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,
Ventura Hall,
Stanford, CA 94305,
USA
aac@csli.stanford.edu

Ted Briscoe
Computer Laboratory
Cambridge University
Pembroke St.
Cambridge
CB2 3QG, UK
ejb@lcl.cam.ac.uk

November 14, 1996

Abstract

Several recent theories of linguistic representation treat the lexicon as a highly structured object, incorporating fairly detailed semantic information, 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, 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 entry be involved. Copestake and Briscoe (1995) claimed that these cases could be explained by discourse coherence, but did not describe how this might work. We remedy this here, by formalising a general pragmatic principle which encapsulates discourse effects on word meaning. We demonstrate how it contributes to the creation 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 interpreters prefer not to make inferences that have an adverse effect on the coherence of the discourse. This principle can affect how words in the discourse are interpreted, and one consequence is 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 known as zeugma, and also for the impossibility of crossed readings: (2) and (3) cannot have readings where John has a tool designed for shaping and Bill has a folder of documents.

(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 in many cases 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) a. The books on the top shelf are about syntax.

b. That thesis has thousands of pages and is quite unreadable.

Some recent theories of the lexicon attempt to account for this by treating the lexicon as a highly structured object, 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, without distinct lexical entries for the physical object/abstract representation usages of book, there is no formal explanation for the oddness of sentences such as (5) and the absence of crossed readings in (6), which cannot refer to a situation where Kim sold two works to a publisher and Sandy sold two paperbacks to a book dealer.

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

(6) Kim and Sandy both sold two books.

Copestake and Briscoe (1995) 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 on the traditional one.

We will first discuss zeugma and crossed readings and their use as a test for ambiguity, and briefly discuss the MLE account and illustrate its shortcomings. 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 explain how zeugma and non-crossed readings can occur in the absence of MLEs. 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 interpretation system.

2 Zeugma and Crossed Readings

For the purposes of this paper, we will assume that zeugma is a word punning effect or a figure of speech, in which two terms are inappropriately linked together. Sentences (7) to (10) are all examples of zeugma, and they indicate that it can occur with a wide variety of syntactic constructions.

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

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

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

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

In the MLE account, zeugma in all of these cases arises from the use of an ambiguous lexical item: *bank* in (7), *take* in (8) and (10), and *with* in (9). Zeugma arises from linking two mutually satisfiable assertions—Mr. Pickwick taking his hat and taking his leave are consistent, for example – which cannot be appropriately linked.

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 do not 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*. Thus on the assumption that lexical ambiguity implies MLEs, (2) and (3) indicate MLEs for *file*. On the other hand, *teacher* must be vague or general between male and female, since *teachers* in (11a,b) can refer to a mixture of male and female teachers.

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

The phenomena of zeugma and the absence of crossed readings are very closely related: intuitively zeugma occurs when incompatible crossed readings are (strongly) suggested by other information in the sentence.

2.1 The MLE Account of Zeugma and Non-crossed Readings

We will now go through the MLE explanation of zeugma in (12) and the lack of crossed readings in (13).

(12) Bank robbers and dam busters blow up banks.

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

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. Under standard assumptions about syntax and semantic processing, two logical forms are built for (12) and (13), arising from the lexical ambiguity of *bank*—one features the predicate $\text{bank}_{\text{ground}}$, and the other the predicate bank_{org}. So (13) cannot refer to a mixture of earth mounds and financial institutions, because neither logical form would be satisfied in this case.[1]

The explanation of the zeugma in (12) is more complex, but intuitively runs along the following lines. As with lexical ambiguity in general, the interpreter uses pragmatic information in an attempt to decide which logical form corresponds to the natural interpretation of (12); the one which features $\text{bank}_{\text{ground}}$ or the one which features bank_{org}. Knowledge about bank robbers favours interpreting

[1] Throughout this paper we will use MLE to describe the case where multiple lexical entries exist with distinct semantics. Some theories allow different lexical entries to have the same semantics but different syntactic properties, but this situation would not give rise to zeugma.

bank as bank_{org}, so this favours the latter logical form. But knowledge about dam busters favours the former logical form. Thus the pragmatic clues conflict, and the interpreter cannot decide which clue should take precedence in order to resolve this conflict. This produces a zeugmatic effect: the irresolvable conflict in pragmatics means the semantic ambiguity can't be resolved. The assumption of MLEs alone does not explain zeugma because it does not formalize this line of reasoning. 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 §3.2.

In contrast, the predicate for *teacher* is vague as to whether the teacher referred to is male or female. So there will be single logical forms for each of (11a) and (11b). Therefore (11a) will not be interpreted as zeugmatic and (11b) 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 is adequate for homonymous words such as *bank*, where senses are unrelated (synchronically). Even theories which assume a structured lexicon, such as that presented in Copestake and Briscoe (1995), assume MLEs in such 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. Unfortunately, zeugma and non-crossed readings often do not provide reliable ambiguity tests when applied to related senses; nor do MLEs provide satisfactory explanations.

2.2 Ambiguity Tests, Lexical Representation and Discourse Coherence

As is well known, ambiguity tests have flaws and are are difficult to apply consistently (see, for example, Catlin and Catlin, 1972). 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 (14):

(14) 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.

The word *brush* is being used quite generally in (14a), 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 test for MLEs, examples (14b,c) indicate no such entry for *brush*. So the results of the zeugma and crossed reading tests for ambiguity in (14b–c) conflict with the evidence of (14a).

It appears that the zeugma in (14b,c) must be explained with more complex machinery than the MLEs alone. It would be highly undesirable to postulate distinct senses for all the possible uses of brushes, and under standard assumptions about syntactic processing, (14a) must involve a general entry for *brush*. In a lexicon which is encoded as a multiple inheritance hierarchy, such as that of Copestake and Briscoe (1995), multiple aspects of meaning can be combined in a single entry. We refer to this as *constructional polysemy* (in contrast to *sense extension*, where two senses are related but have distinct lexical entries). Since lexical entries themselves are represented as feature structures in which some attributes can remain underspecified, a vague (i.e. underspecified) entry can be given for a word usage which subsumes more specific usages in which this attribute takes a specific value. The vague or general senses discussed above are treated in this way; for example, the vague sense of *brush* is represented by leaving the telic ('purpose') attribute (e.g. Pustejovsky, 1991) underspecified. Thus an indefinitely large number of usages can be represented in the lexicon.

But, under the assumption that a general entry would provide a predicate which could be used in the logical form for (14b,c), we have to find an alternative explanation for zeugma. Our explanation will be based on the way that pragmatics affects interpretation. By default, (15a) is interpreted as

Rembrandt used a paint brush, because of real world knowledge and assumptions about what the speaker would do if she wanted to communicate something different.

(15) a. Rembrandt used a brush.
b. The janitor used a brush.

Of course, there are many contexts where (15a) may be interpreted differently. The point here is that in the absence of information to the contrary, we jump to the conclusion that it means he used a paint brush. Similarly, (15b) is interpreted by default as *the janitor used a cleaning brush*. If we model these defaults, then we can explain why (14b) is zeugmatic: there are conflicting default messages about the pragmatic interpretation of *brush*, one coming from *Rembrandt* and the other from *janitor*. Note that this account of zeugma does not appeal to MLEs; it appeals to pragmatic knowledge resources, and their effect on how we interpret words. The general idea is that even though there is a single lexical entry for *brush*, and so sentences like (14b) are not semantically ambiguous, there is irresolvable conflict in the pragmatic component—namely, conflict about how to interpret *brush*—making (14b) pragmatically ambiguous.

Another slightly different set of examples involves 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 and to the abstract contents (as illustrated in (4)). This gives rise to zeugma in some sentences and not in others: (16a) and (16b) involve predication over both the physical object (*has thousands of pages* and *is orange*) and the content (*is unreadable*). And yet (16a) and (16c) are better than than (16b):

(16) 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, (16a) indicates that there is a lexical entry for *thesis* that is general between physical object and content, whereas (16b) indicates that there is not.

Under standard assumptions about coordination and constraint-based processing, it is impossible to account for (16a) if there are two different lexical entries for the different usages of *thesis*. Pustejovsky (e.g. 1991) and Copestake and Briscoe (1995) therefore assume that a single entry for *thesis* contains distinct attributes for form and constituency, which reflects 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 content. This is regarded as another subtype of constructional polysemy. One possible explanation of the contrast between (16a) and (16b) is that whether coordinating constituents is zeugmatic or not will depend not only on whether the lexical structure permits such co-predication—i.e., conjunction of predicates—on different aspects of *thesis*, but also on how the assertions rhetorically link together, to contribute to the coherence of discourse. Intuitively, the size of books is clearly relevant to their readability, and the colour of their covers is not. We are therefore drawing an analogy between the contrast in acceptability of (16a,b) and that in the discourses (17a,b):

(17) a. That thesis has thousands of pages. It's unreadable.
b. ?That thesis is orange. It's unreadable.

Just as in (17), a close rhetorical connection can be found in (16a) but not in (16b), because in (16a), the causal connection between a book's size and its readability is inferrable, and no such causal connection is salient in (16b). In contrast, (16c) does not involve co-predication, so there is no need to find a rhetorical connection between the book's colour and its readability.

One further argument against the MLE approach is that it fails to explain why zeugma and the absence of crossed readings tend to remain when the word form is repeated, as in (18) and (19) (see e.g. van Deemter 1990):

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

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

The zeugmatic effect in (18) is not as pronounced as in (12), but there is still a punning effect which the MLE account fails to predict, since the second use of *bank* could contribute a different predicate from the first. Furthermore, two of the candidate logical forms for (19) invoke crossed readings, but intuitively these seem to be excluded, or, at least, strongly dispreferred. Our explanations for these examples will be similar to those outlined above: constraints on *discourse coherence* are imposed by the way the phrases are rhetorically linked together. Interpretations which involve weak discourse coherence will be dispreferred: since the rhetorical connection will be weak under the crossed readings, the non-crossed readings will be preferred. If the default interpretations triggered by the content of the rest of the sentence are different for each use of the word, and there is no way of resolving the conflict, then zeugma will result. We will formalise this discourse approach in the rest of the paper.

3 Discourse Structure and Pragmatic Reasoning

Formalising the discourse approach to zeugma and the absence of crossed readings rests on providing general axioms that guide the information flow between lexical structure, pragmatic reasoning and constraints on discourse coherence. In this section, we review the representation of discourse and pragmatics that we use, and introduce a new discourse constraint on word meaning, which generalises the one for lexical sense disambiguation given in Asher and Lascarides (1995a). We discuss this rule in detail here because it will contribute to the explanation of zeugma.

3.1 Rhetorical Relations

We need a theory of discourse structure which imposes coherence constraints, in order to predict that (17a) is better than (17b). We will use Segmented Discourse Representation Theory (SDRT) (Asher, 1993a), which is an extension of DRT (Kamp and Reyle, 1993). A text is represented by a segmented DRS (SDRS), which is a recursive structure of DRSs with rhetorical relations linking SDRSs. The rhetorical relations are modeled after those proposed by Hobbs (1985), Polanyi (1985) and Thompson and Mann (1987). The different rhetorical relations impose different constraints on the semantic contents of the SDRSs they link together. We will mention five: *Narration* (e.g., (20a)), *Result* (e.g., (20b)), *Elaboration* (e.g., (20c)), *Contrast* (e.g., (20d)) and *Parallel* (e.g., (20e)).

(20) a. Max stood up. John greeted him.

b. John pushed Max. Max fell.

c. Max painted a picture. He used acrylics and oils.

d. Max has black hair, but Mary has brown hair.

e. Max has black hair, and Mary has black hair too.

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* does not require a distinct common topic, but it does require a causal relation between the events described in the constituents. *Elaboration* entails that the event which is elaborated is the topic of the text segment. These constraints are specified as axioms of a logic, but the formal details are not central to the account of zeugma. So here we will rely on intuitions about when the constraints are satisfied or not, and concentrate on what happens to word meaning when the constraints are not met.

Contrast and *Parallel* both require a partial structural isomorphism between the SDRSs. In addition, *Contrast* requires a contrasting theme, and *Parallel* a common theme. Asher (1993a) defines precisely how to calculate whether these constraints are met. We will explicate the process in §4, when we use it in our formal analysis of zeugma.

3.2 Pragmatic Reasoning

We need to represent default rules which encapsulate pragmatic interpretations of sentences such as (15a). We also need a theory which computes the rhetorical connections between meanings of segments of text (i.e. SDRSs) via the interpreter's background knowledge. For example, we want to be able to predict that the conjuncts in (17a) are connected together with *Result* and *Parallel*, whereas in (17b) the relation is *Parallel* alone. For then we can draw on the analogy discussed earlier, between computing rhetorical connections between the sentences of (17a,b), and checking that the coordinated constituents in (16a,b) link together in a coherent way.

The discussion of zeugma in §2.2 casts light on the kind of logic for pragmatics that we need. It must calculate the interaction between various knowledge resources such as lexical knowledge, real world knowledge and Gricean style pragmatic maxims, and in particular, predict when conflict among these different knowledge resources is resolvable, and when it is not. The pragmatic theory DICE (Discourse in Commonsense Entailment) (Lascarides and Asher 1991, 1993), which accompanies SDRT, is specified in a logic that does this.

DICE represents as default rules the interpreter's background knowledge about how syntax, compositional semantics, Gricean style pragmatic maxims and real world knowledge affect the interpretation of discourse. In particular, DICE is used to compute the pragmatic interpretation of constituents and the rhetorical relations between them in the 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 $A > B$ roughly means "If A, then normally B". CE exploits nonmonotonic validity ($\mid\!\approx$), which has the following attractive properties:

1. $\mid\!\approx$ 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.

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

3. $\mid\!\approx$ validates Skepticality: in cases of conflicts among defaults where there is no default that has a most specific antecedent relative to the others, $\mid\!\approx$ 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.

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

5. For each deduction $A \mid\!\approx B$ 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 $A|\!\approx\! B$. 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) and Asher (1993b).

We have examined in detail elsewhere how to represent different pieces of background knowledge as formulas of CE in DICE (Lascarides and Asher 1991, 1993; Lascarides and Oberlander, 1993; Asher and Lascarides, 1994, 1995a), so as to model the pragmatic interpretation of discourse. Here we discuss only those aspects of the representation that are crucial to zeugma. First, consider (15a).

(15) a. Rembrandt used a brush.
b. The janitor used a brush.

Since Rembrandt using a paint brush is the default interpretation of (15a) in the absence of further context, we can represent this information as a $>$-rule in CE, which will include some compositional semantic information in the antecedent—such as Rembrandt is the agent—and the information that the brush in question is a paint brush in the consequent. We will call this rule Specialisation$_{\text{p}}$, since it 'specialises' the meaning of (general) brush to paint brush. We assume this rule will also capture the implicatures that the brushes are paint brushes in (21):

(21) a. Rembrandt didn't use a brush.
b. If Rembrandt used a brush, he will regret it.

There will be a similar $>$-rule which captures the default interpretation of (15b)—call this Specialisation$_{\text{c}}$.[2] The detailed specification of these rules does not concern us here, since all that matters for our purposes is their logical relation to the other default rules for pragmatic interpretation. What is important is that Specialisation$_{\text{p}}$ and Specialisation$_{\text{c}}$ have unrelated antecedents, because one will refer to Rembrandt and the other to janitors. Consequently, if the rules apply and conflict—i.e., there is some reason why Rembrandt and the janitor cannot be using different types of brushes—then the conflict will be irresolvable because of Skepticality in CE.

DICE also computes the rhetorical relations that connect the DRSs together in SDRT. DRSs for clauses are built in a monotonic fashion via information from syntax, much as in standard DRT. However, in contrast to traditional DRT, one then reasons in DICE about how this DRS attaches to the preceding discourse structure via a rhetorical relation. And unlike the monotonic construction of the DRS representing the clause, the inferences in DICE underlying the choice of rhetorical relation are nonmonotonic, because rhetorical relations are not always linguistically marked but may be inferred on the basis of clues from background knowledge. A further contrast to traditional DRT is that once the semantic representation of the 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 for inferring the rhetorical relation. In standard DRT, semantic interpretation does not occur until the representation of the whole discourse is built.

The default rules Narration, Result and Elaboration in DICE are used for computing rhetorical relations.

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

[2] Indeed, Specialisation$_{\text{p}}$ and Specialisation$_{\text{c}}$ could be instances of a rule schema—one with Rembrandt substituted in the antecedent of the schema, and the other with the janitor.

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

For example, Result states: 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*. In general, axioms in DICE that are for 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. The way one infers $\mathit{Subtype}(\alpha, \beta)$ (glossed as the event condition in α is a subtype of that in β), or $\mathit{cause}(e_\alpha, e_\beta)$ in DICE is explained in detail in Asher and Lascarides (1995a).

3.3 A Discourse Constraint on Word Interpretation

To explicate how rhetorical structure affects word meaning, we give a detailed analysis of a case which involves the rhetorical relation *Elaboration*. The rules in Asher and Lascarides (1995a) ensure that $\mathit{Subtype}$(22a, 22b) holds, because learning to use a paint brush is a subtype of learning to paint, and the janitor is plausibly a member of school staff.

(22) 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 and Elaboration predicts that $\mathit{Elaboration}$(22a, 22b) holds. But inferring this crucially uses the information that the brush in (22b) is a paint brush, since learning to use a (general) brush is *not* a subtype of learning to paint. So what about the discourse (23), where there is no explicit stipulation about the brush, and indeed, where knowledge encoded in Specialisation$_c$ favours cleaning as opposed to paint brushes?

(23) 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 we infer that the brush used is a paint brush because of the discourse context. We explain here how this discourse effect on word meaning is encoded formally in DICE, since this discourse effect can also create zeugma.

Asher and Lascarides (1995a) provided rules in DICE which encoded lexical sense disambiguation in a discourse context. They argued in favour of a rule they called Lexical Impotence. This rule ensured that one avoids disambiguating a word according to the clues from within a sentence if it ultimately leads to discourse incoherence. But Lexical Impotence cannot explain the interaction between words and discourse in (23), because if the brush is a cleaning brush (as the clues in the sentence (23b) favour), then a narrative relation can be inferred. Because there is no incoherence, in that a rhetorical connection has been successfully computed, Lexical Impotence will not apply when interpreting (23).

But intuitively, although (23) is coherent when the brush is a cleaning brush, the coherence would be much better if *brush* is interpreted as referring to a paint brush. We assume that, in general, interpretations of words that give discourses which are only weakly coherent are avoided. This is a generalisation of the principle encapsulated in Lexical Impotence, since it will apply to incoherent discourses as the extreme case.

So, we want to formalise the principle that we do not make inferences that have an adverse effect on the quality of coherence, even when there are some clues from with the sentences that we should. We can capture this interpretation strategy as a rule in DICE. First, we assume that $\mathit{weak}(\tau)$

means that the SDRS τ is at best only weakly coherent. We will discuss how we might begin to formalise the truth conditions for $weak(\tau)$ below, in §3.4. Let $Info(\alpha)$ and $Info(\beta)$ be a gloss for all monotonic information about the constituents α and β.[3] 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 rhetorical 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) $\langle \tau, \alpha, \beta \rangle \wedge Info(\alpha) \wedge Info(\beta) \wedge$
  (b) $\sqsupset_{KB}(\beta', weak(\tau \cup \beta)))$
  (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 (i.e. $Info'(\beta) > \beta'$ is a rule in the KB, where $Info'(\beta)$ is some subpart of $Info(\beta)$). 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 $Info'(\beta) > \beta'$, and it is also more specific than this rule. So although β' would follow via DMP from a KB that contained just β, β' does not follow in the context of τ, because the Interpretation Constraint blocks it via Specificity.

The Interpretation Constraint models the way in which the discourse (23) affects the interpretation of *brush*. It leads to an inference that the brush in question is a paint brush, thereby overriding the default specified by Specialisation$_c$. We will go through this example in detail. Let the DRSs representing the compositional semantic content of (23a,b) be α and β respectively. Then let β' be the condition that the brush is a cleaning brush. Specialisation$_c$ can be glossed as $Info'(\beta) > \beta'$, and this rule applies. But the condition (b) of the Interpretation Constraint is verified with this β'. Interpreting the brush as a cleaning brush yields a nonmonotonic inference via DMP in DICE that the rhetorical relation is *Narration*, just as in (24a,b):

(24) a. All the school staff learned to paint.
b. (And) the janitor learned to use a cleaning brush.

Narration has coherence constraints: there must be a distinct common topic, which we will assume in this case is *school staff learn to do something*. This is the topic computed by the formal procedure of generalisation described by Grover *et al.* (1994), which is adequate for the examples in this paper, although it does not solve the problem of computation of topics in general. But intuitively, although the narrative is found to be coherent, it is only weakly so in this context, since the topic seems strange, as we will discuss in §3.4. If β' yields a weak discourse, $\sqsupset_{KB}(\beta', weak(\tau \cup \beta))$ holds. So the Interpretation Constraint applies, and its consequent is $\neg\beta'$. This conflicts with Specialisation$_c$, but the Interpretation Constraint is more specific, so $\neg\beta'$ is inferred. Thus we infer that the brush is not a cleaning brush.

The antecedent of the Interpretation Constraint applies for each β' where the brush is a tooth brush, a hair brush and so on, but does not apply for the β' where the brush is a paint brush. When this β' is added to the KB, then $Elaboration(\alpha, \beta)$ is inferred, just as in (22), and the resulting discourse interpretation is perfectly acceptable. So we correctly predict that (23) is an elaboration, and the brush in question is a paint brush.

[3] For example, if α is (15a), $Info(\alpha)$ will contain the information that Rembrandt used a general brush, because this is monotonically inferrable from the truth conditional content of α. However, $Info(\alpha)$ will not contain the information that Rembrandt used a paint brush, since this is only nonmonotonically inferrable.

Thus the Interpretation Constraint extends the Lexical Impotence Rule of Asher and Lascarides (1995a) to allow disambiguation to prefer a particular interpretation if other possibilities lead to a weak, but not necessarily incoherent, discourse. In §4 we will show that the Interpretation Constraint also plays a crucial part in creating zeugma in the absence of multiple lexical entries in the lexicon.

3.4 Weak coherence

In the discussion above, we relied on intuition to motivate the assumption that a discourse is weak. Of course, providing formal truth conditions for $weak(\tau)$ would involve complex reasoning about the semantic content of τ and the cognitive states of the participants in the discourse (cf. Grice, 1975), and this is well beyond the scope of this paper. But we can provide a preliminary sketch for the examples we discuss, where it seems that the discourse is weak because the topic is strange. In the example above, for instance, the topic computed when *brush* was interpreted as *cleaning brush* was *school staff learning to do something*. The rule below is intended to capture the intuition that the discourse will be weak if no explanation for this being the topic of conversation can be nonmonotonically deduced from the KB.

- `Weak Discourse from Strange Topics:`
  $(\sqsupset_{KB}(\tau, \gamma \Downarrow \tau) \wedge \neg\sqsupset_{KB}(\tau \wedge \gamma \Downarrow \tau, \delta \wedge Explanation(\gamma \Downarrow \tau, \delta))) \rightarrow weak(\tau)$

This rule states that if augmenting the interpreter's KB with the SDRS τ leads to the nonmonotonic conclusion that γ is the topic of τ (this is the gloss for the formula $\gamma \Downarrow \tau$), and it's not possible to compute from this information and the rest of the interpreter's KB any explanation δ for why this is the topic, then τ is weak.

To formalise the computation of the topic of the discourse from the KB—i.e., choosing the γ such that $\sqsupset_{KB}(\tau, \gamma \Downarrow \tau)$ is true—we can assume the method described by Grover *et al.* (1994). This uses the subsumption relations between entities that are recorded in the KB to compute the generalisation of the propositions being connected, and this generalisation serves as the topic. Thus we will assume that $\sqsupset_{KB}(\tau, \gamma \Downarrow \tau)$ holds just in case γ is the generalisation of the propositions featured in the SDRS τ. Note that the strangeness of the topic is not absolute, there may be discourses and contexts for which the topic *school staff learning to do something* would be acceptable. But here we assume that the hearer has no special knowledge about the situation, and that thus there is no apparent explanation for the speaker initially describing the entire school staff painting, and then weakening the topic by stating that the janitor learned an unrelated activity.

The above rule uses the semantics of the rhetorical relation *Explanation* to define weak discourse. Explanation is defined in Asher (1993a), where he formalises the proposal of Bromberger (1962) that an explanation β of a constituent α serves an answer to the question *why* α*?* Of course, Weak Discourse from Strange Topics doesn't fully specify the truth conditional content of $weak(\tau)$, since we have said nothing about the content of KB which prevents the interpreter from inferring a reason δ for why the topic of the discourse τ is γ. But this is a general (and very hard) problem for any formal account of discourse interpretation, and further discussion takes us away from our main concern in this paper, which is to exploit the logical relationship between the $>$-rules that encode what to do if τ is weak and the other rules in DICE, rather than providing detailed specifications of exactly when τ is a weak discourse. In essence, we show in this paper that modelling zeumga reduces to this well known open problem, of specifying the quality of discourse coherence.

4 Integrating the Components

In the previous section we described a rule that encodes discourse effects on word meaning. We now use it in this section to show how discourse effects create zeugma and the absence of crossed

readings, even when there are no MLEs. But first, we start with an example that is zeugmatic due to homonymous ambiguity.

4.1 Homonyms

Consider sentence (12) and its logical forms once more.

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

For the sake of brevity, we ignore the collective readings and concentrate on the distributive ones. Two logical forms thus arise because of the two lexical entries for *bank*: $\mathrm{LF}_{\mathrm{ground}}$ features the predicate $\mathrm{bank}_{\mathrm{ground}}$, and $\mathrm{LF}_{\mathrm{org}}$ features the predicate $\mathrm{bank}_{\mathrm{org}}$. To check the coherence of this sentence, we must resolve this semantic ambiguity.

To do this, the interpreter assumes that one of the logical forms is true, but not both—i.e., $\mathrm{LF}_{\mathrm{ground}} \vee \mathrm{LF}_{\mathrm{org}} \wedge \neg(\mathrm{LF}_{\mathrm{ground}} \wedge \mathrm{LF}_{\mathrm{org}})$ is part of her KB—and she attempts to draw a nonmonotonic inference via the $>$-rules in DICE about which logical form is the preferred reading. 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, then in the absence of further information we infer 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 can be represented in DICE as $>$-rules similar to the Specialisation rules in §3.3. They 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.

How do these rules affect the pragmatic interpretation of (12)? We are assuming that (12) is distributive, so the embedding functions that verify the DRSs $\mathrm{LF}_{\mathrm{ground}}$ and $\mathrm{LF}_{\mathrm{org}}$ 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: $\mathrm{LF}_{\mathrm{ground}} \vee \mathrm{LF}_{\mathrm{org}} \wedge \neg(\mathrm{LF}_{\mathrm{ground}} \wedge \mathrm{LF}_{\mathrm{org}})$ entails that either all the banks involved are financial or all are earth banks. So even though the antecedents of the default rules could be verified by different agents, different banks and different events, the rules are nevertheless dependent. Because the logical forms do not permit bank robbers to blow up financial institutions while the dam busters blow up earth banks, the real world knowledge conflicts. Moreover, it is irresolvable conflict, because the antecedents of the rules are logically unrelated: one is about dam busters and the other about bank robbers. Consequently by Skepticality in CE, the interpreter comes to no conclusions about whether *bank* should be interpreted as $\mathit{bank}_{\mathrm{org}}$ or $\mathit{bank}_{\mathrm{ground}}$. That is, Skepticality produces a zeugmatic effect in (12), because the semantic ambiguity is irresolvable. The interpreter does not begin to check how the coordinated constituents rhetorically link together, because there was no resolution of the truth conditional content of the constituents.

4.2 Syntax and Pragmatics: A Single Lexical Entry

Now consider an example where there is a single lexical entry, but nevertheless there is zeugma:

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

Because there is a general sense of *brush* which subsumes information in the other senses, there is only one logical form for (14c), which features the predicate derived from the vague lexical entry for *brush*.

The interpreter checks the coherence of coordinating the constituents in (14c) together by checking that a rhetorical connection can be computed between relevant propositions, where these are derived from the syntagmatic representation of the coordinated sentence. The two propositions are formed

by joining each property specified in the coordinate structure—in this case, the two generalised quantifiers *Rembrandt* ($\lambda Q \exists x(Rembrandt(x) \wedge Q(x))$) and *our janitor* ($\lambda Q \exists x(janitor(x) \wedge Q(x))$)—to the property specified by the rest of the sentence—in this case, the property $\lambda x(x$ *used a brush*) given by the VP. So the two propositions to be linked in the analysis of (14c) correspond to *Rembrandt used a brush* and *our janitor used a brush*. Computing the propositions to be connected in this way ensures we capture the analogy between the coherence of coordination and the coherence of discourse we alluded to earlier. Checking the coherence of (14c) is essentially the same as computing the rhetorical relation between the sentences in (25):

(25) Rembrandt used a brush. Our janitor used a brush.

Using the mechanisms available in SDRT, the interpreter infers in the monotonic component of CE that the relation in (14c) is *Parallel*. The task of checking the coherence of the *Parallel* relation is exactly that for checking the coherence of *Parallel* in (26) if the interpreter makes the assumption, β', that Rembrandt uses a paint brush and the janitor a cleaning brush:

(26) ?Rembrandt used a paint brush. Our janitor used a cleaning brush.

But, although it is logically consistent, (26) is at best a weak discourse, if not incoherent. We assume that this is because *Parallel* demands a common theme—which in this case could be glossed as *people doing something with a brush*. In a similar way to the example discussed in §3.4, this would be a bad theme, because there is nothing to explain why the speaker intends to talk about people using brushes for very different activities. Context could ameliorate the problem, as in (27), for example.

(27) 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.

But we assume that (14c) is not uttered in such a context.

Thus, assuming β' leads to $weak(\tau \cup \beta)$ being true and clause (b) of the Interpretation Constraint is verified. Indeed, the whole antecedent of the Interpretation Constraint is verified. This yields conflict between three default rules in CE: Specialisation_p, Specialisation_c and the Interpretation Constraint. They conflict because the former two rules yield β' (that Rembrandt uses a paint brush and the janitor a cleaning brush) and the Interpretation Constraint yields $\neg\beta'$. But the conflict between these rules is irresolvable in CE. Hence the interpreter 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 (14c) is not *semantically* ambiguous as there is only one logical form, it is *pragmatically* ambiguous. This pragmatic ambiguity cannot be resolved and this produces the zeugmatic effect. Ironically although the Interpretation Constraint, in general, improves discourse coherence, here it creates incoherence, because it causes Skepticality in CE in conjunction with other pragmatic knowledge resources. In contrast, in (27), the Interpretation Constraint would not apply: the common theme is acceptable in this discourse context. So the zeugma is ameliorated since there's no irresolvable conflict among the defaults.

As in (14c), there are real world knowledge clues in (28) that the entities involved probably have different properties (one teacher is male and the other female).

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

However, our theory correctly predicts that (28) is not zeugmatic. As in (14c), the relevant constituents must be attached with *Parallel* in (28). But even if the managing director marries a female teacher and the typist a male one, the theme (marrying general—i.e., male or female—teachers) is perfectly acceptable. So unlike (14c), interpreting α and β in the way real world knowledge clues

would predict does not lead to weak coherence. So the Interpretation Constraint doesn't block the pragmatic implicatures, and DMP on the relevant laws yields the inferences that the managing director married a female teacher, and the typist a male one.[4]

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. Consider (16b).

(16) b. ? That thesis is orange and unreadable.

In this case, using the above method for deriving the propositions to be rhetorically connected from the syntagmatic represenation of the coordinated sentence, the propositions are *that thesis is orange* and *that thesis is unreadable*. The only candidate relation between them is *Parallel*, and in order to satisfy its coherence constraints there must be a common theme. As before, we compute this 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. Since 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 *the thesis has (general) properties*. So the common theme constraint is met, but this common theme is very general relative to the semantic content of the constituents being connected, and thus the rhetorical connection is weak in (16b) and (17b).

In this example, the weak rhetorical connection arises when only the *monotonic* or compositional semantics of the constituents are assumed. That is, $weak(\tau \cup \beta)$ is inferrable from the KB alone. So $\exists_{KB}(\beta', weak(\tau \cup \beta))$ doesn't hold for any β' and clause (b) in the antecedent of the Interpretation Constraint doesn't hold. Thus there is no scope for shifting the interpretation of the constituents so that coherence is improved.

Now compare (16b) with (16a).

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

As before, (16a) is coherent only if one can compute a rhetorical connection between the relevant propositions, derived from the syntagmatic representation. Just as in (16b), the same linguistic clues allow the interpreter to compute a *Parallel* relation between them, with the same very general common theme. But real world knowledge also allows the interpreter to infer a causal link between the constituents using the default rules specified in Asher and Lascarides (1995a), and this in turn yields $Result(\alpha, \beta)$. So there is a stronger rhetorical connection between the constituents in (16a) than there was in (16b) and thus our theory correctly predicts that (16a) is more acceptable than (16b).

4.3 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 (14b).

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

We assume that the interpreter calculates 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 (14b), this

[4] We believe that we can provide an account of all the effects attributed by Cruse (1995) to different degrees of distinctness in readings of words such as *brush* contrasted to *teacher* along these lines, without having to make structural distinctions in the lexicon, but we will not discuss this further here.

means that the interpreter calculates the pragmatic implicatures of *Rembrandt used a brush*, before parsing *and so did our janitor.*[5]

So, using a monotonic compositional semantic construction procedure the interpreter builds the DRS α for the first clause *Rembrandt used a brush*, and α is not only evaluated against the model at this stage, but the pragmatic implicatures of this semantic content are calculated in DICE as well. So, *brush* is specialised to *paint brush*. Only then does the interpreter begin to parse *and so did our janitor*. Again, the grammar is used to construct the DRS β for this clause:

(β)

z, e', t'
$janitor(z)$
$do(e', z, c_1)$
$holds(e', t')$
$t' \prec now$
$c_1 \approx ?$

The condition $c_1 \approx ?$ marks the fact that (14b) contains an elided clause, which must be resolved via the discourse context. That is, the anaphor c_1 must be resolved, so that we know what the janitor z is doing. Attaching β to α with a discourse relation will help resolve it. In this case, coordination constrains the applicable relations to *Parallel* or *Contrast* and the use of *and* selects *Parallel*.

The constraints on *Parallel* require a partial structural isomorphism (as defined precisely in Asher (1993a)) and a common theme between the linked propositions. Because of the isomorphism constraint, the anaphor c_1 must be resolved to a property which corresponds to the VP information in the first clause, roughly speaking $\lambda x(x$ use a brush). So the VP ellipsis is resolved as a byproduct of checking the coherence constraints on *Parallel*, and the elided clause is expanded to *our janitor did something, which was use a brush* (this is equivalent to *our janitor used a brush* via axioms on the semantics of the auxillary *do*).

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 after the first clause is retracted because, just as in (14c), the default rules Specialisation$_p$ and Specialisation$_c$ apply, and if one assumes their consequents then the demands of the coherence constraints on *Parallel* yields a weak discourse. So the Interpretation Constraint will also apply, with β' being that Rembrandt uses a paint brush and the janitor a cleaning brush. So just as in (14c), there is irresolvable conflict between Specialisation$_p$, Specialisation$_c$ 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.

Calculating pragmatic implicature as one processes constituents, rather than doing everything at the end of the discourse, 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, leading to zeugma. In addition, the effect is more pronounced in (14b) than it was in (14c), exactly because reinterpretation as well as irresolvable conflict occurs.

Now consider an example involving word repetition.

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

[5] 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.

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 $bank_{\text{deposit}}$ (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 the interpreter to infer a *Parallel* relation (as well as *Narration*). But if we assume that the senses of the two occurrences of *banked* in (18) are governed by the real world knowledge about banking money and banking planes, then the task of computing a rhetorical relation for (18) is essentially identical to that for (29):

(29) ?John deposited the money. He turned the plane.

At best, (29) is coherent only in some weak sense. Therefore, the Interpretation Constraint applies for (18) when β' stipulates that John $\text{bank}_{\text{deposit}}$ed the money and $\text{bank}_{\text{turn}}$ed the plane. The Interpretation Constraint thus blocks both pieces of real world knowledge from applying simultaneously, as it did in (14c). This is because the relevant world knowledge rules have unrelated antecedents, so by Skepticality there is irresolvable conflict. Again, two things contribute to the zeugmatic effect: reinterpretation of the word *bank* is forced, and there is ultimately irresolvable conflict about how to interpret it.

Exactly the same line of reasoning explains the effect in (30):

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

So (30) and (14c), which are truth conditionally equivalent, are both zeugmatic but for slightly different reasons. (30) involved reinterpretation where (14c) did not, but both ultimately yield irresolvable pragmatic ambiguity.

4.4 The Absence of Crossed Readings

In cases where there are MLEs, such as for *file* and *bank*, the MLE account of non-crossed readings when there is a single occurrence of the word form still stands.

(2) John and Bill each have a file.

(13) 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 cases where there is only a single lexical entry, or where there is word repetition. Consider (14d).

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

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 forces one. In (14d), we do not 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 interpretation can be obtained:

(14) 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.

(19) 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 $file_{\text{dossier}}$. The only candidate rhetorical relation to attach *John had a file* and *Bill had a file* is *Parallel*. If the reading is crossed, then the connection is weak: cf. the weak *Parallel* relation in text (31):

(31) John had a dossier. (And) Bill had a tool.

So the Interpretation Constraint applies when a crossed reading is assumed, and thus blocks it via DMP. In this way, the interpreter nonmonotonically rules out two of the four logical forms but cannot choose between the remaining two 'non-crossed' logical forms.

5 Conclusion

We have argued that the view that all cases of zeugma and non-crossed readings are caused by the existence of discrete lexical entries is too simplistic and captures only part of the story. 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. Our treatment predicts that crossing readings of words with multiple lexical entries will always result in zeugma. Zeugma can also arise in cases of constructional polysemy but only when the context is such that the discourse would be weak or incoherent if more specific contrasting interpretations are assumed. Thus the test for distinct lexical entries is not the presence of zeugma alone, but the presence of zeugma even where the discourse would be coherent with contrasting interpretations. For instance, (10) is evidence of ambiguity in *take*, because of the coherence of the discourse even with the different interpretations illustrated by (10a):

(10) Mr. Pickwick took his hat and his leave.

(10a) Mr. Pickwick took his hat and took his leave.

If a word is repeated, crossed readings will not occur if this would result in a weak discourse. Furthermore, if there are conflicting clues in the sentence as to the correct reading, repeated words can produce zeugma. There are other rhetorical effects of word repetition which we have not accounted for, but we assume that these are independent of polysemy.

In all these cases, zeugma is caused by a mixture of lexical organisation, grammatical structure and pragmatics causing Skepticality (or irresolvable conflict) in CE. More specifically, the causes are (a) irresolvable conflict in pragmatic interpretation, created by axioms that link lexical interpretation with the interpreter's knowledge about how to interpret discourse; (b) weak rhetorical connections between the constituents; and (c) reinterpreting incremental interpretation, which is forced by subsequent information in the discourse. These effects 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 is one important common thread: pragmatic reasoning is central to zeugma, regardless of the presence of MLEs. Our account introduces only one piece of new machinery: the Interpretation Constraint which generalises the rule Lexical Impotence described by Asher and Lascarides (1995a). This not only allows us to account for many examples of zeugma but also extends the extent to which discourse coherence constraints can resolve lexical ambiguity.

References

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.

Bromberger, S. (1962) 'An Approach to Explanation' in Butler, R. (ed.), *Analytical Philosophy,* Oxford University Press, pp. 72–105.

Catlin, J-C. and J. Catlin (1972) 'Intentionality: a source of ambiguity in English?', *Linguistic Inquiry,* **3**.

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.

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.

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.

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 Eilfort, W. H., P. D. Kroeber and K. L. Peterson (ed.), *Papers from the General Session at the Twenty-First Regional Meeting of the Chicago Linguistics Society,* Chicago.

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.

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.