Integrating Symbolic and Statistical Representations: The Lexicon Pragmatics Interface

Ann Copestake
Center for the Study of Language and Information,
Stanford University,
Ventura Hall,
Stanford, CA 94305,
USA
aac@csli.stanford.edu

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

Abstract

We describe a formal framework for interpretation of words and compounds in a discourse context which integrates a symbolic lexicon/grammar, word-sense probabilities, and a pragmatic component. The approach is motivated by the need to handle productive word use. In this paper, we concentrate on compound nominals. We discuss the inadequacies of approaches which consider compound interpretation as either wholly lexico-grammatical or wholly pragmatic, and provide an alternative integrated account.

1 Introduction

When words have multiple senses, these may have very different frequencies. For example, the first two senses of the noun *diet* given in WordNet are:

1. (a prescribed selection of foods)
   => fare – (the food and drink that are regularly consumed)

2. => legislature, legislative assembly, general assembly, law-makers

Most English speakers will share the intuition that the first sense is much more common than the second, and that this is (partly) a property of the word and not its denotation, since near-synonyms occur with much greater frequency. Frequency differences are also found between senses of derived forms (including morphological derivation, zero-derivation and compounding). For example, *canoe* is less frequent as a verb than as a noun, and the induced action use (e.g., *they canoed the kids across the lake*) is much less frequent than the intransitive form (with location PP) (*they canoed across the lake*).[1] A derived form may become established with one meaning, but this does not preclude other uses in sufficiently marked contexts (e.g., Bauer's (1983) example of *garbage man* with an interpretation analogous to *snowman*).

Because of the difficulty of resolving lexical ambiguity, it is usual in NLP applications to exclude 'rare' senses from the lexicon, and to explicitly list frequent forms, rather than to derive them. But this increases errors due to unexpected vocabulary, especially for highly productive derivational processes. For this and other reasons it is preferable to assume some generative devices in the lexicon (Pustejovsky, 1995). Briscoe and Copestake (1996) argue that a differential estimation of the productivity of derivation processes allows an approximation of the probabilities of previously unseen derived uses. If more probable senses are preferred by the system, the proliferation of senses that results from unconstrained use of lexical rules or other generative devices is effectively controlled. An interacting issue is the granularity of meaning of derived forms. If the lexicon produces a small number of very underspecified senses for a wordform, the ambiguity problem is apparently reduced, but pragmatics may have insufficient information with which to resolve meanings, or may find impossible interpretations.

We argue here that by utilising probabilities, a language-specific component can offer hints to a pragmatic module in order to prioritise and control the application of real-world reasoning to disambiguation. The objective is an architecture utilising a general-purpose lexicon with domain-dependent probabilities. The particular issues we consider here are the integration of the statistical and symbolic components, and the division of labour between se-

[1] Here and below we base our frequency judgements on semi-automatic analysis of the written portion of the tagged British National Corpus (BNC).

Arzttermin	*doctor appointment	doctor's appointment
Terminvorschlag	* date proposal	proposal for a date
Terminvereinbarung	* date agreement	agreement on a date
Januarhälfte	* January half	half of January
Frühlingsanfang	* spring beginning	beginning of spring

Figure 1: Some German compounds with non-compound translations

mantics and pragmatics in determining meaning. We concentrate on (right-headed) compound nouns, since these raise especially difficult problems for NLP system architecture (Sparck Jones, 1983).

2 The grammar of compound nouns

Within linguistics, attempts to classify nominal compounds using a small fixed set of meaning relations (e.g., Levi (1978)) are usually thought to have failed, because there appear to be exceptions to any classification. Compounds are attested with meanings which can only be determined contextually. Downing (1977) discusses *apple juice seat*, uttered in a context in which it identifies a place-setting with a glass of apple juice. Even for compounds with established meanings, context can force an alternative interpretation (Bauer, 1983).

These problems led to analyses in which the relationship between the parts of a compound is undetermined by the grammar, e.g., Dowty (1979), Bauer (1983). Schematically this is equivalent to the following rule, where R is undetermined (to simplify exposition, we ignore the quantifier for y):

$$(1)\quad \begin{array}{ccc} \text{N0} & \longrightarrow \text{N1} & \text{N2} \\ \lambda x[P(x) \wedge Q(y) \wedge R(x,y)] & \lambda y[Q(y)] & \lambda x[P(x)] \end{array}$$

Similar approaches have been adopted in NLP with further processing using domain restrictions to resolve the interpretation (e.g., Hobbs et al (1993)).

However, this is also unsatisfactory, because (1) overgenerates and ignores systematic properties of various classes of compounds. Overgeneration is apparent when we consider translation of German compounds, since many do not correspond straightforwardly to English compounds (e.g., Figure 1). Since these exceptions are English-specific they cannot be explained via pragmatics. Furthermore they are not simply due to lexical idiosyncrasies: for instance, *Arzttermin/*doctor appointment* is representative of many compounds with human-denoting first elements, which require a possessive in English. So we get *blacksmith's hammer* and not **blacksmith hammer* to mean 'hammer of a type conventionally associated with a blacksmith' (also *driver's cab*, *widow's allowance* etc). This is not the usual possessive: compare (((his blacksmith)'s) hammer) with (his (blacksmith's hammer)). Adjective placement is also restricted: *three English blacksmith's hammers/ *three blacksmith's English hammers.* We treat these as a subtype of noun-noun compound with the possessive analysed as a case marker.

In another subcategory of compounds, the head provides the predicate (e.g., *dog catcher*, *bottle crusher*). Again, there are restrictions: it is not usually possible to form a compound with an agentive predicate taking an argument that normally requires a preposition (contrast *water seeker* with **water looker*). Stress assignment also demonstrates inadequacies in (1): compounds which have the interpretation 'Y made of X' (e.g., *nylon rope*, *oak table*) generally have main stress on the righthand noun, in contrast to most other compounds (Liberman and Sproat, 1992). Stress sometimes disambiguates meaning: e.g., with righthand stress *cotton bag* has the interpretation *bag made of cotton* while with leftmost stress an alternative reading, *bag for cotton*, is available. Furthermore, ordering of elements is restricted: e.g., *cotton garment bag/ *garment cotton bag.*

The rule in (1) is therefore theoretically inadequate, because it predicts that all noun-noun compounds are acceptable. Furthermore, it gives no hint of likely interpretations, leaving an immense burden to pragmatics.

We therefore take a position which is intermediate between the two extremes outlined above. We assume that the grammar/lexicon delimits the range of compounds and indicates conventional interpretations, but that some compounds may only be resolved by pragmatics and that non-conventional contextual interpretations are always available. We define a number of schemata which encode conventional meanings. These cover the majority of compounds, but for the remainder the interpretation is left unspecified, to be resolved by pragmatics.

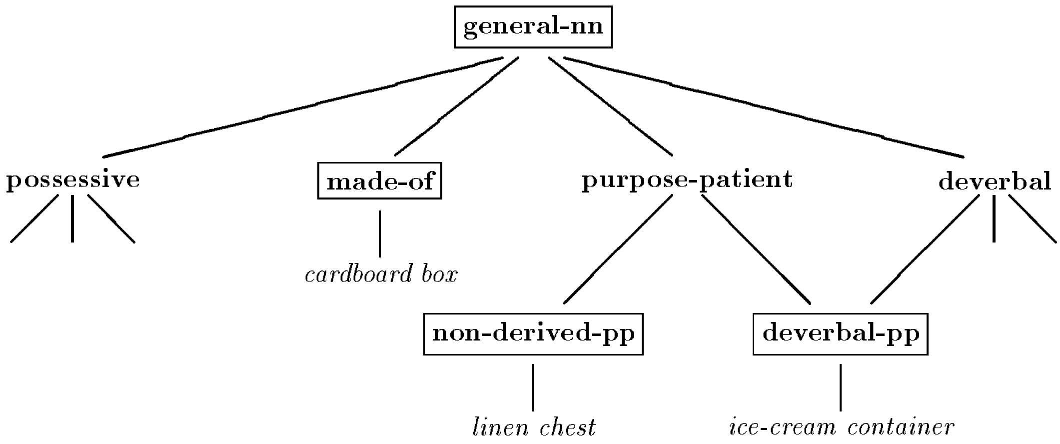

Figure 2: Fragment of hierarchy of noun-noun compound schemata. The boxed nodes indicate actual schemata: other nodes are included for convenience in expressing generalisations.

general-nn	N0 $\lambda x[P(x) \wedge Q(y) \wedge R(x,y)]$ R = /general-nn	->	N1 $\lambda y[Q(y)]$ anything /stressed	N2 $\lambda x[P(x)]$ anything
made-of	R = made-of		substance	physobj /stressed
purpose-patient	R = TELIC(N2)		anything	artifact

Figure 3: Details of some schemata for noun-noun compounds. / indicates that the value to its right is default information.

Space limitations preclude detailed discussion but Figures 2 and 3 show a partial default inheritance hierarchy of schemata (cf., Jones (1995)).[2] Multiple schemata may apply to a single compound: for example, *cotton bag* is an instantiation of the **made-of** schema, the **non-derived-purpose-patient** schema and also the **general-nn** schema. Each applicable schema corresponds to a different sense: so *cotton bag* is ambiguous rather than vague. The interpretation of the hierarchy is that the use of a more general schema implies that the meanings given by specific subschemata are excluded, and thus we have the following interpretations for *cotton bag*:

1. $\lambda x[\text{cotton}(y) \wedge \text{bag}(x) \wedge \text{made-of}(y,x)]$
2. $\lambda x[\text{cotton}(y) \wedge \text{bag}(x) \wedge \text{TELIC}(bag)(y,x)] = \lambda x[\text{cotton}(y) \wedge \text{bag}(x) \wedge \text{contain}(y,x)]$
3. $\lambda x[R(y,x) \wedge \neg(\text{made-of}(y,x) \vee \text{contain}(y,x) \vee \ldots)]$

The predicate made-of is to be interpreted as material constituency (e.g. Link (1983)). We follow Johnston and Busa (1996) in using Pustejovsky's (1995) concept of telic role to encode the purpose of an artifact. These schemata give minimal indications of compound semantics: it may be desirable to provide more information (Johnston et al, 1995), but we will not discuss that here.

Established compounds may have idiosyncratic interpretations or inherit from one or more schemata (though compounds with multiple established senses due to ambiguity in the relationship between constituents rather than lexical ambiguity are fairly unusual). But established compounds may also have unestablished interpretations, although, as discussed in §3, these will have minimal probabilities. In contrast, an unusual compound, such as *apple-juice seat*, may only be compatible with **general-nn**, and would be assigned the most underspecified interpretation. As we will see in §4, this means pragmatics

[2] We formalise this with typed default feature structures (Lascarides et al, 1996). Schemata can be regarded formally as lexical/grammar rules (lexical rules and grammar rules being very similar in our framework) but inefficiency due to multiple interpretations is avoided in the implementation by using a form of packing.

$$\text{Unseen-prob-mass(cmp-form)} = \frac{\text{number-of-applicable-schemata(cmp-form)}}{freq\text{(cmp-form)+number-of-applicable-schemata(cmp-form)}}$$

$$\text{Estimated-freq(interpretation}_i \text{ with cmp-form}_j) = \text{Unseen-prob-mass(cmp-form}_j) \times \frac{Prod(cs_i)}{\sum Prod(cs_1),\ldots,Prod(cs_n)}$$

(where $cs_1 \ldots cs_n$ are the compound schemata needed to derive the n unattested entries for the form$_j$)

Figure 4: Probabilities for unseen compounds: adapted from Briscoe and Copestake (1996)

must find a contextual interpretation. Thus, for any compound there may be some context in which it can be interpreted, but in the absence of a marked context, only compounds which instantiate one of the subschemata are acceptable.

3 Encoding Lexical Preferences

In order to help pragmatics select between the multiple possible interpretations, we utilise probabilities. For an established form, derived or not, these depend straightforwardly on the frequency of a particular sense. For example, in the BNC, *diet* has probability of about 0.9 of occurring in the food sense and 0.005 in the legislature sense (the remainder are metaphorical extensions, e.g., *diet of crime*). Smoothing is necessary to avoid giving a non-zero probability for possible senses which are not found in a particular corpus. For derived forms, the applicable lexical rules or schemata determine possible senses (Briscoe and Copestake, 1996). Thus for known compounds, probabilities of established senses depend on corpus frequencies but a residual probability is distributed between unseen interpretations licensed by schemata, to allow for novel uses. This distribution is weighted to allow for productivity differences between schemata. For unseen compounds, all probabilities depend on schema productivity. Compound schemata range from the non-productive (e.g., the verb-noun pattern exemplified by *pickpocket*), to the almost fully productive (e.g., made-of) with many schemata being intermediate (e.g., has-part: *4-door car* is acceptable but the apparently similar **sunroof car* is not).

We use the following estimate for productivity (adapted from Briscoe and Copestake (1996)):

$$\text{Prod(cmp-schema)} = \frac{M+1}{N}$$

(where N is the number of pairs of senses which match the schema input and M is the number of attested two-noun output forms — we ignore compounds with more than two nouns for simplicity). Formulae for calculating the unseen probability mass and for allocating it differentially according to schema productivity are shown in Figure 4. Finer-grained, more accurate productivity estimates can be obtained by considering subsets of the possible inputs — this allows for some real-world effects (e.g., the made-of schema is unlikely for liquid/physical-artifact compounds).

Lexical probabilities should be combined to give an overall probability for a logical form (LF): see e.g., Resnik (1992). But we will ignore this here and assume pragmatics has to distinguish between alternatives which differ only in the sense assigned to one compound. (2) shows possible interpretations for *cotton bag* with associated probabilities. LFs are encoded in DRT. The probabilities given here are based on productivity figures for fabric/container compounds in the BNC, using WordNet as a source of semantic categories. Pragmatics screens the LFs for acceptability. If a LF contains an underspecified element (e.g., arising from **general-nn**), this must be instantiated by pragmatics from the discourse context.

(2) a. Mary put a skirt in a cotton bag

b.

e, x, y, z, w, t, now
$mary(x)$, $skirt(y)$, $cotton(w)$, $bag(z)$, $put(e, x, y, z)$, $hold(e, t)$, $t \prec now$, $made\text{-}of(z, w)$

$P = 0.84$

c.

e, x, y, z, w, t, now
$mary(x)$, $skirt(y)$, $cotton(w)$, $bag(z)$, $put(e, x, y, z)$, $hold(e, t)$, $t \prec now$, $contain(z, w)$

$P = 0.14$

d.

e, x, y, z, w, t, now
$mary(x)$, $skirt(y)$, $cotton(w)$, $bag(z)$, $put(e, x, y, z)$, $hold(e, t)$, $t \prec now$, $R_c(z, w)$, $R_c = ?$, $\neg(made\text{-}of(z, w) \vee contain(z, w) \vee \ldots)$

$P = 0.02$

4 SDRT and the Resolution of Underspecified Relations

The frequency information discussed in §3 is insufficient on its own for disambiguating compounds. Compounds like *apple juice seat* require marked contexts to be interpretable. And some discourse contexts favour interpretations associated with less frequent senses. In particular, if the context makes the usual meaning of a compound incoherent, then pragmatics should resolve the compound to a less frequent but conventionally licensed meaning, so long as this improves coherence. This underlies the distinct interpretations of *cotton bag* in (3) vs. (4):

(3) a. Mary sorted her clothes into various large bags.
   b. She put her skirt in the cotton bag.

(4) a. Mary sorted her clothes into various bags made from plastic.
   b. She put her skirt into the cotton bag.

If the bag in (4b) were interpreted as being made of cotton—in line with the (statistically) most frequent sense of the compound—then the discourse becomes incoherent because the definite description cannot be accommodated into the discourse context. Instead, it must be interpreted as having the (less frequent) sense given by **purpose-patient**; this allows the definite description to be accommodated and the discourse is coherent. In this section, we'll give a brief overview of the theory of discourse and pragmatics that we'll use for modelling this interaction during disambiguation between discourse information and lexical frequencies. We'll use Segmented Discourse Representation Theory (SDRT) (e.g., Asher (1993)) and the accompanying pragmatic component Discourse in Commonsense Entailment (DICE) (Lascarides and Asher, 1993). This framework has already been successful in accounting for other phenomena on the interface between the lexicon and pragmatics, e.g., Asher and Lascarides (1995), Lascarides and Copestake (1995), Lascarides, Copestake and Briscoe (1996).

SDRT is an extension of DRT (Kamp and Reyle, 1993), where discourse is represented as a recursive set of DRSs representing the clauses, linked together with rhetorical relations such as *Elaboration* and *Contrast*, cf. Hobbs (1985), Polanyi (1985). Building an SDRS involves computing a rhetorical relation between the representation of the current clause and the SDRS built so far. DICE specifies how various background knowledge resources interact to provide clues about which rhetorical relation holds.

The rules in DICE include default conditions of the form $P > Q$, which means *If P, then normally Q*. For example, `Elaboration` states: if β is to be attached to α with a rhetorical relation, where α is part of the discourse structure τ already (i.e., $\langle\tau,\alpha,\beta\rangle$ holds), and β is a subtype of α—which by `Subtype` means that α's event is a subtype of β's, and the individual filling some role θ_i in β is a subtype of the one filling the same role in α—then normally, α and β are attached together with *Elaboration* (Asher and Lascarides, 1995). The `Coherence Constraint on Elaboration` states that an elaborating event must be temporally included in the elaborated event.

- `Subtype:`
  $(\theta_i(e_\alpha,\gamma_1) \wedge \theta_i(e_\beta,\gamma_2) \wedge$
  $\textit{e-condn}_\beta \sqsubseteq \textit{e-condn}_\alpha \wedge \gamma_2 \sqsubseteq \gamma_1) \rightarrow$
  $\textit{Subtype}(\beta,\alpha)$

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

- `Coherence Constraint on Elaboration:`
  $\textit{Elaboration}(\alpha,\beta) \rightarrow e_\beta \subseteq e_\alpha$

`Subtype` and `Elaboration` encapsulate clues about rhetorical structure given by knowledge of subtype relations among events and objects. `Coherence Constraint on Elaboration` constrains the semantic content of constituents connected by *Elaboration* in coherent discourse.

A distinctive feature of SDRT is that if the DICE axioms yield a nonmonotonic conclusion that the discourse relation is R, and information that's necessary for the coherence of R isn't already in the constituents connected with R (e.g., $\textit{Elaboration}(\alpha,\beta)$ is nonmonotonically inferred, but $e_\beta \subseteq e_\alpha$ is not in α or in β), then this content can be added to the constituents in a constrained manner through a process known as SDRS *Update*. Informally, $\textit{Update}(\tau,\alpha,\beta)$ is an SDRS, which includes (a) the discourse context τ, plus (b) the new information β, and (c) an attachment of β to α (which is part of τ) with a rhetorical relation R that's computed via DICE, where (d) the content of τ, α and β are modified so that the coherence constraints on R are met.[3] Note that this is more complex than DRT's notion of update. *Update* models how interpreters are allowed and expected to fill in certain gaps in what the speaker says; in essence affecting semantic contet through context and pragmatics. We'll use this information

[3] If R's coherence constraints can't be inferred, then the logic underlying DICE guarantees that R won't be nonmonotonically inferred.

flow between context and semantic content to reason about the semantic content of compounds in discourse: simply put, we will ensure that words are assigned the most freqent possible sense that produces a well defined SDRS *Update* function.

An SDRS S is well-defined (written $\downarrow S$) if there are no conditions of the form $x = ?$ (i.e., there are no unresolved anaphoric elements), and every constituent is attached with a rhetorical relation. A discourse is incoherent if $\neg \downarrow Update(\tau, \alpha, \beta)$ holds for every available attachment point α in τ. That is, anaphora can't be resolved, or no rhetorical connection can be computed via DICE.

For example, the representations of (5a,b) (in simplified form) are respectively α and β:

(5) a. Mary put her clothes into various large bags.

α.	$x, Y, Z, e_\alpha, t_\alpha, n$
	$mary(x)$, $clothes(Y)$, $bag(Z)$, $put(e_\alpha, x, Y, Z)$, $hold(e_\alpha, t_\alpha)$, $t_\alpha \prec n$

b. She put her skirt into the bag made out of cotton.

β.	$x, y, z, w, e_\beta, t_\beta, n, u, B$
	$mary(x)$, $skirt(y)$, $bag(z)$, $cotton(w)$, $made\text{-}of(z, w)$, $u = ?$, $B(u, z)$, $B = ?$, $put(e_\beta, x, y, z)$, $hold(e_\beta, t_\beta)$, $t_\beta \prec n$

In words, the conditions in β require the object denoted by the definite description to be linked by some 'bridging' relation B (possibly identity, cf. van der Sandt (1992)) to an object u identified in the discourse context (Asher and Lascarides, 1996). In SDRT, the values of u and B are computed as a byproduct of SDRT's *Update* function (cf. Hobbs (1979)); one specifies u and B by inferring the relevant new semantic content arising from R's coherence constraints, where R is the rhetorical relation inferred via the DICE axioms. If one cannot resolve the conditions $u = ?$ or $B = ?$ through SDRS update, then by the above definition of well-definedness on SDRSs the discourse is incoherent (and we have presupposition failure).

The detailed analysis of (3) and (52) involve reasoning about the values of u and B. But for reasons of space, we gloss over the details given in Asher and Lascarides (1996) for specifying u and B through the SDRT update procedure. However, the axiom `Assume Coherence` below is derivable from the axioms given there. First some notation: let $\beta[C]$ mean that β contains condition C, and assume that $\beta[C/C']$ stands for the SDRS which is the same as β, save that the condition C in β is replaced by C'.

Then in words, `Assume Coherence` stipulates that if the discourse can be coherent only if the anaphor u is resolved to x and B is resolved to the specific relation P, then one *monotonically* assumes that they are resolved this way:

- `Assume Coherence:`

$$(\downarrow Update(\tau, \alpha, \beta[u = ?, B = ?/u = x, B = P]) \wedge (C' \neq (u = x \wedge B = P) \rightarrow \neg \downarrow Update(\tau, \alpha, \beta[u = ?, B = ?/C']))) \rightarrow (Update(\tau, \alpha, \beta) \leftrightarrow Update(\tau, \alpha, \beta[u = ?, B = ?/u = x, B = P]))$$

Intuitively, it should be clear that in (5a,b) $\neg \downarrow Update(\alpha, \alpha, \beta)$ holds, unless the bag in (5b) is one of the bags mentioned in (5a)—i.e., $u = Z$ and $B = member\text{-}of$. For otherwise the events in (5) are too 'disconnected' to support any rhetorical relation. On the other hand, assigning u and B these values allows us to use `Subtype` and `Elaboration` to infer *Elaboration* (because skirt is a kind of clothing, and the bag in (5b) is one of the bags in (5a)). So `Assume Coherence`, `Subtype` and `Elaboration` yield that (5b) elaborates (5a) and the bag in (5b) is one of the bags in (5a).

Applying SDRT to compounds encodes the effects of pragmatics on the compounding relation. For example, to reflect the fact that compounds such as *apple juice seat*, which are compatible only with **general-nn**, are acceptable only when context resolves the compound relation, we assume that the DRS conditions produced by this schema are: $R_c(y, x)$, $R_c = ?$, and $\neg(made\text{-}of(y, x) \vee contain(y, x) \vee \ldots)$. By the above definition of well-definedness on SDRSs, the compound is coherent only if we can resolve R_c to a particular relation via the SDRT *Update* function, which in turn is determined by DICE. Rules such as `Assume Coherence` serve to specify the necessary compound relation, so long as context provides enough information.

5 Integrating Lexical Preferences and Pragmatics

We now extend SDRT and DICE to handle the probabilistic information given in §3. We want the pragmatic component to utilise this knowledge, while still maintaining sufficient flexibility that less frequent senses are favoured in certain discourse contexts.

Suppose that the new information to be integrated with the discourse context is ambiguous between $\beta_1, \ldots, \beta_n$. Then we assume that exactly one of $Update(\tau, \alpha, \beta_i)$, $1 \leq i \leq n$, holds. We gloss this complex disjunctive formula as

$!\bigvee_{1 \leq i \leq n}(Update(\tau, \alpha, \beta_i))$. Let $\beta_k \succ \beta_j$ mean that the probability of DRS β_k is greater than that of β_j. Then the rule schema below ensures that the most frequent possible sense that produces discourse coherence is (monotonically) favoured:

- `Prefer Frequent Senses:`
  $$(!\bigvee_{1 \leq i \leq n}(Update(\tau, \alpha, \beta_i)) \wedge \\ \downarrow Update(\tau, \alpha, \beta_j) \wedge \\ (\beta_k \succ \beta_j \rightarrow \neg \downarrow Update(\tau, \alpha, \beta_k))) \rightarrow \\ Update(\tau, \alpha, \beta_j)$$

`Prefer Frequent Senses` is a declarative rule for disambiguating constituents in a discourse context. But from a procedural perspective it captures: try to attach the DRS based on the most probable senses first; if it works you're done; if not, try the next most probable sense, and so on.

Let's examine the interpretation of compounds. Consider (3). Let's consider the representation β' of (3b) with the highest probability: i.e., the one where cotton bag means *bag made of cotton*. Then similarly to (5), `Assume Coherence`, `Subtype` and `Elaboration` are used to infer that *the cotton bag* is one of the bags mentioned in (3a) and *Elaboration* holds. Since this updated SDRS is well-defined, `Prefer Frequent Senses` ensures that it's true. And so *cotton bag* means *bag made from cotton* in this context.

Contrast this with (4). $Update(\alpha, \alpha, \beta')$ is not well-defined because the *cotton bag* cannot be one of the bags in (4a). On the other hand, $Update(\alpha, \alpha, \beta'')$ is well-defined, where β'' is the DRS where *cotton bag* means *bag containing cotton*. This is because one can now assume this bag is one of the bags mentioned in (4a), and therefore *Elaboration* can be inferred as before. So `Prefer Frequent Senses` ensures that $Update(\alpha, \alpha, \beta'')$ holds but $Update(\alpha, \alpha, \beta')$ does not.

`Prefer Frequent Senses` is designed for reasoning about word senses in general, and not just the semantic content of compounds: it predicts *diet* has its food sense in (6b) in isolation of the discourse context (assuming $Update(\emptyset, \emptyset, \alpha) = \alpha$), but it has the law-maker sense in (6), because SDRT's coherence constraints on *Contrast* ((Asher, 1993))—which is the relation required for *Update* because of the cue word *but*—can't be met when *diet* means *food*.

(6) a. In theory, there should be cooperation between the different branches of government.
    b. But the president hates the diet.

In general, pragmatic reasoning is computationally expensive, even in very restricted domains. But the account of disambiguation we've offered circumscribes pragmatic reasoning as much as possible. All nonmonotonic reasoning remains packed into the definition of $Update(\tau, \alpha, \beta)$, where one needs pragmatic reasoning anyway for inferring rhetorical relations. `Prefer Frequent Senses` is a monotonic rule, it doesn't increase the load on nonmonotonic reasoning, and it doesn't introduce extra pragmatic machinery peculiar to the task of disambiguating word senses. Indeed, this rule offers a way of checking whether fully specified relations between compounds are acceptable, rather than relying on (expensive) pragmatics to compute them.

We have mixed stochastic and symbolic reasoning. Hobbs *et al* (1993) also mix numbers and rules by means of weighted abduction. However, the theories differ in several important respects. First, our pragmatic component has no access to word forms and syntax (and so it's not language specific), whereas Hobbs *et al*'s rules for pragmatic interpretation can access these knowledge sources. Second, our probabilities encode the frequency of word senses associated with word forms. In contrast, the weights that guide abduction correspond to a wider variety of information, and do not necessarily correspond to word sense/form frequencies. Indeed, it is unclear what meaning is conveyed by the weights, and consequently the means by which they can be computed are not well understood.

6 Conclusion

We have demonstrated that compound noun interpretation requires the integration of the lexicon, probabilistic information and pragmatics. A similar case can be made for the interpretation of morphologically-derived forms and words in extended usages. We believe that the proposed architecture is theoretically well-motivated, but also practical, since large-scale semi-automatic acquisition of the required frequencies from corpora is feasible, though admittedly time-consuming. However further work is required before we can demonstrate this, in particular to validate or revise the formulae in §3 and to further develop the compound schemata.

7 Acknowledgements

The authors would like to thank Ted Briscoe and three anonymous reviewers for comments on previous drafts. This material is in part based upon work supported by the National Science Foundation under grant number IRI-9612682 and ESRC (UK) grant number R000236052.

References

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

Asher, N. and A. Lascarides (1995) 'Lexical Disambiguation in a Discourse Context', *Journal of Semantics, vol.12.1,* 69–108.

Asher, N. and A. Lascarides (1996) *Bridging,* Proceedings of the International Workshop on Semantic Underspecification, Berlin, October 1996, available from the Max Plank Institute.

Bauer, L. (1983) *English word-formation,* Cambridge University Press, Cambridge, England.

Briscoe, E.J. and A. Copestake (1996) 'Controlling the application of lexical rules', *Proceedings of the ACL SIGLEX Workshop on Breadth and Depth of Semantic Lexicons,* Santa Cruz, CA.

Downing, P. (1977) 'On the Creation and Use of English Compound Nouns', *Language, vol.53(4),* 810–842.

Dowty, D. (1979) *Word meaning in Montague Grammar,* Reidel, Dordrecht.

Hobbs, J. (1979) 'Coherence and Coreference', *Cognitive Science, vol.3,* 67–90.

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

Hobbs, J.R., M. Stickel, D. Appelt and P. Martin (1993) 'Interpretation as Abduction', *Artificial Intelligence, vol.63.1,* 69–142.

Johnston, M., B. Boguraev and J. Pustejovsky (1995) 'The acquisition and interpretation of complex nominals', *Proceedings of the AAAI Spring Symposium on representation and acquisition of lexical knowledge,* Stanford, CA.

Johnston, M. and F. Busa (1996) 'Qualia structure and the compositional interpretation of compounds', *Proceedings of the ACL SIGLEX workshop on breadth and depth of semantic lexicons,* Santa Cruz, CA.

Jones, B. (1995) 'Predicting nominal compounds', *Proceedings of the 17th Annual conference of the Cognitive Science Society,* Pittsburgh, PA.

Kamp, H. and U. Reyle (1993) *From Discourse to Logic: an introduction to modeltheoretic semantics, formal logic and Discourse Representation Theory,* Kluwer Academic Publishers, Dordrecht, Germany.

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

Lascarides, A., E.J. Briscoe, N. Asher and A. Copestake (1996) 'Persistent and Order Independent Typed Default Unification', *Linguistics and Philosophy, vol.19:1,* 1–89.

Lascarides, A. and A. Copestake (in press) 'Pragmatics and Word Meaning', *Journal of Linguistics,*

Lascarides, A., A. Copestake and E. J. Briscoe (1996) 'Ambiguity and Coherence', *Journal of Semantics, vol.13.1,* 41–65.

Levi, J. (1978) *The syntax and semantics of complex nominals,* Academic Press, New York.

Liberman, M. and R. Sproat (1992) 'The stress and structure of modified noun phrases in English' in I.A. Sag and A. Szabolsci (eds.), *Lexical matters,* CSLI Publications, pp. 131–182.

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

Polanyi, L. (1985) 'A Theory of Discourse Structure and Discourse Coherence', *Proceedings of the Papers from the General Session at the Twenty-First Regional Meeting of the Chicago Linguistics Society,* Chicago, pp. 25–27.

Pustejovsky, J. (1995) *The Generative Lexicon,* MIT Press, Cambridge, MA.

Resnik, P. (1992) 'Probabilistic Lexicalised Tree Adjoining Grammar', *Proceedings of the Coling92,* Nantes, France.

van der Sandt, R. (1992) 'Presupposition Projection as Anaphora Resolution', *Journal of Semantics, vol.19.4,*

Sparck Jones, K. (1983) 'So what about parsing compound nouns?' in K. Sparck Jones and Y. Wilks (eds.), *Automatic natural language parsing,* Ellis Horwood, Chichester, England, pp. 164–168.

Webber, B. (1991) 'Structure and Ostension in the Interpretation of Discourse Deixis', *Language and Cognitive Processes, vol.6.2,* 107–135.