Lexical Disambiguation in a Discourse Context*

Nicholas Asher
IRIT, Université Paul Sabatier
118 Route de Narbonne
31062 Toulouse Cedex, France
also at: Centre for Cognitive Science *and*
Department of Philosophy
University of Texas, Austin
asher@irit.fr

Alex Lascarides†
Department of Linguistics
Building 100
Stanford University
Stanford, Ca 94305
USA
alex@csli.stanford.edu

October 10, 1997

Abstract

In this paper we investigate how discourse structure affects the meanings of words, and how the meanings of words affect discourse structure. We integrate three ingredients: a theory of discourse structure called SDRT, which represents discourse in terms of rhetorical relations that glue together the propositions introduced by the text segments; an accompanying theory of discourse attachment called DICE, which computes which rhetorical relations hold between the constituents, on the basis of the reader's background information; and a formal language for specifying the lexical knowledge—both syntactic and semantic—called the LKB. Through this integration, we can model the information flow from words to discourse, and discourse to words. From words to discourse, we show how the LKB permits the rules for computing rhetorical relations in DICE to be generalised and simplified, so that a single law applies to several semantically related lexical items. From discourse to words, we encode two novel heuristics for lexical disambiguation: disambiguate words so that discourse incoherence is avoided; and disambiguate words so that rhetorical connections are reinforced. These heuristics enable us to tackle several cases of lexical disambiguation, that have until now been outside the scope of theories of lexical processing.

1 Introduction

How is discourse information used to take lexical decisions, and lexical information used to take discourse decisions? In this paper, we observe data that illustrate the information flow between the semantics of words and the structure of discourse. From discourse to words, we illustrate how constraints on coherent discourse determine lexical sense disambiguation.

*We would like to thank Ted Briscoe, Ann Copestake, Claire Grover, Antonio Sanfilippo, Greg Whittemore, Yorick Wilks and two anonymous reviewers for helpful comments on the work reported here.
†This paper was written while supported by CNRS at IRIT, Université Paul Sabatier, Toulouse.

From words to discourse, we illustrate how the meanings of words affect discourse structure. We go on to explore how a formal theory of discourse interpretation can be augmented with a theory of lexical semantics, so that this information flow can be modelled.

2 The Puzzles

2.1 The Presentational Problem

Consider the disambiguation of *bar* in text (1):

(1) a. The judge asked where the defendant was.
b. The barrister apologised, and said he was at the pub across the street.
c. The court bailiff found him slumped underneath the bar.

A theory of lexical processing that takes only domain information into account can't handle this, because information about the structure of discourse is needed in order to resolve the lexical ambiguity.

To see this, first consider the Sentence Scenario, where we interpret (1c) in isolation of the discourse context and with the anaphor resolved: in other words, the sentence *The court bailiff found the defendant slumped underneath the bar*. In this context, *bar* would be interpreted as the courtroom bar. This is derived in part from the domain knowledge and the word associations of the courtroom bar with defendants and bailiffs.

Now consider the Discourse Scenario: the interpretation of (1c) in the discourse context provided by (1a,b). Now, *bar* refers to the drinking establishment, which is why the continuation (1d) is better than (1d′).

(1) d. He took him to get coffee before returning to the courtroom.
d′. ?He took him out of the courtroom to get coffee.

This disambiguation must take place because of information in the discourse. But what information?

Obviously, domain knowledge still plays an important part in the Discourse Scenario. In addition, we propose that the way this domain information is *presented* in the discourse also has a critical effect. For consider how (1) is interpreted. Via domain knowledge, one infers that the defendant is not in a courtroom when the events described in (1a,b) occur. Now we must interpret (1c) in this context. To maintain discourse coherence, we must calculate how the meaning of (1c) is connected to the meaning of the preceding discourse. One way in which one can do this is to work out the *rhetorical relation*—such as *Explanation*, *Background*, *Narration*, *Contrast* and *Evidence*—which connects the meanings of the segments of text (Hobbs 1985, Thompson and Mann 1987, Asher 1993a). The only candidate discourse relation for connecting (1c) to (1a,b) is *Narration* (Dahlgren 1988, Lascarides and Asher 1991), thus making (1) a narrative story with events described in their temporal order. An alternative interpretation would have required further information in this context; for example, an

Explanation or *Evidence* relation could have been marked by placing (1c) in the pluperfect, but they cannot be inferred on the basis of the information as it stands.

Let's first suppose that (1c) is indeed connected to (1a,b) with *Narration*. This imposes a spatial constraint. Simplistically put, the constraint is as follows: Unless there's explicit information in the constituents which, together with other background information, makes one believe that an actor in the text moves between the end of the eventuality in the first sentence and the start of the eventuality in the second, then the actor is stationary between these two time points. So, because the defendant isn't in the courtroom when the events in (1a,b) occur, he isn't in the courtroom at the start of the finding event either. But *find* is an achievement verb, without temporal extent. So the start and end of the finding refer to one and the same time and place. So this entails that the defendant isn't in the courtroom when he is found. But by the predicate argument structure of *find*, the defendant's at the *bar* when he is found. So the bar referred to cannot be in the courtroom. This precludes interpreting *bar* as having its courtroom sense.

Alternatively, let's suppose that we use the domain information to disambiguate *bar* that was used in the Sentence Scenario. That is, we make domain knowledge about defendants and bailiffs determine the interpretation of *bar* in (1c), and thus assign its courtroom sense. Then the defendant would be in the courtroom. This violates *Narration*'s spatial constraint, and would therefore preclude (1c) from forming a narrative with (1a,b). But linking the segments with *Narration* was the only way of maintaining discourse coherence. So using this domain knowledge here ultimately results in discourse incoherence.

The fact that we don't interpret (1) as incoherent leads one to conclude that we use the following strategy when disambiguating words: avoid disambiguation that results in discourse incoherence. So in the Discourse Scenario, we don't exploit the domain information that we used in the Sentence Scenario. Rather, we use the spatial constraints on narratives to eliminate the courtroom bar as a possibility, and so assign its pub sense. Note that we can even remove the word association between *pub* and the pub bar without affecting the result: replacing (1b) with (1b′) preserves the pub *bar* interpretation in (1c).

(1) b′. The barrister apologised, and said he was talking to his family across the street.

There are four things to observe here. First, the above explanation hinges on two things: the constraints on coherent discourse that rhetorical relations impose; and a general heuristic for avoiding disambiguation that leads to discourse incoherence. The constraints on coherent discourse make crucial use of the *particular* rhetorical relations featured. The spatial and temporal constraints for a narrative are different from those for a contrastive discourse for example. So changing (1) to a contrastive discourse yields different meanings for *bar*, as is illustrated in (2), where now (1d′) is a better continuation than (1d) is.

(2) The judge asked where the defendant was. His barrister apologised, and said he was at the pub across the street. But in fact, the court bailiff found him slumped underneath the bar.

Thus the disambiguation in (1) is ultimately driven by the rhetorical structure of the discourse. It illustrates a presentational problem: we must model how lexical disambiguation is affected by the way information is *presented* in the discourse context.

The second thing to note is that rhetorical relations between sentences are not always linguistically marked. Sometimes they are inferred pragmatically, using a wide variety of information: knowledge about syntax, semantic content, Gricean-style pragmatic maxims, the domain, and communicative intentions all play a part in inferring how segments of text should be connected together (Hobbs 1985). Therefore, to model the information flow between discourse processing and lexical processing, we must use a computationally tractable reasoning mechanism that models how rhetorical relations are pragmatically inferred.

The third thing to note is that the knowledge resources recruited in calculating the meaning of *bar* in (1) give conflicting messages, and the conflict is ultimately resolved. The two items of knowledge that apply and conflict are on the one hand the defeasible domain knowledge about bailiffs and defendants that is used to disambiguate *bar* in the Sentence Scenario, and on the other the defeasible knowledge that the discourse is narrative. The former knowledge entails that the defendant is in the courtroom when he is found, while via *Narration*'s spatial constraints, the latter entails that he isn't. The latter knowledge is favoured in disambiguation. We must explain why the conflict is resolved in this way.

The fourth thing to note is that the above examples go beyond what current theories on lexical processing have attempted. Some techniques have been developed for resolving the ambiguity of words in context (e.g., Wilks 1975, Boguraev 1979, Hirst 1987, Alshawi 1992). Several theories of word meaning have addressed how pragmatic factors, like world knowledge, affect disambiguation (Wilks 1975, Hayes 1977, Schank and Abelson 1977, Hobbs et al. 1990, Charniak 1983, Wilensky 1983, 1990, Wilks et al 1988, Guthrie et al., 1991, McRoy 1992). But this work hasn't attempted to tackle texts like (1) because one needs more than domain knowledge to explain them; one needs knowledge about rhetorical relations too.

These theories present various techniques for modelling how domain information determines lexical disambiguation. Charniak (1983), for example, uses marker passing techniques to establish word associations between items like *bailiff* and *bar*. More recently, Guthrie *et al.* (1991) use statistical techniques to tackle the same problem. They construct neighbourhoods for word senses, which contain the words that occur most frequently with that word sense in the definitions in a machine-readable dictionary. The idea is that *bailiff* is more likely to be in the neighbourhood of the courtroom bar than the drinking establishment bar. And one can use neighbourhoods to drive disambiguation: the word sense that has the largest intersection of its neighbourhood with the words in the text wins.

These techniques can sometimes predict the wrong results, for at least two main reasons. First, the frequencies of word co-occurrences are small, even in large corpora, so that the statistical models of disambiguation built from these frequencies can be unreliable. Second, statistical models of disambiguation using word association don't always handle conflicting word associations in the right manner. For example in (1), there are two conflicting word associations—one from *pub* and one from *court* etc.—which mark different senses of *bar*. In Guthrie *et al.*'s (1991) terms, *pub* is in the neighbourhood of the pub sense of *bar*, whereas *court* is in the neighbourhood of the courtroom sense of *bar*. Given the statistical flavour of these lexical techniques, there are three ways of resolving this conflict.

The first is to favour the word associations that occur more frequently in the corpus. But this won't work in general. Even if the association between *pub* and *bar* is stronger, thereby predicting the right interpretation of *bar* in (1), it would fail to predict that the meaning of *bar* is changed when (3) is inserted between (1b) and (1c).

(1) a. The judge asked where the defendant was.

b. The barrister apologised and said he was at the pub across the street.

(3) But suddenly, his whereabouts were revealed by a loud snore.

(1) c. The court bailiff found him slumped underneath at the bar.

The second technique is to assign a higher weight to associations between words that appear closer together in the text. This wrongly predicts that *bar* in (1) should have the courtroom sense, because *court* and *bailiff* are in closer proximity to *bar* in (1) than *pub*.

The third technique is the one Guthrie *et al.* (1991) adopt, which is to favour the word sense with the most associations in the text. This technique would also disambiguate *bar* in (1) to its courtroom sense, because there are more word associations for this sense (from *court*, *bailiff*, *judge*, *barrister* and *defendant*), than there are for the pub sense (from *pub*).

The problem with these techniques for resolving conflicting word associations is that they ignore the way the information is structured in the context. Obviously, modelling word associations along the lines suggested in the literature is an essential part of lexical processing. But in addition, we need to model how lexical processing is influenced by the rhetorical structure of the discourse.

So we come to the following conclusions about how to solve the Presentational Problem. First, current theories on the pragmatics of lexical processing need to be brought together with theories on discourse interpretation that calculate how rhetorical relations are inferred, and the constraints they impose on coherent discourse. Then, disambiguation strategies like Avoid Discourse Incoherence can be represented. Second, we require a precise and computationally feasible account of how conflict among the various knowledge resources recruited during lexical processing is resolved.

Now, many computational linguists model discourse by linking units of the discourse with rhetorical relations (e.g., Hobbs 1985, Hobbs et al. 1990, Lascarides and Asher 1991, 1993, Cohen 1987, Thompson and Mann 1987, Scha and Polanyi 1988, Hovy 1990, Moore and Paris 1991). However, with the exception of Hobbs et al. (1990) and Lascarides and Asher (1993), these theories don't attempt to resolve conflict among the knowledge resources during interpretation. They therefore don't supply the kind of inferential framework we need to handle conflicting knowledge resources during disambiguation.

TACITUS (Hobbs et al, 1990) tackles conflict resolution by assigning weights to predicates and guiding inference so that the conclusions inferred have the least weight. But there are no general principles behind the assignment of weights, making their account of conflict resolution unsatisfactory. Hobbs et al. (1990;p46) point out that extending the system by adding new knowledge resources requires extensive revisions to the *existing* representation, involving many hours of manual retuning of the weights on the target material. It is difficult to see what a solution to this retuning problem would be. Ideally, conflict resolution should be modelled by a sound logical consequence relation. For if this is achieved, then adding new knowledge resources to the theory is straightforward. The logic will predict how the new knowledge resource interacts with the existing ones, and the interactions among the existing knowledge resources will remain the same. So no retuning will be required. This is essentially the approach adopted by Lascarides and Asher (1991, 1993). They use a logic that can resolve conflict, and exploit this to model how the various pieces of background knowledge contribute to discourse interpretation. Because the logic predicts conflict resolution, they don't have to assign weights to predicates.

2.2 Strengthening Rhetorical Relations

Rhetorical relations also affect disambiguation in (4a,c) and (4b,c) and (4a,d), but in a different way to that illustrated in (1).

(4) a. The EC has been acting decisively lately.

b. The EC has been running meetings on time lately.

c. Last night's meeting came to a conclusion by 8pm.

d. But last nights' meeting came to a conclusion before any significant matters were discussed.

In (4a,c) the word *conclusion* corresponds to *agreement*; in (4b,c) it means *end*. It is plausible to assume that this lexical disambiguation is driven by the interpreter H's knowledge of what the author/speaker S is trying to do in the discourse; i.e. to present (4c) as *Evidence* for (4a) (or (4b)). H infers this rhetorical relation from the dispositional reading of (4a) (and (4b))—primed by the fact that they're generic—vs. the fact that (4c) refers to a particular event.

Having inferred that the rhetorical connection is *Evidence*, H can use this to disambiguate *conclusion*. In (4a,c), (4c) provides better evidential support for (4a) if *conclusion* is interpreted as *agreement* rather than *end*; for H knows that meetings can come to an end without any decisions being made. But in (4b,c), (4c) provides better evidential support if *conclusion* means *end*. Finally in contrast to (4a,c), *conclusion* means *end* rather than *agreement* in (4a,d), because S is trying to do something different in the discourse: he is contrasting two propositions, rather than providing a relation of evidential support between them.

Although both (1) and (4) use discourse information to drive lexical disambiguation, the ways in which they do so are different. In (1), we claim that only the pub sense of *bar* will preserve discourse coherence. A disambiguation strategy of Avoid Discourse Incoherence is being followed. In contrast, the lexical choice in (4a,c) is not driven by the need to avoid discourse incoherence, because an *Evidence* relation is supported, whatever the interpretation of *conclusion*. This is illustrated by the fact that (4a,e) and (4a,f) are both acceptable cases of *Evidence*.

(4) a. The EC has been acting decisively lately.

e. Last night's meeting came to an agreement by 8pm.

f. Last night's meeting came to an end by 8pm.

So the discourse is coherent regardless of how the word is disambiguated. Rather, having inferred which rhetorical relation holds between the sentences, the reader makes decisions about which sense of *conclusion* reinforces that rhetorical relation. So we have pinpointed a further disambiguation strategy involving rhetorical relations: disambiguate so as to reinforce rhetorical connections.

2.3 Lexical Information and Discourse Decisions

Integrating lexical processing and discourse processing is not only of benefit to lexical interpretation. The theory of discourse interpretation benefits too. Without a representation of lexical knowledge, a theory of discourse attachment is unaware of the semantic concepts that underly lexical entries. Consequently, it misses generalisations in interpretation, across discourses that use semantically similar lexical entries (e.g., Lascarides and Asher 1991). For example, in attaching the second sentence to the first in (5), one uses causal preferences about how *fall* and *push* should be connected together.

(5) Max fell. John pushed him.

Without a theory of lexical knowledge, the rule in the domain knowledge that represents this causal preference must use the actual predicates *fall* and *push*, so that this rule applies when interpreting (5). But this misses a generalisation, because we would then need separate laws of a similar nature for each text in (6), in spite of their semantic similarity.

(6) a. Max fell. John shoved him.
  b. Max tripped. John shoved him.
  c. Max stumbled. John bumped into him.

This makes it difficult to encode the rules for discourse attachment in a systematic way, because the rules must be specific to the lexical entries present. By introducing a theory of lexical knowledge of an appropriate kind, we could generalise the discourse attachment laws. As long as we encode in our lexical knowledge base that *push*, *shove* and *bump* are verbs that describe forces which can cause the movement of the patient, and *fall*, *trip* and *stumble* are verbs that describe movement of its subject, we can represent the information we need to do discourse attachment in (5) and (6) with just one law: when trying to attach two constituents together, where the former describes movement, and the latter describes the application of a force that can cause movement to that same individual, then normally, *Explanation* is the preferred discourse relation.

Theories of discourse attachment at present either provide an unsatisfactory account of how the knowledge resources interact during interpretation (e.g., Hobbs 1985, Hobbs et al. 1990); or they fail to integrate lexical and discourse processing so that the laws for discourse attachment capture intuitive generalisations (e.g. Lascarides and Asher 1991, 1993). We aim to solve these problems.

3 Starting Point

In order to place a theory of lexical disambiguation into a discourse context, we require three ingredients. First, we require a theory of discourse structure, which stipulates how rhetorical relations affect the structure of discourse, the constraints they impose on coherent discourse, and the semantic effects they have on the constituents they relate. Second, we require an accompanying theory of discourse attachment, which models how pragmatic knowledge resources are used to infer which rhetorical relations hold between two given discourse constituents. Finally, we need a formal language for representing lexical information; such as *push* is a verb that describes the application of a force that can result in movement of the patient, or the courtroom sense of *bar* is an object in a courtroom at which a barrister addresses the members of the court. These three ingredients must be mixed together in a unified account of NL interpretation, which makes precise the above accounts of how information flows between words and discourse.

As we've mentioned, there are several theories of discourse structure available that use rhetorical relations; e.g., Thompson and Mann's (1987) Rhetorical Structure Theory, Hobbs et al.'s (1990) TACITUS, Scha and Polanyi's (1988) Linguistic Discourse Model (LDM), and Asher's (1993a) Segmented Discourse Representation Theory (SDRT). Only LDM and SDRT provide a model-theoretic interpretation of the representations of discourse structure. And only SDRT accounts for the semantic effects that rhetorical relations have on the constituents being related. Consequently, SDRT is the only theory that can calculate the impact that the coherence constraints of the various rhetorical relations have, on the semantics of the constituents between which those rhetorical relations obtain. Furthermore, only TACITUS and SDRT come in tandem with a theory of discourse attachment, which computes which rhetorical relations underly the text, given the reader's background knowledge. We therefore will use SDRT as our discourse structure ingredient. And the theory of discourse attachment that accompanies SDRT is called Discourse In Commonsense Entailment (DICE) (Lascarides and Asher, 1991, 1993). This will be the second ingredient we use.

DICE utilises a logic called Commonsense Entailment (CE) (Asher and Morreau 1991), that is designed to handle reasoning with conflicting knowledge resources. As we've mentioned, it refines the tools used in TACITUS, in that it supplies a logical consequence relation for resolving conflict among the knowledge resources during discourse interpretation. DICE can explain how linguistic strings can be interpreted differently in different discourse contexts. But it is not equipped with a theory of lexical knowledge, and so as it stands it cannot model the reasoning that underlies the lexical disambiguations in (1) and (4).

For the third ingredient—a language for representing lexical information—several candidates exist. Arguably from our point of view, the most promising of these is Copestake and Briscoe's (1991) Lexical Knowledge Base (LKB). They use typed feature structures (FSs) to represent lexical information. The semantic type hierarchy, and the accompanying subsumption relation, represents information like **cheese** is a subtype of **food**, or a **trans-verb** is a subtype of **verb**. One reason why using FSs is advantageous from our perspective, is that through exploiting reentrancy, we gain a tight interface between syntax and semantics. This is an essential requirement if we are to explore interactions between the syntactic and semantic knowledge resources recruited during text processing. The subsumption relation on the type hierarchy will also prove useful at the discourse level, for it represents in a succinct way the syntactic and semantic proximity of words. This will enable us to generalise the laws for discourse interpretation.

We can build a general multi-modal model for the langauge of feature structures $\mathcal{L}_{fs}$ (Blackburn, 1992), and the language of CE $\mathcal{L}_{>}$ (where $>$ is the default conditional connective): viz. a model for $\mathcal{L}_{(fs,>)}$. Nevertheless—and this is important—our formulation countenances no interaction between the CE logic of $>$ (and *a fortiori* the nonmonotonic consequence relation $\approx\!\!\!|$) on the one hand, and Blackburn's modal operators on the other. So from the perspective of DICE, we can translate $\mathcal{L}_{(fs,>)}$ into $\mathcal{L}_{>}$ by treating each feature structure description and statement about subsumption relations, as atomic formulae of $\mathcal{L}_{>}$. The semantics of $>$, its logic, and the properties of the nonmonotonic consequence relation $\approx\!\!\!|$ are as discussed in Lascarides and Asher (1993), or more generally (for the predicate case), as in Asher and Morreau (1991).

We now describe SDRT, DICE and the LKB in more detail.

3.1 A Description of SDRT and DICE

SDRT is a semantically-based theory of discourse structure (Asher 1993a). This theory extends Kamp's (1981) Discourse Representation Theory (DRT) to represent the rhetorical relations that hold between the propositions introduced in a text. SDRT takes the basic building blocks of discourse structure to be propositions with a dynamic content, which are represented as DRSs—the representation scheme in DRT. A simple discourse structure consists of DRSs related by discourse relations—like *Narration*, *Background* and *Evidence*, among others. More generally, an NL text is represented by a recursive structure called a segmented DRS (or SDRS). An SDRS is a pair of sets containing respectively: the DRSs or SDRSs representing respectively sentences or text segments, and discourse relations between them. These structures are constructed in a dynamic, incremental fashion. The default assumption is that the sentence boundary marks the unit of information to be attached to the SDRS for the preceding discourse.

Discourse relations modelled after those proposed by Hobbs (1985) and Thompson and Mann (1987) link together the constituents of an SDRS. We will use seven discourse relations: *Narration*, *Elaboration*, *Explanation*, *Background*, *Evidence*, *Consequence* and *Contrast*. The first four of these constrain temporal structure: *Narration* entails that the descriptive order of events matches their temporal order; an *Explanation* or *Elaboration* entail they mismatch; and *Background* entails temporal overlap. *Narration* and *Elaboration* also constrain spatial structure (Asher, 1993b). Further details of the semantics of these relations, and NL examples, can be found in (Hobbs 1985, Thompson and Mann 1987, Asher 1993, Lascarides and Asher 1991).

The recursive nature of SDRSs give discourse structures a hierarchical configuration. The subordinating relations are *Elaboration* and *Explanation*, and the constituents to which new information can attach are a subset of the so-called *open constituents*, which are a subset of those constituents on the right frontier of the discourse structure (cf. Polanyi 1985, Grosz and Sidner 1986, Webber 1991), assuming that it is built in a depth first left to right manner.

SDRT specifies where in the preceding discourse structure the proposition introduced by the current sentence can attach with a discourse relation. DICE (Lascarides and Asher 1991, 1993, Lascarides and Oberlander 1992, 1993) is a formal theory of discourse attachment, which provides the means to *infer* from the reader's knowledge resources which discourse relation should be used to do attachment. Here, we assume the reader's knowledge base (KB) contains: the SDRS for the text so far; the logical form of the current sentence; an assumption that the current sentence must attach at an open site (i.e., the text is coherent); all defeasible and indefeasible world and pragmatic knowledge; and the laws of logic.

Lascarides and Asher (1991) argue that the rules introduced below are manifestations of Gricean-style pragmatic maxims and world knowledge; and as we've just mentioned, they form part of the reader's KB. A formal notation makes clear both the logical structure of these rules, and the problems involved in calculating rhetorical relations. Let $\langle \tau, \alpha, \beta \rangle$ be the update function, which means "the representation τ of the text so far, of which α is an open node, is to be updated with the representation β of the current sentence via a discourse relation with α". Let $\alpha \Downarrow \beta$ mean that α is a topic for β; let e_α be a term referring to the main eventuality described by α; and let $fall(e_\alpha, m)$ mean that this event is a Max falling. Let $e_1 \prec e_2$ mean the eventuality e_1 precedes e_2, and $cause(e_1, e_2)$ mean e_1 causes e_2. Finally, we represent the defeasible connective as a conditional $>$ (so $\phi > \psi$ means 'if ϕ, then normally ψ'). The maxims for modelling implicature are then represented as schemas:[1]

- `Narration`: $\langle \tau, \alpha, \beta \rangle > Narration(\alpha, \beta)$
- `Axiom on Narration`: $\Box(Narration(\alpha, \beta) \rightarrow e_\alpha \prec e_\beta)$
- `Background`: $(\langle \tau, \alpha, \beta \rangle \wedge state(e_\beta)) > Background(\alpha, \beta)$
- `Axiom on Background`: $\Box(Background(\alpha, \beta) \rightarrow overlap(e_\alpha, e_\beta))$
- `Push Explanation Law`: $(\langle \tau, \alpha, \beta \rangle \wedge fall(e_\alpha, m) \wedge push(e_\beta, j, m)) > Explanation(\alpha, \beta)$
- `Axiom on Explanation`: $\Box(Explanation(\alpha, \beta) \rightarrow cause(e_\beta, e_\alpha))$
- `Causes Precede Effects`: $\Box(cause(e_2, e_1) \rightarrow \neg e_1 \prec e_2)$

The rules for Narration and its Axiom convey the pragmatic effects of the textual order of events; by default, textual order mirrors temporal order. Background and its Axiom convey the pragmatic effects derived from aktionsart (states normally provide background information). The Push Explanation Law is a mixture of world knowledge (WK) and linguistic knowledge (LK). Given that β is to be attached to α with a discourse relation, the events they describe must be connected somehow. Given the kinds of events that they are, the reader normally concludes that the pushing caused the falling. Therefore, the pushing explained why Max fell. Finally, that Causes Precede their Effects is indefeasible world knowledge.

In fact, in some cases these rules are slightly modified versions of the rules in Lascarides and Asher (1991, 1993). We have made modifications to simplify the inferences underlying discourse attachment. The complexities we ignore have no bearing on the information flow between words and discourse, which is our main concern in this paper.

[1] Discourse structure and $\alpha \Downarrow \beta$ are given model theoretical interpretations in Asher (1993a); e_α abbreviates $me(\alpha)$, which is formally defined in Asher (1993a) in an intuitively correct way.

We also have laws relating the discourse structure to the topic structure (Lascarides and Asher, 1993). A Common Topic for Narrative states that any constituents related by *Narration* must have a distinct, common (and perhaps implicit) topic, and Topic for Elaboration states that the elaborated constituent is the topic:

- `A Common Topic for Narrative:`
  $\Box(Narration(\alpha,\beta) \rightarrow (\exists\gamma)(\gamma \Downarrow \alpha \wedge \gamma \Downarrow \beta) \wedge \neg(\alpha \Downarrow \beta) \wedge \neg(\beta \Downarrow \alpha))$
- `Topic for Elaboration:` $\Box(Elaboration(\alpha,\beta) \rightarrow \alpha \Downarrow \beta)$

The logic on which DICE rests is Asher and Morreau's (1991) Commonsense Entailment (CE). Motivation for choosing CE over other candidate nonmonotonic logics is discussed in detail in (Lascarides and Asher, 1993). Three patterns of nonmonotonic inference are particularly relevant. The first is Defeasible Modus Ponens: if one default rule has its antecedent verified, then the consequent is nonmonotonically inferred. The second is the Penguin Principle: if there are conflicting default rules that apply (conflicting in the sense that the consequents can't all hold in a consistent KB), and the antecedents to these default rules are in logical entailment relations, then the consequent of the rule with the most specific antecedent is nonmonotonically inferred. The third is the Nixon Diamond: if there are conflicting default rules that apply but no logical relations between the antecedents, then no conclusions are inferred.

Two further features of CE are essential to DICE. First, all the monotonic inferences from the premises are retrieved before embarking on the nonmonotonic inferences. The second feature is irrelevance: each one of these inference patterns also holds when information that is irrelevant to the predicates involved in the pattern is added to the premises (see Morreau (1992) for a precise definition; also Lascarides and Asher (1993; appendix) have shown the role of irrelevance in DICE).

Let us illustrate how DICE works by means of two simple examples. In interpreting (7), the KB contains α, β and $\langle \alpha, \alpha, \beta \rangle$, where α and β are respectively the logical forms of the first and second sentences.

(7) Max stood up. John greeted him.

(α)

e_1, t_1
$standup(e_1, m)$ $hold(e_1, t_1)$ $t_1 \prec now$

(β)

e_2, t_2
$greet(e_2, j, m)$ $hold(e_2, t_2)$ $t_2 \prec now$

The only rule that applies is Narration, and its consequent is inferred via Defeasible Modus Ponens. Hence by Narration's Axiom, the standing up precedes the greeting. By contrast in text (5), the KB verifies the antecedents to *two* conflicting defeasible laws: Narration and the Push Explanation Law ($Narration(\alpha, \beta)$ and $Explanation(\alpha, \beta)$ can't both hold given their Axioms).

(5) Max fell. John pushed him.

By the Penguin Principle, the Push Explanation Law wins, because its antecedent entails Narration's. Hence its consequent—*Explanation*—is inferred.

3.2 Typed Feature Structures for Lexical Information

As mentioned, we require a formal language for representing lexical information. We use typed FSs, similar to those described in Carpenter (1992). As shown in Copestake and Briscoe (1991), such a language allows a tight interface between syntax and semantics, and the lexicon and interpreter.

Features are represented by words in *italics* and types by words in **bold**. The type hierarchy is a partial ordering defined by the type subsumption relation $\sqsubseteq$. Subtypes inherit all the properties of all their supertypes. So, for example, if **salmon**$\sqsubseteq$**food** (meaning salmon is a type of food), then **salmon** has all the properties specified by the type constraints on **food**, plus perhaps some more properties of its own. We return to this shortly. The models for typed FSs correspond to directed acyclic graphs (DAGs). One can define the semantics of typed FSs by thinking of features as modal operators that label arcs between the nodes of a DAG, types as propositions holding at nodes, and constraints on types as conditionals (cf. Blackburn, 1992). As we have explained, since there is no interaction between Blackburn's models for FSs and those for $>$, we can think of Blackburn's WFFs for defining FSs as atomic WFFs of the language of DICE.

Our semantic framework is based on DRT, which defines the accessibility constraints on anaphora in discourse, as well as truth-conditional content. Because we exploit this property of DRT in SDRT and DICE, we augment the semantic component of Copestake and Briscoe's lexical entries, by replacing the predicate logic formulae with DRSs. These will yield the logical forms of sentences as DRSs, and consequently, we can maintain DRT's explanation of anaphora resolution in discourse.

We will assume, in line with Pustejovsky (1991), that the lexical entries that describe causes and their effects specify in which dimension this takes place. The dimensions of causation we use stem from Aristotle. There are four of them: Location, Form, Matter, and Intention. Locative (LOC) causes are efficient causes. They include movement, for example, such as pushing Max. Formal (FORM) causes create and destroy objects: for example, building a house, or eating a sandwich. Matter (MATTER) causes change the shape, size, matter or colour of an object: for example, treading on a cake, or painting a block red. Finally, Intentional (INT) causes change the propositional attitudes of individuals: for example, persuading John. It is, however, by no means clear that these should be universal categories of a KB, or the only categories. But they do for our purposes here.

Using notation similar to Sanfilippo (1992), we distinguish between those verbs that describe change, and those that cause it. So $loc(change(e, y))$ means that the event e describes a change in the individual y in the dimension LOC; e.g., $fall(e, y)$ will make $loc(change(e, y))$ true.[2] And $loc(\textit{cause-change-force}(e, x, y))$ means that x, through action e, causes a change in the LOC forces acting on y. For example, under one of its senses, $push(e, x, y)$ will make $loc(\textit{cause-change-force}(e, x, y))$ true.

Consider the LOC sense of *push*. This sense of pushing involves the application of a force through contact between the agent and patient. And depending on the strength of the force and the type of object, this force can result in the patient moving—that is, the patient can undergo a LOC change. In line with Sanfilippo (1992), we assume that this sense of *push* appears in the semantic type hierarchy at the meet of the types **force** and **contact**, and so would inherit features from both. The LOC sense of *push* is represented as follows:[3]

$$
\begin{bmatrix}
\textbf{causal-transverb} \\
orth : \textbf{push} \\
syn : \begin{bmatrix}
cat : \textbf{transverb-cat} \\
subj : \begin{bmatrix} syn : \begin{bmatrix} cat : \textbf{np} \end{bmatrix} \\ sem : \begin{bmatrix} agent : \boxed{1} \end{bmatrix} \end{bmatrix} \\
obj : \begin{bmatrix} syn : \begin{bmatrix} cat : \textbf{np} \end{bmatrix} \\ sem : \begin{bmatrix} patient : \boxed{2} \end{bmatrix} \end{bmatrix}
\end{bmatrix} \\
sem : \begin{bmatrix}
\textbf{contact-force} \\
qs : \begin{bmatrix} loc : \begin{bmatrix} \textit{cause-change-force}(\boxed{3}, \boxed{1}, \boxed{2}) \end{bmatrix} \end{bmatrix} \\
form : \begin{bmatrix} \begin{array}{|l|} \hline \boxed{1}\, x,\ \boxed{2}\, y,\ \boxed{3}\, e, t \\ \hline push(e, x, y) \\ hold(e, t) \\ \hline \end{array} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

This FS relates syntactic elements to semantic ones via reentrancy. Whatever fills the agent and patient positions in the syntax must also fill respectively the arguments in semantics for the individual who applies the force through contact, and the individual affected by this. In the semantic component, we see that *push* is classified as a contact-force verb; in other words

[2] We consider *loc* to be a modal operator, because it will correspond to a feature in the lexical entry for verbs, and using Blackburn's (1992) semantics, features correspond to modal operators.

[3] For the sake of simplicity, we don't represent the possible coercions of *push*. For example, the change in possession that occurs in *Max pushed Bill the book*.

it's a verb describing force through contact. *qs* stands for *qualia structure*, and the value of this feature supplies information about the dimension of causation (if any), the telic roles (i.e., the purpose) of objects, the form of objects, and their agentive roles (cf. Pustejovsky 1991, Copestake and Briscoe 1991, this volume). From the fact that *push* describes a LOC force through contact, we'll learn from the type hierarchy that the pushing action causes a change in the LOC forces acting on the patient y, thereby making *loc*(*cause-change-force*(e, x, y)) true, as we require. Finally, the value of the feature *form* is a DRS, and via the sentence grammar, this is used to build the DRS representing the logical form of any sentence containing the word *push*.

Feature structures can describe partial information, in the sense that more than one DAG may satisfy the feature structure. We can use this to represent semantic ambiguity. For example, *John pushed Max* can mean John applies a force to Max which, if large enough, will cause Max to move (*push*'s LOC sense), or that he encouraged Max to p, where p is some proposition (*push*'s INT sense).[4] The semantic ambiguity between the LOC and INT senses of *push* is represented by a feature structure that looks just like that given above, save that (8) is removed from the *qs*, and replaced with a FS that is a supertype of the FSs in (8) and (9):

(8) *loc* : *cause-change-force*($\boxed{3}$, $\boxed{1}$, $\boxed{2}$)

(9) *int* : *encourage*($\boxed{3}$, $\boxed{1}$, $\boxed{2}$, p)

We assume a hierarchy of types ensures that the resultant feature structure is well typed; in particular, the type of the semantic component of the FS will be the more general type **force**, rather than **contact-force**. Or to put it another way, some semantically ambiguous words form a type hierarchy, such that each sense is a subtype of the ambiguous word. Any DAG that satisfies the LOC sense of *push* will also satisfy the feature structure that represents the ambiguity between the LOC and INT senses of *push*.

When translating the NL verb *push* into logical form, we use the feature structure that represents the ambiguous sense until the word can be disambiguated by other information. Given the syntactic and semantic information in (10), only the LOC or INT senses of *push* will produce a well-formed feature structure representing (10).

(10) John pushed Max

We assume that the grammar will produce a feature structure where the logical form of (10) is the standard DRT representation (10′), that DICE used in Lascarides and Asher (1991).[5]

[4] There are other senses of *push*, but they're incompatible with either the sentence syntax of *John pushed Max*, or with the semantic type of the object *Max*. Syntax rules out things like the sense of *push* in *John pushed against the waves*. And the semantic type of the object, rules out the sense of *push* in *John pushed heroin*. These senses can be eliminated using techniques described in Boguraev (1979) and Alshawi (1992), and so they're not considered here.

[5] Re-entrancy and unification will play a role traditionally played by λ-abstraction in deriving this logical form, as described in Moore (1989).

e, x, y, t
$john(x)$ $max(y)$ $push(e, x, y)$ $hold(e, t)$ $t \prec now$

(10′)

A grammar such as HPSG (Pollard and Sag, 1994) or LFG (Kaplan and Bresnan, 1982) could be used here. We assume at this stage that the word *push* hasn't been disambiguated, because the intra-sentential information in (10) alone doesn't suffice to do this. Other information will be needed; for example the preceding discourse context.

However, the typed FSs represent more information than just the DRS (10′): For example, that *push* involves a LOC or INT force where the subject John is the agent and the object Max is the patient. This additional information can be exploited when reasoning about discourse attachment, since the typed FS is assumed to form part of the reader's KB.

How modular are the representations of lexical and discourse information? The rules for NL interpretation in TACITUS (Hobbs *et al.* 1990) are non-modular, in that a single rule can represent lexical information, world knowledge and linguistic knowledge as inextricably linked together. Consequently, there are no lexical entries in TACITUS; lexical information and more general background information aren't represented in separate modules. In DICE, the rules for NL interpretation are also represented so that the various knowledge resources seem to be inextricably linked. But in contrast to TACITUS, lexical knowledge is added to DICE via a separate mechanism: typed FSs. We will put information into a lexical entry only if it is useful for what have been thought of as lexical processes—such as coercion, metonymy (Pustejovsky 1991), and sense extensions (Copestake and Briscoe 1991, this volume). The logic will permit complex interactions between the lexical and discourse modules, and through this, integration among knowledges resources will be achieved.

4 From Words to Discourse

4.1 Lexical Semantics in Service to Explanations

We have extended DICE with typed FSs which represent lexical information. This extension can be used to generalise laws like the Push Explanation Law.

- `Push Explanation Law`: $\langle \tau, \alpha, \beta \rangle \wedge \mathit{fall}(e_\alpha, m) \wedge \mathit{push}(e_\beta, j, m) > \mathit{Explanation}(\alpha, \beta)$

This law was used to interpret (5). But it wouldn't apply when interpreting the texts in (6), in spite of the fact that they are so closely related in meaning.

(5) Max fell. John pushed him.

(6) a. Max fell. John shoved him.

b. Max tripped. John shoved him.

c. Max stumbled. John bumped into him.

Our classification of the dimensions of causation in the lexical entries have utility at the level of discourse attachment, and can overcome this limitation in DICE. First, we re-encode causation's effect on discourse attachment, so that a single, new Explanation law applies to the texts in (5) and (6):

- `Explanation:`
  $\langle \tau, \alpha, \beta \rangle \wedge \pi(\mathit{change}(e_\alpha, y)) \wedge \pi(\mathit{cause\text{-}change\text{-}force}(e_\beta, x, y)) > \mathit{Explanation}(\alpha, \beta)$

In words, the above law expresses a mixture of pragmatic knowledge and lexical knowledge about causation. It states that if β is to be attached to α with a discourse relation, and the event e_α (which is introduced in the lexical entry of the main verb in α) is a change in the individual y along the causal dimension π, and e_β is an action brought about by x which changes the forces along dimension π that act on y, then normally β explains α (and so by Explanation's Axiom, e_β caused e_α, thereby leading to the inference that the π force on y was sufficient to cause the π change in y described by e_α).[6]

Now suppose that we specify the lexical entries of *fall*, *stumble*, *trip*, *push*, *shove* and *bump* so that the Explanation Law applies when attaching the sentences in (5) and (6). This is guaranteed if these lexical entries have an *rqs* that looks like that of the LOC sense of *push* defined above. Then by the Penguin Principle, *Explanation* will be inferred in each of these texts: the conflicting laws being Explanation and Narration. By encoding lexical knowledge in such a way that the antecedent of the above Explanation Law is verified, we simplify the theory of discourse attachment. A single Explanation law for discourse attachment applies to all the above texts, rather than requiring a separate law for each lexical entry.

It is essential that this new version of the Explanation law specifies the causal dimensions of change. This is because we must ensure that a relation of causation is inferred only if the changes to the individual that are described in the sentences to be attached are compatible. We don't, for example, wish to infer a relation of causation between treading on a cake, which causes *deformation* of the cake, to the cake falling, which describes *movement*:

(11) The cake fell off the table. Max trod on it.

Explanation won't apply to (11) because the causal dimensions of the verbs mismatch. And so one will infer (11) is narrative. So using causal dimensions in the representation of lexical information is crucial if lexical processing is to influence discourse attachment in appropriate ways.

How do we ensure that lexical knowledge is specified in such a way that the antecedent to Explanation is verified, when doing discourse attachment in (5)? There is a further problem

[6] It would be useful to distinguish between those verbs that describe change and are amenable to external causation from those that aren't amenable to external causation (cf. Sanfilippo 1992). But we gloss over this for now.

that must be addressed if the above story is to work. As we've already mentioned, the word *push* is ambiguous in the sentence *John pushed Max* between its LOC and INT senses. But in the discourse context (5); *push* must be disambiguated to its LOC sense for the Explanation Law to apply. How does this disambiguation take place?

Intuitively, when the context fails to provide a proposition p that can be used to interpret *John pushed Max* as *John encouraged Max to p*, and when the context fails to tell you that the agent and patient can't be in physical contact with each other, then the LOC meaning of *push* is preferred. No such proposition appears in (5), and there's no inference that John and Max can't possibly be in contact with each other. So *push* is assigned the LOC sense. Thus Explanation applies. How are we to formalise this reasoning process?

In CE the nonmonotonic consequence relation is $\models\!\!\approx$. So if the context provides an appropriate proposition p such that John encouraged Max to p holds, then the following holds: $\langle\tau, \alpha, \beta\rangle, \tau, \beta \models\!\!\approx (\exists p) encourage(e, x, y, p)$.[7] Similarly, if the context provides information which leads to the inference that x and y can't possibly be in contact with each other, then the following holds: $\langle\tau, \alpha, \beta\rangle, \tau, \beta \models\!\!\approx \neg contact(x, y)$. CE is a logic that can represent directly in the object language the $\models\!\!\approx$ relation (Asher, 1993c). It is a nested conditional with the premises in the antecedent, and the conclusion in the consequent. We gloss over which formulae appear in between the antecedent and consequent in the nested conditional, since they are complex, and irrelevant to our purposes here. We simply refer to the nested conditionals that encode the above two $\models\!\!\approx$ relations as Π_{int} and Π_{dist} respectively. So, the law below captures the intuition that the LOC sense of *push* is chosen in (10), unless discourse context provides the appropriate proposition p such that John encouraged Max to p, or the discourse context leads to the inference that John and Max can't be in physical contact.

- `Locative Push:`
  $\Box(((\beta \rightarrow push(e, x, y)) \wedge \neg\Pi_{int} \wedge \neg\Pi_{dist}) \rightarrow loc(cause\text{-}change\text{-}force(e, x, y)))$

Note that we assume that the semantics of $loc(cause\text{-}change\text{-}force(e, x, y))$ is such that this rule doesn't entail that Max actually moves. Rather, it entails that the force applied to him is one that, if of sufficient strength, would cause him to move. We also assume that the metaphoric uses of *John pushed Max* block the rule from firing by making Π_{dist} true.

Now consider how the lexical preference stated in Locative Push will interact with the other laws in DICE, in the analysis of (5). The logical forms of the sentences in (5) are respectively α and β:

	x, e_1, t_1
(α)	$max(x)$ $fall(e_1, x)$ $hold(e_1, t_1)$ $t_1 \prec now$

[7] For the sake of simplicity, we gloss over the fact that encourage could be replaced by *persuade* or *force*.

(β)

y, e_2, t_2
$john(y)$ $push(e_2, y, x)$ $hold(e_2, t_2)$ $t_2 \prec now$

But α and β together don't provide a proposition p such that we can infer $encourage(e_2, x, y, p)$. There is no inference from the discourse context to $encourage(e_2, x, y, p)$, because the candidate propositions—$john(y)$, $hold(e_1, t_1)$, $fall(e_1, x)$—aren't of the appropriate kind; and none of the lexical entries in α or β feature the semantic type *encourage* (at least, not unambiguously). So $\neg\Pi_{int}$ holds. Moreover, α and β don't allow us to infer that John and Max can't be in contact with each other. So $\neg\Pi_{dist}$ also holds. Hence Locative Push's antecedent is verified. This is an indefeasible law, and in CE, monotonic reasoning takes place before nonmonotonic reasoning. Therefore it's consequent, that *push* is assigned its LOC sense, is inferred in DICE. Therefore, in the nonmonotonic component of DICE, the laws for discourse attachment that apply are Explanation and Narration. And by the Penguin Principle, $Explanation(\alpha, \beta)$ is inferred.

We have represented the preference for the LOC sense of *push* as an indefeasible law, with the exception to this preference explicitly represented in the antecedent: the exception being that there's a proposition p in the discourse context such that x encouraged y to p, or that x and y can't be in contact. This appears to go against the current trend in lexical processing, for using defeasible reasoning as an abbreviatory mechanism, allowing one to delete the exception statement (cf. Daelemans et al. 1992).

But such an abbreviatory strategy may be dangerous when applied here. For suppose that we stated the law as defeasible, and deleted the exception statement as shown below:

- `Defeasible Version of Locative Push:`
  $push(e, x, y) > loc(cause\text{-}change\text{-}force(e, x, y))$

Then we would fail to infer in the monotonic component of DICE that *push* is assigned the LOC sense. The laws that apply in the nonmonotonic component would then be Narration, and the defeasible version of the Locative Push Law. But there is conflict between them, because $Narration(\alpha, \beta)$ and $Explanation(\alpha, \beta)$ are inconsistent, and the latter is inferrable if the defeasible version of the Locative Push Law and Explanation fire in turn. But the antecedents of Narration and the Locative Push Law are unrelated. Consequently, the conflict is irresolvable, and no discourse relation between the sentences in (5) would be inferred, thus predicting—contrary to intuitions—that the text is incoherent.

If we assume that lexical and discourse information are modelled by the same logic of defaults, and that lexical and discourse processing can be interleaved, we obtain the above undesirable interactions. We could avoid this by either assuming that lexical information and discourse information are modelled by two different logics of defaults, or by assuming that lexical processing is always done before discourse processing, so that the conclusion of the Defeasible Version of Locative Push is inferred before discourse attachment is attempted. We don't want to make either of these assumptions, however, because this would preclude doing justice to the complex interactions between lexical and discourse processing. In particular, we would fail to formalise the resolution of conflict between the lexical and discourse information, that we illustrated in our informal analysis of text (1).

We were able to present the lexical preference for *push* in (10) via an indefeasible law, because the exceptions to the preference can be explicitly stated. This contrasts with defaults for discourse attachment, where exceptions to preferences cannot be exhaustively listed. If exceptions to lexical preferences can be listed, then default statements should be avoided, or else spurious knowledge conflicts will occur.

The above framework for specifying the semantics of lexical entries ensures a straightforward interaction with the rules for discourse attachment in DICE. Inferences about causation at the lexical level help us to infer the rhetorical relation *Explanation* at the textual level. The analyses of the texts in (6) will be similar to that of (5), assuming that the lexical entries are similar to *fall* and *push*. So through lexical processing, we have achieved a generalisation in DICE about how causation affects discourse attachment.

4.2 Lexical Semantics in Service to Elaborations

A similar problem to DICE's original treatment of explanations arises in elaborations. Consider the following example, taken from Lascarides and Asher (1991):

(12) a. Guy enjoyed a lovely meal.
 b. He ate salmon.
 c. He devoured lots of cheese.

Lascarides and Asher (1991) use the following rules to infer that (12b) and (12c) elaborate (12a): they capture the intuition that if Guy eating a meal and Guy eating salmon (or cheese) are to be connected, then a part/whole relation is preferred, and that this in turn normally yields *Elaboration*:

- `The Salmon Law:` $\langle \tau, \alpha, \beta \rangle \wedge eat(e_\alpha, g, meal) \wedge eat(e_\beta, g, salmon) > part(e_\beta, e_\alpha)$
- `The Cheese Law:` $\langle \tau, \alpha, \beta \rangle \wedge eat(e_\alpha, g, meal) \wedge eat(e_\beta, g, cheese) > part(e_\beta, e_\alpha)$
- `Elaboration:` $\langle \tau, \alpha, \beta \rangle \wedge part(e_\beta, e_\alpha) > Elaboration(\alpha, \beta)$

These laws allow one to infer that the text is an elaboration. But this analysis misses a generalisation: it is the fact that salmon and cheese are a *subtype* of food, which in turn constitutes the substance of the meal, that permits us to infer that (12b) and (12c) elaborate (12a). The fact that we needed two laws—the Cheese and Salmon Laws—to analyse the above misses this generalisation.

We have seen how specifying causal information in lexical semantics permitted generalisation at the textual level for explanations. Here, we see how subsumption relations on semantic types in lexical entries permit generalisation for elaborations.

(12a) is a sentence where logical metonymy takes place, as described in Briscoe et al. (1990). One must ensure that the object of *enjoy* is of the right type, namely, an event. Using techniques described in (Pustejovsky, 1991), we can assume that the meal in (12a) is coerced into the event *eating a meal*, because this is the telic role or purpose of a meal. Coercing *meal* to *eating the meal* is a default, and hence further discourse information could override this preference (cf. Copestake and Briscoe, this volume). We gloss over this here, however. So the relevant FSs for building the logical forms of (12a) and (12b) are respectively as follows:[8]

$$\begin{bmatrix} \textbf{event-noun} & \\ \textit{orth}: \textbf{meal} & \\ \textit{sem}: & \begin{bmatrix} \textit{eventstr}: & \begin{bmatrix} \textbf{event} \\ \textit{agent}: \boxed{1}\,\textbf{animate} \\ \textit{patient}: \boxed{2}\,\textbf{food} \\ \textit{process}: \boxed{3}\,\textbf{eat} \end{bmatrix} \\ \textit{form}: & \left[\begin{array}{|l|} \hline \boxed{1}\,x,\ \boxed{2}\,y,\ \boxed{3}\,e \\ \hline eat(e,x,y) \\ \hline \end{array}\right] \end{bmatrix} \end{bmatrix}$$

edible-fish-substance⊑food

$$\begin{bmatrix} \textbf{lexical-mass-noun} & \\ \textit{orth}: \textbf{salmon} & \\ \textit{syn}: & \begin{bmatrix} \textit{cat}: \textbf{noun} \\ \textit{count}: - \end{bmatrix} \\ \textit{sem}: & \begin{bmatrix} \textbf{edible-fish-substance} \\ \begin{array}{|l|} \hline x \\ \hline salmon(x) \\ \\ \hline \end{array} \end{bmatrix} \end{bmatrix}$$

In words, *meal* refers to an event of an **animate** individual eating **food**. *Salmon* is a mass noun and **edible-fish-substance** is a subtype of **food**.[9]

The above FSs encode semantic type information on the individuals involved. We can use this to generalise the laws at the discourse level that are used to infer *Elaboration*. To do

[8] For the sake of simplicity, we have glossed over the values of *qs* in these entries. In the case of *salmon*, some properties of its *qs* will be inferred via type inheritance from **food**.

[9] Some comments about the above lexical entry for *salmon* are in order. This is an example of sense extension, as described in Briscoe, Copestake and Lascarides (1993). We assume a lexical rule of *animal grinding* turns an animal into its food substance, where the orthography remains unchanged. In sentence (12b), this food sense of *salmon* is used because part of the meaning of *eat* is that the object is of type **food**, and so the preferred interpretation of *salmon* in (12b) is as food. Consequently, when the logical form for (12b) is fixed, we already know *salmon* is assigned its food sense, thus ensuring Subtype and Elaboration apply in the logic CE.

this, we first extend $\sqsubseteq$ to take DRSs γ_1 and γ_2 as arguments: $\gamma_1 \sqsubseteq \gamma_2$ if there is a discourse reference $x_1 \in U_{\gamma_1}$ such that the conditions on x_1 are a subtype of those on the discourse referents in U_{γ_2}, and γ_2 has no conditions on its discourse referents that are a subtype of a condition on a discourse referent in γ_1. So, for example, $\gamma_2 \sqsubseteq \gamma_1$ below, because *salmon* is an **edible-fish-substance**, which is a subtype of **food**.

$$(\gamma_1) \quad \begin{array}{|l|}\hline \text{x} \\ \hline \text{food(x)} \\ \hline \end{array}$$

$$(\gamma_2) \quad \begin{array}{|l|}\hline \text{x} \\ \hline \text{salmon(x)} \\ \hline \end{array}$$

We now state the relevant laws:

- `Subtype:`
  $\Box((\theta_i(\text{e-condn}_\alpha, \alpha, \gamma_1) \wedge \theta_i(\textit{e-condn}_\beta, \beta, \gamma_2) \wedge \textit{e-condn}_\beta) \sqsubseteq \textit{e-condn}_\alpha \wedge \gamma_2 \sqsubseteq \gamma_1) \rightarrow \textit{subtype}(\beta, \alpha))$
- `Elaboration:` $\langle \tau, \alpha, \beta \rangle \wedge \textit{subtype}(\beta, \alpha) > \textit{Elaboration}(\alpha, \beta)$

In words Subtype states the following: if (a) the DRSs γ_1 and γ_2 respectively identify the thematic role θ_i in α and β, with respect to the event conditions e-condn$_\alpha$ and e-condn$_\beta$ on particular events in α and β; (b) this condition on an event of β is a subtype of that of α (for example in this case, *devour* is a subtype of *eat*); and (c) γ_2 is a subtype of γ_1; then β is a subtype of α. Elaboration states that if β is to be attached to α, and β is a subtype of α, then normally *Elaboration*(α, β) holds.

The above lexical information and rules for discourse attachment have an impact on the analysis of (12a,b). We assume that the sentence grammar derives the logical forms α and β respectively for (12a) and (12b) from the above lexical entries, and in the FS for β, the patient *eat* is of type **edible-fish-substance** which, as we've stated, is a subtype of **food**:

$$(\alpha) \quad \begin{array}{|l|}\hline x, e_1, t_1, e, f \\ \hline max(x) \\ enjoy(e_1, x, e) \\ hold(e_1, t_1) \\ eat(e, x, f) \\ t_1 \prec now \\ \hline \end{array}$$

$$(\beta) \quad \begin{array}{|l|}\hline e_2, t_2, s \\ \hline eat(e_2, x, s) \\ salmon(s) \\ hold(e_2, t_2) \\ t_2 \prec now \\ \hline \end{array}$$

According to the above lexical entries for *meal* and *salmon*, the DRSs γ_1 and γ_2 above are in the same theta role *patient* to α's and β's eating events respectively. Furthermore, $\gamma_2 \sqsubseteq \gamma_1$ and **eat**$\sqsubseteq$**eat**, and therefore $e\text{-}condn_{\beta}) \sqsubseteq e\text{-}condn_{\alpha}$. So, the antecedent to Subtype is verified, and so in the monotonic component of CE, we infer $subtype(\beta, \alpha)$. In the nonmonotonic reasoning for discourse attachment, therefore, two laws apply: Elaboration and Narration. By the Penguin Principle, $Elaboration(\alpha, \beta)$ is inferred. The same laws for discourse attachment will apply when analysing (12c), because **cheese**$\sqsubseteq$**food**, and **devour**$\sqsubseteq$**eat**.

Through exploring how to spread reasoning in interpretation between the lexicon and pragmatics, we have learnt about the kind of lexical structures that we need. Lexical entries must be structured along causal dimensions, to ensure the right flow between the semantics of verbs at the lexical level and explanations at the textual level. Lexical entries must also encode semantic type subsumption relations, to ensure the right flow between thematic role information at the lexical level and elaborations at the textual level.

5 From Discourse to Words

We have seen how lexical processing can work in service to a theory of discourse attachment. Now we investigate the other side of the coin. How can the knowledge resources encoded in a theory of discourse attachment be used in lexical processing? In particular, how should we encode the affects of discourse context on lexical disambiguation in (1) and (4)?

(1)
  a. The judge demanded to know where the defendant was.
  b. The barrister apologised, and said that he was drinking across the street.
  c. The court bailiff found him slumped at the <u>bar</u>.

(4)
  a. The EC are decisive.
  b. The EC run meetings on time.
  c. Last night's meeting came to a <u>conclusion</u> by 8pm.

We suggested earlier that *bar* in (1) is disambiguated to its pub sense on the basis of constraints on coherent discourse. In contrast, *conclusion* in (4a,c) and (4b,c), are disambiguated on the basis of strengthening the rhetorical link between the sentences. We now show how these proposals can work in formal detail. We analyse each of the above texts in turn.

5.1 Lexical Information and Discourse Coherence

When disambiguating *bar* in text (1), we argued that the discourse information on how to disambiguate *bar*—which results from the coherence constraints on discourse—conflicts with the information on how to disambiguate *bar* which resides in the sentence (1c). The former favours the pub sense of *bar*, and the latter favours the courtroom sense, and the former ultimately wins. We must represent these discourse and sentential knowledge resources, and show how the conflict between them is resolved in the logic in the appropriate way.

So, let's first consider the sentential knowledge resources. The rational process that underlies the way the syntactic and semantic information in (1c), and the domain knowledge about courtroom scenarios, favour the courtroom sense of *bar* is not well understood. So we simply make the following crucial assumption about lexical disambiguation using word association. If words in the sentence favour by default a particular sense of an ambiguous word, we can encode this in DICE in the form of a >-rule, where the antecedent is the FS of the sentence that's constructed from the FS of the ambiguous word, and the consequent is the FS of the sentence that's now constructed from the FS of the sense of the word that's favoured. Some notation clarifies this assumption. Let FS_{bar} be the feature structure that represents the sentence (1c), which is constructed from the FS for the word *bar* before it's disambiguated. Let FS_{bar_1} be the same FS as FS_{bar}, save that the part of the FS that represents the ambiguous word *bar* is replaced with the FS that represents the courtroom sense of *bar*, which we label bar_1. Then our assumption is that these FSs are related by >, as shown in the Bar Rule.

- `Bar Rule:` $FS_{bar} > FS_{bar_1}$

This rule states the following: the information in the sentence normally leads one to infer that the meaning of *bar* is bar_1.[10]

Assuming that the default disambiguation of senses by sentential information can be expressed as a >-rule is a relatively weak assumption, but it poses a problem. The Bar Rule is specific to (1c). So how is the Bar Rule acquired from more general principles about disambiguation in a sentential context? It appears that we have replaced DICE's problem of requiring very specific laws like the Push Explanation Law concerning discourse attachment, with the problem of requiring very specific laws concerning sense preference in a sentential context. Ideally, we should be able to exploit general principles about word sense disambiguation in a sentential context. If these principles predicted correctly that the courtroom bar is favoured in the sentential context (1c), then we would be assuming merely that this preference is stated as a >-rule. We could, in essence, systematically generate rules like the Bar Rule from the general principles of lexical processing. But as we've mentioned, these principles aren't well understood in the lexical literature, although it has been appreciated for a long time that such principles are needed in a full treatment of NL interpretation. So for now, we simply assume that the preferences that an adequate theory of sense disambiguation in a sentential context would give, will be stated in DICE as a >-rule.

Now we turn to the discourse information that applies when interpreting text (1). First, we argued at the beginning of the paper that the spatial constraints on narrative discourse played an important role in the disambiguation of *bar*. What are the spatial constraints?

Asher (1993b) investigates the spatial constraints of narratives in DICE, and uses these constraints to reason about how objects move in narrative discourse. First he defines how to calculate when and where eventualities start (written *source*(e)) and when and where they stop (written *goal*(e)). *source*(e) is calculated by projecting e onto the space line and time line, and the temporal part of *source*(e) is the first point on the time line in the projection, and the spatial part of *source*(e) is the place where e occurs at this point on the time line.

[10] Unlike LOC push, we assume we really do need >-rules here, because the exceptions to inferring the consequent cannot be exhaustively listed. This is a reasonable assumption to make, since unlike (10), the disambiguation of (1c) is dependent on domain knowledge.

$goal(e)$ is similarly defined. The spatial constraints in narrative that Asher (1993b) proposes then constrain the values of $source(e)$ and $source(e)$.

The text (1), however, is more complex than the narratives considered in Asher (1993b), in that it contains the indirect speech report (1b), and this plays a significant role in reasoning about the defendant's whereabouts during the discourse. Asher (1993b) doesn't consider how indirect speech reports of this kind affect the inferences about space in narratives.

Now, in (1), the defendant is an actor in all three sentences, and in (1b), the barrister makes a claim about where he is. Upon interpreting (1b), therefore, we gain an expectation: the defendant is at the pub across the street when the barrister is speaking. This expectation is a manifestation of Grice's Maxim of Quality; we assume that the barrister is being sincere, and we expand our beliefs with his belief about the defendant's whereabouts. It must be stressed that this is an *expectation* though, and not added to the truth conditional content of the constituents in the discourse. For if it were added, then it would be impossible to refute the barristers speech report by subsequent information in the discourse, without reducing the discourse to inconsistency.

In narratives, expectations cannot be cancelled by subsequent information in the discourse: *Contrast* is the rhetorical relation that plays this function. So if (1c) is attached to (1b) with narration, then the expectation that the defendant is at the pub when the barrister spoke survives when interpreting what (1c) means. But what does this tell us about where the defendant was found?

There is a spatial constraint on narratives, that as long as there is no information in the compositional semantics of β that, together with the rest of the KB, leads to an inference that actors have moved between the end of the first eventuality and the start of the second, then they don't move between these points. One example of such a signal in the compositional semantics of β is a temporal adverbial like *twenty years later*, which indicates such a long lapse in time that actors are likely to have moved. Furthermore, given that expectations can't be cancelled in narratives, if the expectations include information about where an actor is, then this can fix the location of the actor in subsequent eventualities in the narrative. Let Π_{loc} be the embedded default that asserts the following relationship: the compositional semantics of β and the KB $\approx\!\!\!\!|$ the actors have moved between the end of α's eventuality and the start of β's. Then, these ideas are about locations in narrative, and in particular how expectations about locations are exploited, are captured in the following constraint:

- `Spatial Constraint on Narrative:`
  $\Box((Narration(\alpha,\beta) \wedge actor(x,\alpha) \wedge actor(x,\beta) \wedge Expt(\tau \cup \alpha, \gamma) \wedge \neg\Pi_{loc}) \rightarrow$
  $loc(x, source(e_\beta)) = loc(x, source(e_\alpha + e_\gamma)))$

This law states: if $Narration(\alpha,\beta)$ holds; α and β share an actor x; γ is an expectation from the discourse context; and β contains no information that leads you to believe things have moved between the goal of α and γ's eventualities combined, and the the source of β's; then the location of x at these spatio-temporal points are the same.

More needs to be said about the source and goals of complex eventualities, but the idea is that the source and goal of these complex entities is calculated by projecting the eventualities onto a time line and space line, in a similar manner to simple eventualities. The situation is complicated when we consider states: what's the endpoint of the state as far as the discourse context is concerned? This is a difficult issue, and requires extensive research in the interaction between aktionsart and discourse. So we gloss over this issue here, and will simply fix the values of $goal(e_\alpha + e_\gamma)$ as intuitions would dictate.

Now, the fact that the above spatial constraint allows e_γ to play a role in calculating where x is guarantees that if the expectation γ gives information about where x is, then this fixes where he is in the subsequent narrative. In this way, we exploit the expectation of where x is when interpreting the subsequent discourse. But it should be stressed that this is not possible with all rhetorical relations. Narratives can't cancel expectations, but contrastive discourses do.

We will see how the above spatial constraint plays a crucial role in our interpretation of (1c). One might think that domain knowledge would be adequate here to reason about the movement of objects. After all, many AI systems of knowledge representation represent a persistence axiom, that states that unless there's information to the contrary, objects remain stationary. But there are two reasons why encoding such a constraint is inadequate for our purposes. First, as we showed at the beginning of the paper, different rhetorical relations have different spatial constraints. An AI style 'spatial persistence' axiom fails to reflect this. Second, AI persistence axioms are default rules, and our spatial constraint on Narration is an indefeasible rule, which applies in quite specific circumstances. It will shortly become clear why it is important that the spatial constraint is an indefeasible law. If it weren't, we would get spurious irresolvable knowledge conflicts when reasoning about discourse attachment.

Now, in the analysis of (1) we aim to formalise the following line of reasoning. The discourse context wins over the sentential context when disambiguating *bar*, in order to avoid discourse incoherence. How do we encode this?

Discourse incoherence occurs when no rhetorical relation can be inferred for attaching the current constituent. The rule Narration is the only rule that *always* applies when attaching β to α, regardless of their content. So consider the case when Narration is the *only* rule that applies. Then if sentential information about disambiguation conflicts with Narration—such conflict is possible because $Narration(\alpha, \beta)$ constrains the semantic relation between α and β—then a Nixon Diamond will form, no discourse relation will be inferred and the discourse will be incoherent. So we wish to capture the intuition that these cases of irresolvable conflict are avoided, by ensuring that discourse information wins over sentential information.

Any Nixon Diamond between Narration and sentential information about disambiguation must be avoided, if Narration is the only rule for discourse attachment that applies. We avoid this Nixon Diamond, by transforming it into a Penguin Principle. More specifically, suppose sentential information favours a particular sense of a word *lex*, and this preference is encoded as a >-rule $FS_{lex} > FS_{lex_1}$. Suppose furthermore, that Narration is the only discourse attachment rule that applies. Then under these circumstances, we derive a new narrative rule which is 'strengthened', in that this sentential information FS_{lex} is added as a conjunct in the antecedent of Narration. So we have a new rule $(FS_{lex} \wedge \langle \tau, \alpha, \beta \rangle) > Narration(\alpha, \beta)$. Then, if there is conflict between this discourse rule and the sentential rule for disambiguation, $Narration(\alpha, \beta)$ will be inferred via the Penguin Principle, since the new rule's antecedent entails FS_{lex}, making it more specific. This ensures that discourse information wins, and discourse incoherence is avoided.

This general law about the interaction between words and discourse is formally represented in DICE as follows:

- `Lexical Impotence:`
  $\Box(((FS_{lex} > FS_{lex_1}) \wedge \neg((Info_{lex_1}(\alpha,\beta) \wedge \langle\tau,\alpha,\beta\rangle) > (R(\alpha,\beta) \wedge R \neq Narration))) \rightarrow$
  $(FS_{lex} \wedge \langle\tau,\alpha,\beta\rangle) > Narration(\alpha,\beta))$

In words, it states the following: suppose there is information in the sentence FS_{lex} that normally leads one to conclude the meaning of a particular lexical item *lex* is lex_1. Suppose furthermore that even if you interpret *lex* as lex_1, then you still don't add information to the two constituents α and β that help you infer a non-default discourse relation between them (i.e., a relation other than *Narration*). Then, indefeasibly, one can assume that updating α with β normally leads to *Narration*, whatever *lex* means. That is, the intra-sentential information doesn't change the defaults for discourse attachment.

In essence, if Lexical Impotence strikes, then discourse attachment has priority over the lexical disambiguation from sentential information. This prioritisation is ensured, because specificity will favour the strengthened narrative rule—which is inferred via Lexical Impotence—when it conflicts with the disambiguation rule $FS_{lex} > FS_{lex_1}$.

Lexical Impotence carries with it the heuristic we required of disambiguating words so as to avoid discourse incoherence. To see this, we'll show how it underlies the analysis of (1), and how it ensures that we choose to preserve discourse coherence, rather than disambiguating *bar* in favour of the intra-sentential information. Let the logical forms of (1a,b) be respectively α_1 and α_2. We assume that when interpreting α_1, WK is stated as intuitions would dictate, so that given the reader's KB, the reader has an expectation that the judge is in court, and the defendant isn't, and hence not in bar_1 (written respectively as $\neg overlaps_s(c,d,t_{\alpha_1})$ and $\neg overlaps_s(bar_1,d,t_{\alpha_1})$, where c is the court, d is the defendant, $overlaps_s$ is "spatially overlaps", and t_{α_1} is the time where e_{α_1} holds).[11]

Now consider attaching α_2 to α_1. The only law that applies is Narration, and its consequent—$Narration(\alpha_1,\alpha_2)$—is consistent with the rest of the KB. So by Defeasible Modus Ponens, it is inferred. And by logical omniscience and the Axiom on Narration, $e_{\alpha_1} \prec e_{\alpha_2}$.

As we've mentioned, we also gain an expectation from the indirect speech report in α_2 about the defendant's whereabouts: he is at the pub across the street at the time when e_{α_2} (the saying event) occurs. Let the DRS representing *The defendant is at the pub across the street* be γ. Then, the reader infers $Expt(\alpha_2,\gamma)$, and assuming that $goal(e_\alpha + e_\gamma)$ is defined as intuitions would dictate, this spatio-temporal referent point fixes the whereabouts of the defendant, as at the pub across the street. If he's at the pub across the street, he's not in the courtroom. So $loc(d, goal(e_\alpha + e_\gamma)) \not\subseteq c$ holds, where d is the defendant and c is the courtroom. But, as we said before, because this was just an expectation about the defendant's whereabouts, it's not added to the truth conditional content of the discourse. But it will have discourse effects via the spatial constraint.

Now the task is to attach α_3—the logical form of (1c)—to the preceding SDRS. The only open constituent is α_2, and so $\langle\tau,\alpha_2,\alpha_3\rangle$ is added to the KB. So which rules apply when

[11] $\neg overlaps_s(c,d,t)$ is a gloss for $\neg overlaps_s(stref(c), stref(d), t)$ as used in Asher (1993b), where *stref* stands for spatio-temporal reference. The semantics of $stref(x)$ is defined in Vieu (1991).

attaching α_3? First, we must consider the monotonic reasoning component of CE. Lexical Impotence is verified. This is because first, we have the Bar Rule; and second, the courtroom sense of *bar* doesn't affect the candidate discourse relations that can be used to attach α_3 to α_2. More specifically, there is no rule of the form below, where $Info_{bar_1}(\alpha_2, \alpha_3)$ is a gloss for "information about the semantic content of α and β, where *bar* is interpreted as bar_1":

(13) $(Info_{bar_1}(\alpha_2, \alpha_3) \wedge \langle \tau, \alpha_2, \alpha_3 \rangle) > (R(\alpha_2, \alpha_3) \wedge R \neq Narration)$

So in the monotonic component of CE, we conclude the following law via Lexical Impotence:

- `Strengthened Narration:` $(FS_{bar} \wedge \langle \tau, \alpha_2, \alpha_3 \rangle) > Narration(\alpha_2, \alpha_3)$

By the logic CE, these monotonic inferences are performed before any nonmonotonic inferences, and so Strengthened Narration forms part of the premises to the nonmonotonic component. The rules that apply in this component are: Narration, the Bar Rule and Strengthened Narration.

Given the Spatial Constraint on Narration, $Narration(\alpha_2, \alpha_3)$ and the consequent of the Bar Rule are inconsistent, given the rest of the contents of the reader's KB. For on the one hand, if FS_{bar_1} represents the predicate argument structure of (1c) as intuitions would dictate, then it entails that the defendant is in the courtroom when he's found. On the other hand, the Spatial Constraint on Narration entails that if $Narration(\alpha_2, \alpha_3)$ holds, then the location of the defendant at the source of the finding event is given by $loc(d, goal(e_\alpha + e_\gamma))$. So if $Narration(\alpha_2, \alpha_3)$ holds, then this location *does* now form part of the truth conditional content of the discourse; the defendant is at the pub at the source of the finding event. Hence, if the discourse is narrative, the reader infers from the spatial constraint that the defendant is not in the courtroom at the source of the finding event. But *find* is an achievement verb, and as such it has no temporal extent, and its source is its goal. So the defendant is not in the courtroom when he's found, if the discourse is narrative. This conflicts with the Bar Rule, which predicted that the defendant is in the courtroom when he was found.

So, since the Spatial Constraint on Narration makes FS_{bar_1} and $Narration(\alpha_2, \alpha_3)$ inconsistent in the reader's KB, the Bar Rule and Bar Strengthened Narration conflict. But the latter is more specific, because it's antecedent has the added conjunct $\langle \tau, \alpha_2, \alpha_3 \rangle$. So by the Penguin Principle, $Narration(\alpha_2, \alpha_3)$ is inferred. And the consequence of the Bar Rule is not inferred.

Having inferred $Narration(\alpha_2, \alpha_3)$, the Spatial Constraint on Narration applies. So by logical omniscience, the *bar* where the court bailiff finds the defendant is not bar_1. Therefore *bar* must be bar_2 (i.e., its pub sense) rather than bar_1.

In this example, we saw how coherence constraints on narrative can drive lexical disambiguation. If the intra-sentential information about how a word should be disambiguated conflicts with discourse coherence constraints, then the need for discourse coherence wins. At least, this is the case so long as the intra-sentential information prefers a particular sense *by default*, rather than indefeasibly. This preference for discourse coherence was modelled in a general rule that encoded the interaction between sentential information and discourse information. Lexical Impotence in essence captured the following: if the disambiguation favoured by sentential information doesn't affect which rules for discourse attachment apply, then one can assume that the rules for discourse attachment have priority during NLP.

Now consider the contrastive discourse (2), introduced at the beginning of the paper, and compare it's interpretation in DICE with that of (1).

(2) a. The judge asked where the defendant was.
  b. His barrister apologised, and said he was at the pub across the street.
  c. But in fact, the court bailiff found him slumped underneath the bar.

The logical forms of (2a,b) are as before—α_1 and α_2. And of course, reasoning about their discourse attachment is the same as before: the reader infers $Narration(\alpha_1, \alpha_2)$, $e_{\alpha_1} \prec e_{\alpha_2}$, and $Expt(\alpha_2, \gamma)$, where γ is the DRS representing *The defendant is at the pub across the street.*

Now we must attach (2c) to the SDRS for the preceding discourse. Let the logical form of (2c) be α'_3. First consider the monotonic component. In contrast to α_3, Lexical Impotence *isn't* verified, because the information in α'_3 allows a rule for discourse attachment to apply, whose consequent is a relation other than *Narration*. Namely, the presence of *But* means that a *Contrast* relation must be inferred. So in this example, Strengthened Narration isn't inferred. Instead we infer $Contrast(\alpha_2, \alpha'_3)$. Now consider the nonmonotonic component. Two rules apply: Narration, and the Bar Rule. The *Contrast* relation already inferred conflicts with Narration, and so its consequent isn't inferred. Therefore, the Spatial Constraint on Narration plays no role in interpreting (2). But what about the Bar Rule? In DICE, it is necessary to check that whenever a *Contrast* is inferred (by the presence of *but*, for example), that it has been coherently used: the information in α'_3 must violate an expectation that arose from the discourse context. And indeed, if the Bar Rule fires, then such an expectation is violated. For then, α'_3 entails that the defendant is in the courtroom when he's found, whereas our expectation from the preceding discourse is that he wasn't in the courtroom. So in contrast to (1), the Bar Rule doesn't conflict with the discourse information. It fires in the interpretation of (2), and so *bar* is now assigned its courtroom sense.

5.2 Lexical Information and Strengthening Rhetorical Connections

In text (1), discourse coherence constraints influenced disambiguation. We now examine a further type of discourse-word interaction. Intuitively, the discourse relations *Evidence* and *Consequence* are *scalar*. One can have weak or strong evidential support; and weak or strong consequential links. Here, we analyse texts that feature *Evidence* and *Consequence*, and show that if a word can be disambiguated in favour of the stronger support, then this takes place.

Typically, when attempting to attach constituents together with *Evidence*, the reader is prepared to assume information that's not stated, so as to achieve an appropriate relationship of logical or causal support between the constituents. The strength of evidential support can be measured in terms of the amount of this new information that the reader needs to infer. But the relationship between *Evidence*, new information, and strength of connection is very complex, and forms a focus of current research for the authors. We gloss over the complexities here, by approximating their effect in a plausibility scale. $Pl(\beta \mid \alpha_1) \succ Pl(\beta \mid \alpha_2)$ means "The plausibility of β given α_1 is greater than the plausibility of β given α_2".[12] The plausibility

[12] We assume four axioms characterise the behaviour of the *Pl* scale, which amount to: Validity is maximally plausible; Contradiction is minimally plausible; Asymmetry; and Transitivity. These axioms are compatible but weaker than the axioms that define distributive probability functions.

scale measures the strength of evidential support according to the following assumption: if $Pl(\alpha \mid \beta_1) \succ Pl(\alpha \mid \beta_2)$, then β_1 would provide better evidential support for α than β_2 would.

How does the plausibility scale affect lexical disambiguation? The following law is a general principle about how lexical and discourse information interact. It states the following: if (a) a word *lex* is ambiguous, with (at least) two senses represented by the FSs FS_1 and FS_2; (b) *Evidence* is a candidate relation for attaching β to α (because a default rule for inferring *Evidence* applies), and (c) the sense of the ambiguous word *lex* provided in FS_1 would strengthen the evidential support, compared to interpreting *lex* as FS_2, then disambiguation takes place in favour of the sense FS_1.

- `Strengthening Evidential Support:`

(a) $\Box((FS_1 \rightarrow orth(lex) \wedge FS_2 \rightarrow orth(lex) \wedge$
(b) $X(\alpha) \wedge Y(\beta) \wedge \langle \tau, \alpha, \beta \rangle \wedge$
$(X(\alpha) \wedge Y(\beta) \wedge \langle \tau, \alpha, \beta \rangle > Evidence(\alpha, \beta))$
(c) $\wedge Pl(\alpha_{FS_1} \mid \beta_{FS_1}) \succ Pl(\alpha_{FS_2} \mid \beta_{FS_2}))$
$\rightarrow FS_1)$

The notation β_{FS_1} stands for the logical form of β obtained using FS_1. Similarly, for α_{FS_1}, β_{FS_2} and α_{FS_2}. So (c) above says what we wish, that assuming *lex* is FS_1 provides better evidential support than FS_2 would. Consequently, Strengthening Evidential Support says: the sense of *lex* is FS_1, because this reinforces the relation *Evidence*, that is inferrable from the semantic content of α and β.

Now we consider the impact of this law on the analysis of (4a,c).

(4) a. The EC are decisive.
b. The EC run meetings on time.
c. Last night's meeting came to a conclusion by 8pm.

Let the logical forms of the sentences (4a) and (4c) be respectively α and β. The lexical ambiguity of *conclusion* in β is as yet unresolved, because the intra-sentential information in (4c) alone fails to disambiguate *conclusion* in the monotonic reasoning component.

(4a) is a generic, and α must reflect this. In particular, if α is true, then by WK it follows that all EC meetings are normally (or 'generically') decisive. In contrast, β doesn't quantify over EC meetings; it is about a particular event token of the EC meeting, which happened last night. It says of this event token that it came to a conclusion.

Now, if *conclusion* is interpreted as agreement, then the meeting was decisive. On the other hand, if *conclusion* is interpreted as *end*, then by the linguistic knowledge (LK) in (4c) we can infer that the meeting was over relatively quickly, and by WK we can in turn infer that therefore, the meeting was decisive. Either way, there appears to be a nonmonotonic inference from the reader's KB when interpreting β that the meeting referred to was a decisive one, regardless of how *conclusion* is interpreted.

So, let *meeting*(e) mean that e is an event of the EC meeting, and *decisive*(e) mean that this meeting was decisive. Then we obtain the following relationships between α, β, generic laws and particular statements:

1. By WK, α entails EC meetings are decisive:
   $\Box(\alpha \rightarrow (\forall e)(meeting(e) > decisive(e)))$

2. β strictly entails there is a meeting e':
   $\Box(\beta \rightarrow meeting(e'))$

3. β and LK and WK nonmontonically entail that the meeting in question was a decisive one, regardless of how *conclusion* is interpreted:
   $\beta \cup$ *laws representing* WK *and* LK $\mathrel{|\!\!\approx}$ $decisive(e)$

We declaratively specify a rule, which states that the three above properties are normally sufficient to infer *Evidence*. This amounts to inductive evidential support, for condition (1) above states that α entails a generic statement, condition (2) states that β describes an instance of the antecedent to that generic statement, and condition (3) states that β and other knowledge resources in the KB lead to the consequence of the generic statement. So (2) and (3) together mean that β is a particular example of the generic relationship entailed by α. Thus assuming β is *Evidence* for α is tantamount to assuming a step of induction, from a particular case—β—to the general rule—α. This inductive step is stipulated in Inductive Evidence below, where $\Pi_{Z(a)}$ stands for the nested conditional which specifies in the object language that β and the laws characterising WK and LK together $\mathrel{|\!\!\approx} Z(a)$ (that is, the object language specification of the information in 3 above):[13]

- `Inductive Evidence:`
  $((\langle\tau, \alpha, \beta\rangle \wedge \Box(\alpha \rightarrow (\forall x)(Y(x) > Z(x)))$
  $\wedge \Box(\beta \rightarrow Y(a)) \wedge$
  $\Pi_{Z(a)})$
  $> Evidence(\alpha, \beta)$

This law will apply when attaching (4c) to (4b) or (4a). It will also apply when trying to attach (4e) or (4f) to (4a):

(4)
e. Last night's meeting was over very quickly.
f. Last night's meeting was successful.
g. ?Last night they had a meeting.
h. ?Last night's meeting ended without any agreement.
i. ?Last night's meeting was in Brussels.

But it won't apply when attaching (4g), (4h) or (4i) to (4a), because $\Pi_{decisive(e)}$ will be false.

Recall the logical forms of (4a) and (4c) are α and β. In line with intuitions, we assume a plausibility scale where reaching agreement quickly provides better evidential support for being decisive than meetings coming to an end quickly. So let FS_1 be the lexical entry for *conclusion* corresponding to its *agreement* sense, and FS_2 the lexical entry corresponding to its *end* sense. Then all the following hold: $FS_1 \rightarrow orth(conclusion)$, $FS_2 \rightarrow orth(conclusion)$ and $Pl(\alpha_{FS_1} \mid \beta_{FS_1}) \succ Pl(\alpha_{FS_2} \mid \beta_{FS_2})$. This plausibility scale, together with the fact that

[13] A similar rule is specified in further detail in Asher and Lascarides (1994).

the above Inductive Evidence rule applies for α and β, ensures that when analysing (4a,c), the antecedent of Strengthening Evidential Support is verified in the monotonic component of CE. So it's consequence is inferred: *conclusion* means *agreement*. This shows how the heuristic to strengthen evidential support can cause lexical disambiguation.

Now we turn to the nonmonotonic reasoning, in which a discourse relation between α and β will be inferred: the laws that apply are Inductive Evidence and Narration. The tense structure of (4a,c) conflicts with the consequence of Narration, because e_α holds at the time of speech, whereas e_β held earlier. So Evidence wins, and $Evidence(\alpha, \beta)$ is inferred.[14]

Now consider (4b,c). Suppose the logical form of (4b) is γ. Then the contents of the KB are the same as those above where α is substituted with γ, save that now the following holds: $Pl(\gamma_{FS_2} \mid \beta_{FS_2}) \succ Pl(\gamma_{FS_1} \mid \beta_{FS_1})$. So in the monotonic component of CE, FS_2 is inferred via Strengthening Evidential Support (and so *conclusion* means *end*), and in the nonmonotonic component, by a similar pattern of reasoning, $Evidence(\gamma, \beta)$ is inferred.

We now consider a text where the discourse relation is *Consequence*.

(14) a. They put a plant there.
b. It ruined the view.
c. It improved the view.

Consider first text (14a,b): let the logical forms of the sentences be α and β respectively. The verb *ruin* is a causative verb (Sanfilippo, 1992). Moreover, we assume that the selectional restrictions on *ruin* mean that the pronoun in β must be resolved to the event e_α of putting a plant there. So $cause(e_\alpha, e_\beta)$ is inferred via the lexical semantics of *ruin*. This has an effect at the textual level via Consequence:[15]

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

Finally, there is an analogous law to Strengthening Evidential Support for Consequence:[16]

- `Strengthening Consequential Support:`

$$(FS_1 \rightarrow orth(lex) \wedge FS_2 \rightarrow orth(lex) \wedge \\ X(\alpha) \wedge Y(\beta) \wedge \langle \tau, \alpha, \beta \rangle \wedge \\ (X(\alpha) \wedge Y(\beta) \wedge \langle \tau, \alpha, \beta \rangle > Consequence(\alpha, \beta)) \\ \wedge Pl(\beta_{FS_1} \mid \alpha_{FS_1}) \succ Pl(\beta_{FS_2} \mid \alpha_{FS_2})) \\ \rightarrow FS_1$$

We're now in a position to see how disambiguation occurs in (14a,b). We have already mentioned that when attaching β to α, $cause(e_\alpha, e_\beta)$ is inferred in the monotonic component,

[14] Even if *Narration* were consistent with the facts in the KB, Inductive Evidence would still win over Narration thanks to the Penguin Principle.

[15] The law for inferring Consequence can be generalised, so that *Consequence* is inferred when the constituents are logically related, as well as when they are causal related, but we gloss over this for the sake of simplicity.

[16] Using schemas, we could reduce Strengthening Evidential Support and Strengthening Consequential Support to one rule, but we gloss over this here.

via the lexical semantics of *ruin*. This entails that the antecedent to Consequence is also verified. Let FS_1 be the factory sense of *plant* and FS_2 its fauna sense. Then the plausibility scale reflects the relative aesthetics of factories and flora as long as $Pl(\beta_{FS_1} \mid \alpha_{FS_1}) \succ Pl(\beta_{FS_2} \mid \alpha_{FS_2})$. In other words, it is more plausible for a factory to ruin a view than flora. So the antecedent to Strengthening Consequential Support is also verified, and consequently FS_1 is inferred. That is, *plant* in β is assumed to have its factory sense. In the nonmonotonic component, Consequence and Narration apply. These don't conflict, and so *Consequence* and *Narration* are both inferred. The analysis of (14a,c) is similar, save that the plausibility scale ensures one infers that *plant* has its fauna sense FS_2.

6 Conclusion

In this paper, we have investigated how lexical disambiguation takes place in a discourse context. We showed that domain knowledge and word association are insufficient to account for lexical disambiguation in many cases. By augmenting a formal theory of discourse attachment with lexical knowledge, two important goals were achieved. First, we were able to model how discourse information affects lexical decisions. We offered a theory of lexical processing that uses new knowledge resources: knowledge about rhetorical relations, and the constraints they impose on coherent discourse. And we were able to specify very general heuristics for disambiguation that used this discourse information: Avoid Discourse Incoherence, and Strengthen Rhetorical Connections.

The second important goal was that through adding lexical knowledge to a theory of discourse attachment, many of the pragmatic heuristics and causal laws for deriving discourse structure can be simplified and generalised. We were able to replace lexical items in the laws with the underlying semantic concepts that make those laws plausible, thereby allowing a single law to apply to many closely related lexical items. In essence, we showed how lexical information is used to take decisions about discourse attachment.

Although integrating discourse and lexical processing provides a forum where world knowledge and lexical knowledge interact in precise ways, answers to the long-standing question of where lexical semantics ends and world knowledge starts remain elusive. But nevertheless, we can draw general conclusions about the kinds of lexical structures we require, if lexical processing is to work in service to discourse attachment. In order to infer explanations and elaborations at the textual level, the lexicon must include typed semantic information about causation and part/whole relationships, as well as typed syntactic information. This is unsurprising. But it's a fortunate discovery, since it is compatible with current approaches to computational lexical semantics (e.g., Pustejovsky 1991, Copestake and Briscoe 1991, this volume, Sanfilippo 1992).

Many research questions remain unanswered, the most pressing of which is to reconstruct in DICE general principles about how intra-sentential information affects lexical disambiguation. Furthermore, it is hoped that the techniques used in this paper will extend to an account for how lexical disambiguation is affected by discourse structural constraints on anaphora resolution: in (15a,b), *newspaper* refers to to the paper object, whereas in (15a,c), it refers to the organisation.

(15) a. Max got angry at the newspaper.

b. He had thrown it onto the table, and in the process he had spilt his coffee all over the contract.

c. He had received a rude reply from the editor to his letter accusing him of libel.

This disambiguation is determined by the information connected to the anaphor *it*, which is constrained by the discourse structure to be resolved to the antecedent *newspaper*. Our theory is in a good position to model this line of reasoning, since SDRT already represents constraints on anaphora resolution imposed by discourse structure. These are the topics of current research.

7 References

Alshawi, H. [1992] (ed.) *The Core Language Engine*, MIT Press.

Asher, N. [1993a] *Reference to Abstract Objects in English: A Philosophical Semantics for Natural Language Metaphysics*, Kluwer Academic Publishers, Holland.

Asher, N. [1993b] Temporal and Locational Anaphora in Texts. Research Report, IRIT, Université Paul Sabatier, Toulouse, France.

Asher, N. [1993c] Extensions for Commonsense Entailment, in C. Boutilier and J. Delgrande (eds.), *IJCAI Workshop on Conditionals*

Asher, N. and Lascarides, A. [1994] Intentions and Information in Discourse, *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL94)*, pp34–41, Las Cruces New Mexico, June 1994.

Asher, N. and Morreau, M. [1991] Common Sense Entailment: A Modal Theory of Non-monotonic Reasoning, in *Proceedings to the 12th International Joint Conference on Artificial Intelligence*, Sydney Australia, August 1991.

Blackburn, P. [1992] Modal Logic and Attribute Value Structures, in M. de Rijke (ed.), *Diamonds and Defaults*, Studies in Logic, Language and Information. Kluwer, Dordrecht, Holland (available as University of Amsterdam, ITLI, LP-92-02).

Boguraev, B. [1979] *Automatic Resolution of Linguistic Ambiguities*, PhD Thesis, Computer Laboratory, University of Cambridge.

Briscoe, T., Copestake, A., and Boguraev, B. [1990] Enjoy the Paper: Lexical Semantics via Lexicology, in *Proceedings of COLING90*, volume 2, pp42–47.

Briscoe, T., Copestake, A. and Lascarides, A. [1993] Blocking, in St. Dizier, P. (ed.) *Computational Lexical Semantics*, Cambridge University Press.

Carpenter, R. [1992] *The logic of typed feature structures* Tracts in Theoretical Computer Science, Cambridge University Press, Cambridge, England.

Charniak, E. [1983] Passing Markers: A Theory of Contextual Influence on Language Comprehension, *Cognitive Science*, **7** 3, pp171–190.

Copestake, A. and Briscoe, E. J. [1991] Lexical Operations in a Unification-based Framework, in Pustejovsky, J. and Bergler, S. (eds.) *Proceedings of the ACL SIGLEX Workshop on Lexical Semantics and Knowledge Representation*, pp101–119, Springer Verlag.

Copestake, A. and Briscoe, E. J. [1994] Semi-productive Polysemy and Sense Extension, this volume.

Daelemans, W., De Smedt, K. and Gazdar, G. [1992] Inheritance in Natural Language Processing, *Computational Linguistics*, **18** 2, 205–219.

Grosz, B. J. and Sidner, C. L. [1986] Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, **12**, 175-204.

Guthrie, J., Guthrie, L., Wilks, Y. and Aldinejad, H. [1991] Subject-Dependent Co-Occurrence and Worse Sense Disambiguation, in *Proceedings of the 29th Association for Computational Linguistics*, 55–63, Berkeley, June 1991.

Hayes, P. [1977] *Some Association Based Techniques for Lexical Disambiguation by Machine*, PhD thesis, University of Rochester, New York.

Hirst, G. [1987] *Semantic Interpretation and the Resolution of Ambiguity*, Studies in Natural Language Processing, Cambridge University Press.

Hobbs, J. R. [1985] On the Coherence and Structure of Discourse. Report No: CSLI-85-37, Center for the Study of Language and Information, October 1985.

Hobbs, J. R., Stickel, M., Appelt, D. and Martin, P. [1990] Interpretation as Abduction. Technical Note No. 499, Artificial Intelligence Center, SRI International, Menlo Park.

Kamp, H. [1981] A Theory of Truth and Semantic Representation, in Groenendijk, J. A. G., Janssen, T. M. V. and Stokhof, M. B. J. (eds.) *Formal Methods in the Study of Language* **136**, 277–322. Amsterdam: Mathematical Centre. Tracts.

Kamp, H. and Reyle, U. [in press] *From Discourse to Logic; Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer Academic Publishers, Dordrecht Holland.

Kaplan, R. M., and Bresnan, J. [1982] Lexical-functional Grammar: A Formal System for Grammatical Representation, in Bresnan, J. (ed.) *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge MA.

Lascarides, A. [1992] Knowledge, Causality and Temporal Representation, *Linguistics*, **30** 5, pp941–973.

Lascarides, A. and Asher, N. [1991] Discourse Relations and Defeasible Knowledge, in *Proceedings of the 29th Association for Computational Linguistics*, 55–63, Berkeley, June 1991.

Lascarides, A. and Asher, N. [1993] Temporal Interpretation, Discourse Relations and Commonsense Entailment, to appear in *Linguistics and Philosophy.*

Lascarides, A. and Oberlander, J. [1992] Abducing Temporal Discourse, in Dale, R., Hovy, E., Rösner, D. and Stock, O. (eds.) *Aspects of Automated Natural Language Generation*, Springer-Verlag.

Lascarides, A. and Oberlander, J. [1993] Temporal Coherence and Defeasible Knowledge, *Theoretical Linguistics*, **19**, 1.

McRoy, S. W. [1992] Using Multiple Knowledge Sources for Word Sense Discrimination *Computational Linguistics*, **18** 1, 1–30.

Moore, J. D. and Pollack, M. [1992] A Problem for RST: The Need for Multi-Level Discourse Analysis, *Computational Linguistics*, **18** 4, pp537–544.

Moore, R. C. [1989] Unification-based Semantic Interpretation, in *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pp33–41.

Morreau, M. [1992] *Conditionals in Philosophy and AI*, PhD. thesis, IMS, Universität Stuttgart.

Norvig, P. and Wilensky, R. [1990] A Criticial Evaluation of Commensurable Abduction Models for Semantic Interpretation. In Karlgren, H. (ed.) *Proceedings of COLING90*, 225–230.

Polanyi, L. [1985] A Theory of Discourse Structure and Discourse Coherence. In Eilfort, W. H., Kroeber, P. D. and Peterson, K. L. (eds.) *Papers from the General Session at the Twenty-First Regional Meeting of the Chicago Linguistics Society*, Chicago, April 25-27, 1985.

Pollard, C. and Sag, I. A. [1994] *Head-Driven Phrase Structure Grammar*, University of Chicago Press.

Pustejovsky, J. [1991] The Generative Lexicon. *Computational Linguistics*, **17**, 4, 409–441.

Sanfilippo, A. [1992] Grammatical Relations in Unification Categorial Grammar, *Lingua & Stile*, Fall issue.

Schank, R. C. and Abelson, R. P. [1977] Scripts, Plans and Knowledge, in Johnson-Laird, P. N. and Wason, P. C. (eds.) *Thinking*, Cambridge University Press.

Thompson, S. and Mann, W. [1988] Rhetorical Structure Theory: A Framework for the Analysis of Texts. In *IPRA Papers in Pragmatics*, **1**, 79–105.

Vieu, L. [1991] *Semantique de Deplacement et de la Localisation en Francais: Une Etude des Verbes, des Prepositions et de leur Relations dans la Phrase Simple*, PhD. dissertation, IRIT, Université Paul Sabatier, Toulouse.

Webber, B. [1991] Structure and Ostension in the Interpretation of Discourse Deixis, *Language and Cognitive Processes*, Volume **6** Number 2, pp 1-7–135.

Wilensky, R. [1983] *Planning and Understanding: A Computational Approach to Human Reasoning*, Addison-Wesley.

Wilensky, R. [1990] Extending the Lexicon by Exploiting Subregularities in *Proceedings of COLING90*, 407–412, Helsinki, Finland, July 1990.

Wilks, Y. [1975] A Preferential, Pattern Seeking Semantics for Natural Language Inference, *Artificial Intelligence*, **6**, 53–74.

Wilks, Y., Fass, D., Guo, C., McDonald, J. Plate, T. and Slator, B. [1988] A Tractable Machine Dictionary as a Resource for Computational Semantics. In Boguraev, B. and Briscoe, T. (eds.) *Computational Lexicography for Natural Language Processing*, Harlow, Essex: Longman.