves painting or cleaning brushes from the discourse context.

(30) Several artists, two hairdressers and the janitor have been visited by a representative from the ACME wonder brush company. One artist and our janitor now use ACME brushes.

Since adding β' to the KB produces a weak discourse, clause (b) of the Interpretation Constraint is verified with this β'. Indeed, the whole antecedent of the Interpretation Constraint is verified.[9] This conflicts with Rembrandt and Specialisation taken together, since together these latter rules nonmonotonically yield β'—that the brush Rembrandt uses is a paint brush and the one the janitor uses is a cleaning brush, but the consequent of the Interpretation Constraint is that this doesn't hold. The result is irresolvable conflict between Rembrandt and Specialisation, via the conditions imposed by the Interpretation Constraint. One cannot decide which implicature to drop: whether it should be the one that Rembrandt is using a paint brush, or the janitor using a cleaning brush, or both. So although (19c) is not *semantically* ambiguous in that there is only one logical form, it is *pragmatically* ambiguous. This pragmatic ambiguity cannot be resolved and this produces a word pun effect.

Consider (30) again. As long as discourse attachment is stipulated as intuitions would dictate, the same β' (that the artist uses a paint brush and the janitor a cleaning brush) *doesn't* produce a weak discourse in (30) because the common theme doesn't violate Quantity in this particular context. Therefore, unlike (19c), the Interpretation Constraint *isn't* verified for this same β' with respect

[9] If α involves a paint brush but β involves a general brush, or if α involves a general brush and β a paint brush, then the common theme would be *humans used a (brush-purpose) brush*, and this wouldn't violate Quantity in the way that *paint-or-clean brush* does (because the latter is neither specific enough, nor general enough, whereas the former is general enough). So the β' as stipulated is the only combination of candidate interpretations of α and β that yield weak coherence.

to the KB for interpreting (30), and so the Specialisation rules fire via DMP, rather than being in irresolvable conflict with each other via the Interpretation Constraint. So (30) is correctly predicted by this theory to be a discourse context where the zeugma is ameliorated.

As in (19c), there are real world knowledge clues in (31) that the entities involved probably have different properties (one teacher is male and the other female).

(31) In 1950 the typist and the managing director both married teachers.

However, our theory predicts that, in spite of this common property of (19c) and (31), (31) isn't zeugmatic. As in (19c), the relevant constituents must be attached with *Parallel* in (31). Suppose that the managing director marries a female teacher and the typist a male one. Then generalisation yields the common theme *humans married a (general—i.e., male or female) teacher*. This isn't a weak theme, because Quantity isn't violated. So in contrast to (19c), interpreting α and β in the way real world knowledge clues would predict doesn't produce weak coherence. So the Interpretation Constraint doesn't block the pragmatic implicatures. Consequently, DMP on the relevant laws yields the inferences that the managing director married a female teacher, and the typist a male one.

Sometimes, there is no scope for re-interpreting a constituent in order to improve discourse coherence, because the truth conditional content of the constituents yields weak coherence, rather than the weakness hinging on pragmatic interpretations alone. Consider (21b).

(21) b. That thesis is orange and unreadable.

As with coordinated NPs, we assume that a VP which contains some coordinated constituent (in this case, a coordinated AP), is coherent only if one can find a rhetorical connection that links the relevant constituents together; in this case *that thesis is orange* and *that (same) thesis is unreadable.*

The interpreter's knowledge about syntax, compositional semantics and pragmatics entails that the only candidate is the relation *Parallel*. So one must check that its coherence constraints are satisfied. In other words, the constituents must be structurally isomorphic, and there must be a common theme. We will give more details of the structural isomorphism condition in the next section — it is met here straightforwardly. The common theme must now be computed, by generalising on the two constituents *That thesis is orange* and *That thesis is unreadable.* The first constituent predicates the physical aspect of *thesis*, and the second one predicates its content. Assuming that generalising over the semantics for an individuated object and an abstract representation gives something which is basically empty of content, the generalisation of the two constituents is *Properties of the thesis.* So, the common theme constraint is also met. But this common theme is very general relative to the semantic content of the constituents being connected, and makes the rhetorical connection weak. Compare this theme with one that's more specific, such as *Physical properties of the thesis.* This would have been the common theme for (32), for example:

(32) That thesis is orange and has thousands of pages.

So although (21b) was found to be coherent, the rhetorical connection is weak. Moreover, this connection is weak when only the *monotonic* or compositional semantics of the constituents are assumed. So the Interpretation Constraint fails to allow us a revision of interpretation to improve coherence. This is because clause (b) in the antecedent of the Interpretation Constraint doesn't hold. For suppose β' is instantiated to the truth conditional content of α and β. Then it is indeed the case that β' yields weak discourse coherence, but even so $\sqsupset_{KB}(\beta', \mathit{weak}(\tau \cup \beta))$ doesn't hold. This is because $\sqsupset_{KB}(\beta', \mathit{weak}(\tau \cup \beta))$ means $\mathit{weak}(\tau \cup \beta)$ is nonmonotonically inferred from the KB augmented with β', but *not* from the KB alone. Here, since the KB entails β', $\mathit{weak}(\tau \cup \beta)$ is inferrable from the KB alone. Since the Interpretation Constraint doesn't apply, and since there is no scope for shifting the interpretation of the constituents so that coherence is improved, the sentence sounds odd.

Now compare (21b) with (21a).

(21) a. That thesis has thousands of pages and is unreadable.

As before, (21a) is coherent only if one can compute a rhetorical connection between the VP constituents *has thousands of pages* and *is unreadable*. Just as in (21b), the same linguistic clues allow the interpreter to compute a *Parallel* relation between them, with the same mapping and the same common theme. But, unlike (21b), that's not all. Real world knowledge also allows the interpreter to infer a causal link between the constituents. This can encoded formally in DICE. Let $Info(\alpha, \beta)$ be a gloss for α is *x has thousands of pages* and β is *x is unreadable*. Then Unreadable states that if one tries to connect such constituents together in some way, then normally there is a causal relation between them.

- `Unreadable:`
  $\langle \tau, \alpha, \beta \rangle \wedge Info(\alpha, \beta) > cause(e_\alpha, e_\beta)$

This, together with Result, allow one to infer $Result(\alpha, \beta)$ in CE. So there is a stronger rhetorical connection between the constituents in (21a) than there was in (21b): as well as the weak *Parallel* connection, the interpreter's background knowledge has permitted the computation of a *Result* connection. So our theory correctly predicts that (21a) is better than (21b).

5.3 Discourse Coherence and Incremental Processing

It is well established that interpretation of language proceeds incrementally (e.g., Marslen-Wilson and Welsh 1978, Frazier 1979, Crain and Steedman 1985). Here, we investigate how this affects lexical processing in a discourse context. Consider (19b).

(19) b. Rembrandt used a brush and so did our janitor.

Now, suppose we assume that we calculate the semantic content *and* pragmatic implicature of the text as soon as enough information is available to allow an initial decision to be made. In (19b), this means that the interpreter calculates the pragmatic implicatures of *Rembrandt used a brush*, before parsing *and so did our janitor*.[10]

So, using a monotonic compositional semantic construction procedure such as that in Frank and Reyle (1995), the DRS α for the first clause *Rembrandt used a brush*, and α is not only evaluated against the model at this stage, but also the pragmatic implicatures of this semantic content are calculated in DICE as well.

$$(\alpha) \quad \begin{array}{|l|}\hline x, y, e, t \\ \hline rembrandt(x) \\ brush(y) \\ use(e, x, y, t) \\ holds(e, t) \\ t \prec now \\ \hline \end{array}$$

So, using the default rule Rembrandt, *brush* is specialised to *paint brush*. Only then does the interpreter begin to parse *and so did our janitor*. Again, one uses the constraint-based grammar to construct the DRS β for this clause:

[10]It is not our intention to advance fine-grained hypotheses concerning the temporal organisation of human language processing, since these can only be distinguished by psycholinguistic experimentation. Furthermore, the gross (and uncontroversial) assumption that pragmatic implicatures are calculated at least on a clause by clause basis is sufficient for our purposes.

$$(\beta) \qquad \begin{array}{|l|}\hline z, e', t' \\ \hline janitor(z) \\ do(e', z, c_1) \\ holds(e', t') \\ t' \prec now \\ c_1 \approx ? \\ \hline \end{array}$$

The condition $c_1 \approx ?$ marks the fact that (19b) contains an elided clause, which must be resolved via the discourse context. That is, the anaphor c_1 must be resolved to a property that z is doing. We must attach β to α with a discourse relation. This will then help resolve the 'property' anaphor and hence the elided clause. In this case, coordination constrains the applicable relations to *Parallel* or *Contrast* and the use of *and* selects *Parallel*.

We must check that *Parallel* is being used coherently: there must be a partial structural isomorphism and a common theme. Asher (1993a) defines structural isomorphism precisely, using the notion of an *extended embedding tree* (EET) . An EET amounts to the usual DRS structure, expanded out so that the VP information, which is represented as an abstracted DRS (the subject NP and event and time variables being abstracted out) is subordinate to the DRS as a whole. So, for example, *Rembrandt used a brush* is assigned the following two-level EET, where α is the DRS above, and P_1 is the abstracted DRS given below, which intuitively corresponds to the property of using a brush (in the past):

$$\begin{array}{c} \alpha \\ | \\ P_1 \end{array}$$

$$(P_1) \qquad \lambda e \lambda t \lambda x \begin{array}{|l|}\hline e, x, y, t \\ \hline brush(y) \\ use(e, x, y) \\ holds(e, t) \\ t \prec now \\ \hline \end{array}$$

Similarly, the EET for the second clause is given below, where β is the DRS corresponding to *so did our janitor*, and c_1 is the predicative DRS anaphor in β:

$$\begin{array}{c} \beta \\ | \\ c_1 \end{array}$$

Now, we determine the degree of structural isomorphism, by finding the mapping between the nodes of these EETs that is as close to a bijection as possible. The formal details of how to do this when an actual bijection isn't possible are to be found in Asher (1993a;286–291). These details can be skipped over here, however, because in this case there is a bijection. Here, β is mapped to α and c_1 to P_1. Because c_1 is mapped to P_1, the propositional anaphor is resolved to the property described by P_1; i.e., the property of use a brush. So, the event e' that the janitor did is use a brush (in the past). That is, the DRS in question with the propositional anaphor resolved is β_1:

$$(\beta_1)\qquad \begin{array}{|l|}\hline z, e', t' \\ \hline janitor(z) \\ do(e', z, c_1) \\ holds(e', t') \\ t' \prec now \\ c_1 \approx \lambda x'' \lambda e'' \lambda t'' \begin{array}{|l|}\hline e'', x'', y', t'' \\ \hline use(e'', x, y') \\ brush(y') \\ hold(e'', t'') \\ t'' \prec now \\ \hline \end{array} \\ \hline \end{array}$$

The model theoretic interpretation of β_1 as described in Asher (1993a;251) is equivalent to β_2—that is, one can reduce the abstract property the janitor does to the 'concrete' event that he uses a brush:

$$(\beta_2)\qquad \begin{array}{|l|}\hline z, y', e'', t'' \\ \hline janitor(z) \\ use(e'', z, y') \\ brush(y') \\ holds(e'', t'') \\ t'' \prec now \\ \hline \end{array}$$

In this way, the VP ellipsis is resolved as a byproduct of checking the coherence constraints on *Parallel*. This is compatible with the approach taken in Hobbs (1985), but the mechanisms described in Asher (1993a) go beyond those in Hobbs (1985), in that they handle examples where the clauses have complex structures (e.g., as those to be found in propositional attitude reports).

Now, according to our assumption about the process of incremental interpretation, the interpreter now calculates the pragmatic implicatures of what has been processed so far. Because pragmatic implicature is nonmonotonic in nature, it is important to retract old pragmatic inferences before calculating the new ones in the light of the new information. In this case, the specialisation of *brush* to *paint brush* that occurred earlier isn't considered to be fixed. So new information has the ability to force a reinterpretation of this word. This happens here because, just as in (19c), the default rules Rembrandt and Specialisation apply, and if they fire then the common theme that is demanded by the coherence constraints on *Parallel* violates Grice's Maxim of Quantity. So the Interpretation Constraint applies as well, where β' in that constraint is the information that Rembrandt uses a paint brush and the janitor a cleaning brush. So just as in (19c), there is irresolvable conflict between Rembrandt, Specialisation and the Interpretation Constraint. We cannot decide whether to assume Rembrandt doesn't use a paint brush, or the janitor doesn't use a cleaning brush, or both.

The interpreter has therefore been forced to reinterpret the word *brush*, in the light of processing the subsequent phrase *and so did our janitor*. Now, there is indecision about how to interpret it, where there wasn't indecision before the elided clause was parsed.

Calculating pragmatic implicature as one processes constituents, rather than doing it all at the end, is a risky but necessary strategy, because of limitations of memory, the utility of rapid comprehension, and so forth. Here, the risk has not paid off, and reinterpretation has been forced. Hence there is a word pun effect. In addition, because there are conflicting messages about how to specialise *brush* once all the relevant information has been processed in (19b), this contributes further to the word pun effect. The effect is more pronounced in (19b) than it was in (19c), exactly because reinterpretation as well as irresolvable conflict occurs.

Now consider an example involving word repetition.

(22) John banked the money and then he banked the plane.

As before, we calculate the pragmatic interpretation of *John banked the money*, before parsing the second constituent. Background knowledge allows one to infer in CE that *bank* in this phrase should be interpreted as $bank_{deposit}$ (that is, the verb with the financial sense). Now we parse the second constituent. We must attach it to the first. The similarity in syntactic structure and word forms used prompt one to infer a *Parallel* relation (as well as *Narration*). Both require coherence constraints to be met. We'll concentrate here on those for *Parallel*. We must find a common theme. The common theme for John $\text{bank}_{\text{deposit}}$ed the money and $\text{bank}_{\text{turn}}$ed the plane is *John doing things*, which is very general relative to the semantic content of the constituents being connected, and so yields weak discourse coherence. Checking the coherence constraints on *Narration* is similar, and so yields no improvement. Therefore, the Interpretation Constraint applies where β' stipulates that John $\text{bank}_{\text{deposit}}$ed the money and $\text{bank}_{\text{turn}}$ed the plane. But there is real world knowledge that yields these interpretations of the two occurrences of *banked* in (22). The Interpretation Constraint thus blocks both pieces of real world knowledge from firing simultaneously, much like it did in (19c). But the relevant rules have unrelated antecedents. So there is irresolvable conflict between the real world knowledge that money is $\text{bank}_{\text{deposit}}$ed rather than $\text{bank}_{\text{turn}}$ed and planes are $\text{bank}_{\text{turn}}$ed rather than $\text{bank}_{\text{deposit}}$ed, thanks to the Interpretation Constraint. The listener therefore fails to conclude what the interpretations of the word forms *bank* should be; she cannot work out which implicature from real world knowledge to drop. This produces a word pun effect. Again, two things contribute to it: reinterpretation of the word *bank* is forced, and there is ultimately irresolvable conflict about how to interpret it. As we mentioned in §2.3, MLEs alone don't predict the word pun in (22) because of the word repetition. We have formally encoded discourse interpretation principles and shown how they contribute to the word pun.

Exactly the same line of reasoning explains the word pun in (33):

(33) Rembrandt used a brush and our janitor used a brush.

So (33) and (19c), which are truth conditionally equivalent, are both zeugmatic but for slightly different reasons. (33) involved reinterpretation where (19c) did not, but both ultimately yield irresolvable pragmatic ambiguity via conflicts involving Rembrandt and Specialisation.

5.4 The Absence of Crossed Readings

In cases where there are MLEs, such as for *file* and *bank*, the account of non-crossed readings when there is a single occurrence of the word form in the sentence is the same as the traditional MLE approach:

(2) John and Bill each have a file.

(16) Texas and Alabama have preservation orders on their most beautiful banks.

However, by adding information flow between lexical organisation and discourse inference, we can also provide a complementary account of non-crossed readings in certain cases where there aren't MLEs, or where there is word repetition. Consider (19d).

(19) d. The grandparents gave the children brushes for Christmas.

Our theory predicts that this sentence doesn't have a reading where one child got a paint brush, and the other a hair brush. This is because we have assumed that there is a general lexical entry for *brush*, which refers to paint brushes, hair brushes, tooth brushes and so on. Because of this, the word *brush* fails to have a specialised meaning unless something in the context gives it one. In (19d), there is nothing in the context which causes specialisation. Since *brushes* is general we don't interpret the sentence specifically as one child getting a paint brush, and the other a floor brush, for example. Rather, the sentence is interpreted as vague or general about what kind of brushes the children received. Here, the absence of a crossed reading is epiphenomenal. And with further context, a crossed reading can be obtained:

(19) d. The grandparents gave the two children brushes for Christmas.
One got a hair brush and the other a paint brush.

Now consider an example where there is word repetition.

(23) John had a file and Bill had a file.

There are four possible logical forms of this sentence, corresponding to the four permutations of *file*$_{\text{tool}}$ and *file*$_{\text{dossier}}$ for the two occurrences of *file* in the sentence. The Interpretation Constraint rules out the crossed readings, however. We must attach the constituents *John had a file* and *Bill had a file* together with a rhetorical relation, to check the coherence of the coordination. The only candidate relation is *Parallel*. If the reading is crossed, then the common theme which the *Parallel* relation demands is weak, because it's too general (*humans have physical artefacts*). So, let β' be the information that the files Bill and John have are 'crossed'. Then the Interpretation Constraint applies with respect to the KB which contains the *disjunction* of attachment tasks corresponding to the four alternative logical forms. Thus DMP on the Interpretation Constraint yields $\neg\beta'$; i.e., the files aren't 'crossed'. Consequently, the interpreter non-monotonically infers that two of the four alternative logical forms is ruled out by the Interpretation Constraint. The remaining two logical forms amount to: either John and Bill both have files$_{\text{tool}}$, or they both have files$_{\text{dossier}}$. The interpreter cannot choose between these two alternatives, but he does at least infer that (23) doesn't have a crossed reading.

In contrast, (17b) can have a crossed reading, because there is only one logical form for (17b) since the predicate *teacher* is neutral about gender, and the Interpretation Constraint doesn't apply when β' is *The teachers John likes are male and those that Bill likes are female*, because the common theme is acceptable (*humans like their (general) teachers*).

(17) b. John and Bill both like their teachers.

6 Conclusion

We have argued that viewing lexical ambiguity as caused by the existence of discrete lexical units is too simplistic and that this view of the lexicon can tell only part of the story concerning word puns and non-crossed readings. Our complementary account formalises the pragmatic reasoning underlying the MLE approach to zeugma and non-crossed readings involving homonymous ambiguity, but also explains how zeugma and non-crossed readings occur in cases involving word repetition and constructional polysemy. In these cases, zeugma is caused by information flow between lexical organisation, grammatical structure, pragmatic reasoning, coherence constraints on discourse structure and incremental interpretation. Similarly, the absence of crossed readings was not due to lexical organisation alone, but the information flow between lexical organisation and discourse coherence constraints. This is in accord with the assumptions in Copestake and Briscoe (1995) that zeugma can arise in the absence of multiple lexical entries.

We studied all these interactions in a formal setting, by providing links between a lexicon with hierarchical organisation, SDRT and DICE and showed how the logic of CE which underlies DICE validates the inferences which predict zeugma and non-crossed readings in the right cases. This led to some general conclusions about linking the various components. First, discourse coherence constraints don't work independently from lexical ambiguity. The Interpretation Constraint records this dependency because it ensures that pragmatic inferences about what a word means can be blocked if they have an adverse effect on the overall coherence of the discourse. Word pun effects are caused by (a) irresolvable conflict in pragmatic interpretation, created by axioms that link lexical interpretation with the interpreter's background knowledge; (b) weak rhetorical connections between the constituents; and (c) reinterpreting incremental interpretation, which is forced by subsequent information in the discourse. In future work, we will show that this account can provide an adequate analysis of the effects attributed to degrees of discreteness in readings of words such as *brush* (compared with *teacher*) discussed in Cruse (1995), without making structural distinctions in the lexicon.

References

Altmann, G. and M. Steedman (1988) 'Interaction with context during sentence processing', *Cognition,* **30**.

Asher, N. (1993a) *Reference to Abstract Objects in Discourse,* Kluwer Academic Publishers.

Asher, N. (1993b) 'Extensions for Commonsense Entailment' in IJCAI Worskhop on Conditionals (ed.), *C. Boutilier and J. Delgrande,* Morgan Kaufmann.

Asher, N. and A. Lascarides (1994) 'Intentions and Information in Discourse', *Proceedings of the the 32nd Annual Meeting of the Association of Computational Linguistics (ACL94),* Las Cruces, New Mexico, pp. 34–41.

Asher, N. and A. Lascarides (1995a) 'Lexical Disambiguation in a Discourse Context', *Journal of Semantics,* **12**, 69–108.

Asher, N. and A. Lascarides (1995b) 'Metaphor in Discourse', *Proceedings of the AAAI Spring Symposium on The Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity and Generativity,* Stanford, CA, pp. 3–7.

Asher, N. and M. Morreau (1991) 'Commonsense entailment — a modal theory of non-monotonic reasoning', *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91),* Sydney, Australia, pp. 387–392.

Asher, N. and P. Sablayrolles (1995) 'A Typology and Discourse Semantics for Motion Verbs and Spatial PPs in French', *Journal of Semantics,* **12**, .

Briscoe, E.J., A. Copestake and A. Lascarides (1995) 'Blocking' in P. St. Dizier and E. Viegas (ed.), *Computational Lexical Semantics,* Cambridge University Press, pp. 273–302.

Copestake, A. (1992) 'The representation of lexical semantic information', Doctoral dissertation, University of Sussex, Cognitive Science Research Paper CSRP 280.

Copestake, A. and E.J. Briscoe (1995) 'Semi-Productive Polysemy and Sense Extension', *Journal of Semantics,* **12**, 15–67.

Crain, S. and M. J. S. Steedman (1985) 'On Not Being Led up the Garden Path: The Use of Context by the Psychological Parser' in D. Dowty, L. Karttunen and A. M. Zwicky (ed.), *Natural Language Parsing,* Cambridge University Press.

Cruse, D. A. (1986) *Lexical semantics,* Cambridge University Press, Cambridge, England.

Cruse, D. A. (1995) 'Polysemy and related phenomena from a cognitive linguistic viewpoint' in P. St. Dizier and E. Viegas (ed.), *Computational Lexical Semantics,* Cambridge University Press, pp. 33–49.

van Deemter, K. (1990) 'The ambiguous logic of Ambiguity', *Proceedings of the First CLIN Meeting,* Utrecht, The Netherlands, pp. 17–32.

Frank, A. and U. Reyle (1995) 'Principle Based Semantics for HPSG', *Proceedings of the Seventh Meeting of the European Chapter of the Association for Computational Linguistics (EACL95),* Dublin.

Frazier, L. (1979) *On Comprehending Sentences: Syntactic Parsing Strategies,* Indiana University Linguistics Club, Bloomington, Indiana.

Grice, H. P. (1975) 'Logic and Conversation' in P. Cole and J.L. Morgan (ed.), *Syntax and Semantics Volume 3: Speech Acts,* Academic Press, New York.

Grover, C., C. Brew, S. Manandhar and M. Moens (1994) 'Priority Union and Generalization in Discourse Grammars', *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL94),* Las Cruces, New Mexico, pp. 17–24.

Hobbs, J. R. (1985) 'On the Coherence and Structure of Discourse', Report No. CSLI-85-37, Center for the Study of Language and Information.

Kamp, H. and U. Reyle (1993) *From Discourse to Logic,* Kluwer Academic Publishers, Dordrecht, Germany.

Kehler, A. (1994) 'Common Topics and Coherent Situations: Interpreting Ellipsis in the Context of Discourse Inference', *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL94),* Las Cruces, New Mexico, pp. 50–57.

Kempson, R.M. (1977) *Semantic Theory,* Cambridge University Press.

Lascarides, A. (1995) 'The Pragmatics of Word Meaning', *Proceedings of the AAAI Spring Symposium on the Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity and Generativity,* Stanford, CA, pp. 75–81.

Lascarides, A. and N. Asher (1991) 'Discourse Relations and Defeasible Knowledge', *Proceedings of the the 29th Annual Meeting of the Association of Computational Linguistics,* Berkeley, Ca, pp. 55–63.

Lascarides, A. and N. Asher (1993) 'Temporal Interpretation, Discourse Relations and Commonsense Entailment', *Linguistics and Philosophy,* **16.5**, 437–493.

Lascarides, A. and A. Copestake (1995, forthcoming) 'The Pragmatics of Word Meaning' in T. Galloway and M. Simons (ed.), *SALT V Proceedings,* Cornell University, Ithaca, NY.

Lascarides, A. and J. Oberlander (1993) 'Temporal Coherence and Defeasible Knowledge', *Theoretical Linguistics,* **19.1**, 1–35.

Link, G. (1983) 'The logical analysis of plurals and mass terms: a lattice-theoretical approach' in Bäuerle, Schwarze and von Stechow (ed.), *Meaning, use and interpretation of language,* de Gruyter, Berlin, pp. 302–323.

Marslen-Wilson, W. and A. Welsh (1978) 'Parsing interactions and lexical access during word recognition in continuous speech', *Cognitive Psychology,* **10**, 29-63.

Polanyi, L. (1985) 'A Theory of Discourse Structure and Discourse Coherence' in Chicago (ed.), *Eilfort, W. H., P. D. Kroeber and K. L. Peterson,* , pp. Papers from the General Session at the Twenty-First Regional Meeting of the Chicago Linguistics Society.

Pollard, C. and I. Sag (1994) *Head Driven Phrase Structure Grammar,* CSLI Lecture Notes Series, Chicago University Press.

Pustejovsky, J. (1991) 'The generative lexicon', *Computational Linguistics,* **17(4)**, 409–441.

Pustejovsky, J. (1995) 'Linguistic constraints on type coercion' in P. St. Dizier and E. Viegas (ed.), *Computational Lexical Semantics,* Cambridge University Press.

Scha, R. and L. Polanyi (1988) 'An Augmented Context Free Grammar for Discourse', *Proceedings of the 12th International Conference on Computational Linguistics and the 24th Annual Meeting of the Association for Computational Linguistics,* Budapest, pp. 22–27.

Sperber, D. and D. Wilson (1986) *Relevance: Communication and Cognition,* Basil Blackwell, Oxford.

Thompson, S. and W. Mann (1987) 'Rhetorical Structure Theory: A Framework for the Analysis of Texts', *IPRA Papers in Pragmatics,* **1**, 79–105.

Zwicky, A. M. and J. M. Sadock (1975) 'Ambiguity tests and how to fail them' in J. P. Kimball (ed.), *Syntax and Semantics IV,* Academic Press, New York, pp. 1–36.