Annotation Guidelines for NomBank –
Noun Argument Structure for PropBank
2007

Adam Meyers

1   Introduction

In the PropBank project [11, 10, 29], a research group at the University of Pennsylvania annotated the verbal argument structure for the Wall Street Journal Corpus of the Penn Treebank (PTB). This document outlines a new effort, called NomBank, undertaken at New York University. In this project, we annotate nominal argument structure on the same corpus, using the same formalism.

Our goal is to annotate each "markable" NP, marking the head, its arguments and "allowed" adjuncts in the style of PropBank. Some sample annotation is given as figure 1. To get a general sense of the annotation, note that for most obvious nominalizations (*knowledge, solicitation, director*), the subjects of the corresponding verbs are (usually) ARG0 and the objects are (usually) ARG1.[1] This generalization also holds even if one takes a fanatically broad view of the definition of nominalization and assume that *picture*, *assembly* and *type* are nominalizations. Under this interpretation, we would take the transitive verb of assemble as a model for *assembly* (*John assembled the shareholders in a room*), the verbs *photograph* or *draw* as the basis of the entry for *picture* in *Mary's picture of John* and a verb like *categorize* as the basis for the lexical entry for *type* in *a type of cheese* even though *type* lacks the ARG0 argument – the one doing the classification. In other words, we attempt to find previously marked predicates that take similar arguments as these nouns, regardless of whether or not there is a morphological connection between the model verb and the noun under analysis. The motivation is that we would like the whole system (PropBank, NomBank, EtcBank)[2] to be as consistent

---

[1]ARG0 and ARG1 actually reflect some notion of typical subject and object. This is the same approach to role labels that was taken in PropBank, as well as in Relational Grammar of the 1970s and 1980s.

[2]We propose to extend this formalism for predicate argument structure to other parts of speech including adjectives, degree words, quantifiers and idioms. We will henceforth refer to this proposed

1. *students' knowledge of two-letter consonant sounds*
   ARG0 = students, REL = knowledge, ARG1 = two-letter consonant sounds

2. *a solicitation of shareholder consents*
   REL = solicitation, ARG1 = shareholder consents

3. *the museum's director*
   REL = director, ARG0 = director, ARG2 = the museum's

4. *her husband*
   REL = husband, ARG0 = husband, ARG1 = her

5. *a set of tasks*
   REL = set, ARG1 = of tasks

6. *a cascade of genetic damage*
   REL = cascade, ARG1 = of genetic damage

7. *the early part of this century*
   REL = part, ARG1 = of this century, ARGM-TMP = early

8. *a picture of the quake's impact*
   REL = picture, ARG1 = of the quake's impact

9. *Carl Bernstein's book about Watergate*
   REL = book, ARG0 = Carl Bernstein's, ARG1 = about Watergate

10. *an assembly of shareholders*
    REL = assembly, ARG1 = of shareholders

11. *a different type of filter*
    REL = type, ARG1 = of filter, ARGM-MNR = different

Figure 1: Sample Annotation from PTB WSJ Corpus

as possible with respect to the way the arguments are marked across predicates. We are less concerned with whether the nominal predicate and the parallel verbal argument structure can paraphrase one another, although this frequently is the case.

The adjuncts that are marked include only adjunct types that have also been marked for verbs. In figure 1, manner and temporal adjuncts have been marked. However, there are many types of nonarguments of nouns, which we do not mark. For example, the determiners *the* and *a* are never marked. Adjectives can only be marked when they fill an argument role or one of the allowed adjunct roles. For example, *bad, slow, lunar* are markable in the phrases *his bad behavior*, *a slow decrease* and *her lunar journey* because these modifiers can be paraphrased in terms of verbs, e.g., one can behave badly, something can decrease slowly and someone can journey to the moon. In contrast, *the* and *rubber* are not markable in the phrase *the rubber bathtub stopper* because these do not correspond to any verb adjunct or argument position (although *bathtub* is the ARG1 of *stopper*).

Of the head nouns in figure 1, *husband*, *set* and *part* do not appear to be nominalizations, even if we stretch our imagination. However, we still use the same set of role types to describe them. This is possible because we create classes of nouns that include both nominalizations and non-nominalizations and extend the roles used to cover all members of the class. Relational nouns include *director* and *husband*. Partitive nouns include *set*, *part* and *cascade*. Our full set of noun classes is described in Section 4. We assume that arguments not covered by these classes are not markable.

Efficient annotation decisions require clear criteria and finite numbers of cases. This is the motivation behind placing limitations on both the classes of markable argument-taking nouns and the set of markable adjunct types.

1.1 A Brief Summary of the Task

Our task involves first creating lexical entries that define the senses and the possible arguments for each noun. These lexical entries, for example determine what constitutes an ARG0, ARG1, etc. for a particular noun. Details are provided in section 4. Given a lexical entry, an annotator must go through all the instances of that noun in the corpus (including plurals, spelling variations, etc.) and identify which instances are markable (section 3). For each markable instance, a roleset[3] is chosen from the lexical entry and the arguments are assigned as per that roleset. Adjuncts are also marked according to the guidelines in section 11. In addition,

---

work as EtcBank. We also assume that EtcBank will include the Penn Discourse Treebank [21] or PDTB.

[3]*Roleset* is a similar concept to *sense*. It is defined in section 4.

the annotator must revise the lexical entries based on what is found in the corpus. Thus lexicon creation and corpus annotation are actually intermingled.

1.2 The remaining sections of these Specifications

The next section suggests how NomBank might be used by the NLP community. The remaining sections provide the details for understanding the argument structure of nouns as represented in NomBank. First we go over how it is determined what is and is not markable. Then we describe the structure of our lexicon, the relation of nominalizations to verbs and adjectives and our various other lexical classes.

2   How will NomBank be Used?

When complete, the work described in this document will include offset annotation for all noun arguments in the PropBank corpus and a set of related lexicons. This section discusses how NLP systems will be able to use these NomBank resources, both independently and in conjunction with PropBank. While many applications are possible, I will concentrate on Information Extraction. Please keep in mind that generalizations of the same sort also improve other types of systems, e.g., Question Answering, Machine Translation, Document Summarization, etc.

By itself, the annotated corpus provides the means for identifying syntactic alternations within NPs and computing various probabilities. For example, there are over 500 instances of the word *director* in the corpus. Suppose that an ARG2 occurs with 375 instances and of these instances, it occurs 220 times as a postnominal *of* phrase, 80 times as a possessive, 27 times as a singular noun left modifier, 22 times as a *for* phrase, 17 times as an adjectival left modifier and 9 times as an *in* phrase. Based on these frequencies and using statistical techniques, it would seem fairly straightforward to extract instances of *director* together with instances of these ARG2s. A system that is trying to determine where people are employed at a particular time (e.g., the MUC management succession scenario) could generalize over ARG2s of *director* rather than using separate patterns for *IBM's finance director*, *the finance director of IBM*, *the director of finance for IBM*, etc.

Our lexicon provides the means for even greater generalization. For nominalizations of verbs, the PropBank annotation can be used in conjunction with the NomBank annotation. Once again, suppose a system used PropBank and Nom-Bank as a basis for generating all the possible propositions from some new corpus. One could assume that the nominalizations *appointment* and *appointee* form an equivalence class with the verb forms of *appoint*. Then, a single IE pattern would

PARTITIVE: *dozens of attorneys who received letters*
    REL = dozens, ARG1 = of attorneys who received letters

SHARE: *each company's share of liability*
    REL = share, ARG0 = each company's, ARG1 = liability

TYPE: *his unique brand of Christianity*
    REL = brand, ARG1 = of Christianity

VERSION: *the House version of the deficit-cutting bill*
    REL = version, ARG0 = House, ARG1 = of the deficit-cutting bill

Figure 2: Some Transparent Noun Examples

cause a system to extract information from *IBM's appointment of John Smith as the director of finance*, *IBM's director of finance appointee* and *Apple appointed Mary Jones (as) director of finance*.[4] In this way, the generalizations implicit in Nom-Bank and PropBank will multiply so that a single pattern will handle variation both in the verb domain and the noun domain. The NomBank lexicon also notes that certain nouns can be used as part of complex discourse connectives, allowing users to make some connections with the PDTB.[5] As the NomBank lexicon also makes reference to nominalizations of adjectives and nominalizations related to adverbs, we are ensuring that NomBank will interact with future EtcBank projects involved in the marking of adjectival and adverbial argument structure.

   The lexicon will also provide the means for other sorts of generalization. For example, systems could use back off models which include both general patterns with NomBank classes as well as specific patterns for particular words. For example the class of RELATIONAL nouns might license several general patterns in the management succession domain, e.g., COMPANY hires PERSON as RELATIONAL-NOUN.

   Finally, certain noun classes have special properties. For example, NPs headed by PARTITIVE nouns typically take the semantic class of the ARG1. Thus, *dozens of attorneys* takes the semantic class PERSON, but *dozens of fiber-end bunches* does not. Similarly, the classes of *a whole bunch of people* and *a wide variety of crops* are determined by the object of *of*, rather than the head noun. This information is very useful for classifying NPs for a wide variety of tasks, including IE.

---

[4]Further regularization will put *director of finance* and *finance director* into an equivalence class.
    [5]NomBank identifies the arguments of the connective, but does not provide some of the other information found in that resource. By and large, however, the connectives marked are supplementary to those marked by the PDTB.

SHARE, VERSION and TYPE nouns can also have this property. Some examples of such "transparent" noun classes are provided as figure 2. In addition, there are some other cases where nouns are "nearly" transparent. For example, *stage* in *its stages of development* acts as links to its argument. Here "its" is the ARG1 of *development*. Thus *stage* links the possessive even though *stages* are pieces of events (PARTITIVE/PIECE) and not identified with their ARG1s like regular partitives (cf. section 4.3 for further details about these classes).

2.1 Spinoff Dictionaries

Annotation in NomBank, like PropBank, is guided by a dictionary of frames listing rolesets for each word. In addition, we have created a simple morphology dictionary (nombank-morph.dict) for nouns which maps not only singles to base forms, but also hyphenated words and alternate spellings to the base noun form. This is intended to be used to map nouns to the appropriate NomBank frame entry. Descriptions of these two dictionaries are found in section 4.

As a side effect of the NomBank effort, we have also created a number of additional dictionaries [20]. These were created in a semi-automatic way using both our lexical knowledge and several heuristics for approximating lexical entries. The focus was on coverage, but not necessarily high precision. The original intention of creating these was for our own internal development purpose. However, we found that other researchers were interested in using them. These dictionaries are all available in a lisp-like notation (similar to COMLEX Syntax and NOMLEX [15, 16]). The other NomBank dictionaries include:

1. NOMLEX-PLUS, which extends the hand-coded (1000 entry) NOMLEX [16] to over 8000 entries including approximately 7000 entries that are a product of semi-automatic procedures. This dictionary is useful for identifying nominal argument structure in much the same way as NomBank. However, no machine learning is necessary. This includes verbal nominalization entries like NOMLEX, adjectival nominalization entries and special entries for the various NOMBANK classes.

2. ADJADV, maps adjectives (*slow*, *probable*) to the corresponding adverbs (*slowly*, *probably*) and lists the appropriate Comlex Syntax modification (:MODIF) classes defined as in Comlex [31].

3. NOMADV, maps nouns to corresponding adverbs in the same manner as ADJADV. Many of these nouns correspond to discourse connective and other adverbial nouns described in the frame files.

6

We will provide brief descriptions of these dictionaries in this document, but more detailed descriptions in the accompanying document entitles *Those Other NomBank Dictionaries*. In particular, sectionnomadv-sect will discuss how these dictionaries were useful for developing NomBank frame files of nouns related to adverbs. Section 11.12 will explain how to use ADJADV and NOMADV to identify adverb-like modifiers of nouns and determine how they correspond to their verbal counterparts.

3   What is Markable and What is not Markable?

The question, "what is markable?" actually breaks down into at least two questions: "Which NPs are markable?" and "Which constituents of NPs are markable?". This section breaks down these questions even further, distinguishing among subtypes of markable constituents of NPs: arguments and "proposition-modifying" adjuncts.

3.1   What is Markable?

Our goal is to annotate all noun phrases (NPs) with nominal argument structure. To be markable, an NP N must satisfy one of the following conditions:

  i. N must contain at least one (unincorporated) argument of the head noun.

  ii. The head of N must be of a "propositional" type (representing an event, state, etc.) and N must contain at least one "proposition-modifying" adjunct.

  iii. The head of N takes an argument in one of the ways described in the later sections of this manual, beginning with section 5. This includes support and other constructions. For pedagogical purposes, we introduce other NomBank concepts before describing these more complex argument taking environments.

A "proposition-modifying" adjunct of a noun is an adjunct of a type that can also be part of verbal argument structure. These include the nominal versions of both traditional verbal adjuncts (locational, temporal, etc.) and so-called sentential adjuncts, e.g., the adjunct *probably* for verbs corresponds to the adjective *probable* for nouns. We assume that other constituents of the NP are not markable parts of its argument structure.

Intuitively, a noun argument is a constituent that can co-occur with a particular head noun, but not with head nouns in general, i.e., the noun "selects" its arguments. More specifically, the argument and head noun must be in a particular relation that is idiosyncratic to a small set of argument/head noun pairs. For example,

math, but not *tall* are arguments of *teacher* in the phrases *the math teacher* and *the tall teacher*. *Tall* can co-occur with a wide variety of head nouns, and the relation between *tall* and these nouns is based primarily on the compatibility of the meaning of *tall* and the head noun.[6] In contrast, *the math teacher* derives its meaning from the argument-taking properties of *teacher*, i.e., teachers teach subjects.

We consider all non-heads within a noun phrase as potential arguments: possessives, prenominal modifiers, adjectives, PPs, post-nominal clauses, etc. Their argument status is determined by their relation to the head noun. If the constituent fits, the whole constituent is considered as the argument. Thus the possessive includes the possessive *'s* and the PP includes the preposition.

We extend our notion of argument to include both: (1) various arguments that are outside the NP, particularly shared arguments of "support" verbs, e.g., "John" is the subject of "walk" in "John took a walk" (later sections discuss the full inventory of cases where arguments can occur outside of the NP); and (2) incorporated arguments, e.g., words like "teacher" and "appointee" are, in a sense, their own subjects or objects on the analogy of verbal argument structure. Although we include incorporated arguments in our representation of argument structure, an incorporated argument by itself is not sufficient for us to consider an NP markable. For example, while *the math teacher* is markable (ARG0 = *teacher*, ARG1 = *math*), the NP *the teacher* is not markable, because the word itself is the only argument. In particular, note that *teacher*, not *John* is the ARG0 in a sentence like *John was a math teacher*. By itself, the phrase *a math teacher* means "one who teaches math", and *John* is identified with that phrase by means of the copula. Typical argument nominalizations include subject nominalizations (*teacher*, *destroyer*, *accelerator*), object nominalizations (*nominee*, *affiliate*, *consultant*) and indirect object nominalizations (*addressee*, *lessee*, *payee*).[7]

A propositional NP represents an event, relation or state. Although most of these NPs include at least one argument, some do not. We decided to only mark propositional NPs that contain either an argument or a proposition-modifying adjunct because: (1) modifiers provide valuable clues for classification, including sense disambiguation – without modifiers, it is more difficult to determine if an NP is "propositional"; and (2) unmodified NPs provide very little information for the user. Thus we decided not to mark bare NPs because that annotation would be of less practical use and of lower accuracy in comparison with the other annotation. In rare cases, a propositional noun may not take any arguments, but still allow

---

[6]The reason that *the tall noise is ill-formed is that the meaning of *tall* is not so compatible with the meaning of *noise* – in fact, one could imagine a Science Fiction story with different rules of reality where *the tall noise* is well-formed.

[7]Some more exotic types of nominalizations may include particles, e.g., *bailout*. Some argument nominalizations also include the particle, e.g., *runaway* is a subject nominalization.

modification by proposition-modifying adjuncts, e.g., *Last year's drought*. Such NPs are markable under these guidelines.

3.2 Unmarkable NPs: Proper Nouns, Titles, Most Idioms, Metaphor, Headless NPs, . . .

To further narrow our task, we eliminate all proper noun phrases from consideration. In principle, some proper noun phrases have argument structure, just like the cases which we are covering. For example, although we would assign argument structure to "a photograph collector", we would not assign structure to "The Photograph Collector", the name of a newsletter. For our purposes, proper noun phrases (also known as named entities) are viewed as unanalyzable wholes. For the case in point, the name does not actually suggest that the newsletter collects photographs, but rather that its readers may. As the significance of the argument structure (if any) of names varies quite a bit, it would seem misleading to assign argument structure to proper noun phrases.

Similarly, most idioms and idiom-like units are removed from consideration. Thus "lieu" and "spite" are not marked when they are part of the multiword prepositions "in lieu of" and "in spite of", even though they are marked as nouns (NN) in the PTB. Similarly "sake" and "midst" are left unmarked in "for the sake of X", "in the midst of X" and their possessive variants "for X's sake" and "in X's midst" [25].[8]

There are certain idiom-like items that we mark, namely combinations of support verbs and nouns, e.g., "take advantage of" and "keep tabs on". See section 5 for details. For idioms that are not anchored by support verb plus noun combinations, a determination must be made as to whether or not they are markable. There are two crucial factors: (1) Can this noun take arguments in the absence of specific collocated lexical items (other than prepositions that follow the head noun), i.e., outside of an idiomatic use?[9] and (2) Is the argument structure (role assignment) of the idiomatic use basically the same as that of the nonidiomatic use? If the answer to either of these questions is "no", than the idiomatic use should not be marked. The second limitation amounts to a prohibition against a frame requiring that one of the arguments be filled by a specific word or phrase. While this eliminates some idioms, it does not eliminate all of them. It also provides a simple

---

[8]There are some arbitrary decisions that have to be made here in defining a complex preposition. For example, we assume that *favor* in *in favor of* is a markable noun, in part because it shares the meaning of the verb *favor*. Similarly, *lot* and *lots* in a *a lot of* and *lots of* act like a normal partitive noun and we decided to mark them.

[9]If in doubt, the question, "Does this noun appear in the corpus in a nonidiomatic use?" can be answered instead.

1. *the butt of so many jokes* [NOT MARKABLE]

2. *with a grain of salt* [MARKABLE]

3. *line of credit* [NOT MARKABLE]

4. *line of work* [MARKABLE]

5. *one hell of a X* [NOT MARKABLE]

Figure 3: A Sample of Idioms Found in the Corpus

criterion for annotators to use to distinguish those items we mark from those we do not. It prevents them from having to decide exactly what an idiom is – a task that is worthy of a separate annotation project.[10]

Consider the sample idioms in figure 3. Arguments of the literal sense of *butt* are meronymic (As per cf. section 4.13, we view this as a subtype of partitive) in nature (*the butt of a rifle*, cigarette butt, etc.). However, *butt of a joke* has a special idiosyncratic meaning. Therefore, this sense of *butt* is unmarkable by criterion 2. In contrast, the idiomatic phrase *with a grain of salt* assigns a metaphoric interpretation to the NP *a grain of salt*, the latter which under its most literal interpretation is perfectly interpretable, and would be markable just like other partitive phrases, including *a gram of salt*. For this reason, we assume that *a grain of salt* is markable.[11] A careful search of WordNet, www.thesaurus.com (and possibly other thesauri) suggests a contrast between the senses of *line* in *line of credit* and *line of work*. A *line of credit* is a special type of loan in which the amount borrowed is variable up to some maximum. Outside of collocations which include *line* and *credit*, neither of these words normally refer to such a specific type of loan (even though the word *credit* is a mass noun that may refer to lending power in general). Therefore, we assume that *line of credit* is unmarkable. In contrast, *line of work* is almost synonymous with "type of work". As *line* also means *type* in many other cases, e.g., *line of products*, we can reasonably assume that it has the same argument structure and is therefore markable (See section 4.17 for details about the class TYPE). *One hell of a X* is an idiom by our criteria because, in addition to *of*, it requires two specific determiners to cooccur: (1) it must be pre-

---

[10]Although job titles often are multiwords, they still usually contain one main word that determines their distribution. For example, *vice president* licenses the same arguments that *president* does. Thus *vice president* would be markable.

[11]We are not making any claims at all about the argument structure of the idiom (*with a grain of salt*) which contains this phrase.

ceded by a determiner from the set {*one*, *a*, *some*}; and (2) the determiner *a* must immediately follow the preposition *of*. It is the latter fact that is the most telling since other nouns that can only occur in the singular form take a very limited set of determiners. However, the requirement that *hell* be followed by the string *of a* makes it a clear idiom.[12] A larger list of idioms taken from the corpus is provided in Appendix A.

Metaphoric language can license arguments or adjuncts that are not possible in other sorts of text. For example, the phrase *intellectual gigolo* conveys an image which relies heavily on metaphor, e.g., what a gigolo would be like if he peddled intellect or if one is to take *gigolo* as a verb, this phrase refers to "one who gigolos intellectually". The annotator should NOT try to figure out if *intellectual* is an ARG1 or an ARGM-MNR. Rather, the phrase should be thrown out immediately as metaphoric language. Either alternative where the phrase is markable would produce a skewed analysis, e.g., it would produce noise for a statistical system. Much the same can be said about phrases like *entrepreneurial buccaneers* and *program-trading goblins*. However, one must distinguish between real metaphors and so-called "dead metaphors". A dead metaphor is an apparent metaphor that is really part of the language.[13] For example, *tapestry* is frequently used as a synonym for *combination*, and as such takes an ARG1 argument, e.g., *a tapestry of Americana*. This example is markable on the theory that the word *tapestry* now has this additional meaning and fails to invoke the image of a rug. Indeed, this is the first sense of *tapestry* found in WORDNET 1.7.1.

Another condition on the markability of an NP is that it contains a head noun. Thus the NPs in figure 4 are not markable, even though they are similar to some of the classes of NPs that we are marking, e.g., partitives. There are a number of important rationales for not marking these: (1) They are somewhat more regular and can be detected automatically quite easily, e.g., *some of* always begins a partitive phrase, whereas noun heads are less reliable in this way (there is a greater need for corpus annotation); and (2) Some of the arguments in these phrases would more properly be classified in the context of the argument structure of other parts of speech including quantifiers, comparatives, superlatives and adjectives.

Finally, there are NPs headed by certain classes of nouns which we claim are always unmarkable. These classes are exemplified in figures 5 through 6. What we are calling places of activity (figure 5) are places (real or virtual) where some action tends to happen, e.g., the stock market is a virtual place where stocks are traded.[14]

---

[12]The same analysis is assumed for *one/a heck of a*.

[13]The theory is that a phrase that is historically a metaphor has been lexicalized over time. However, we only consider usage and not etymological evidence when determining if a metaphor is dead.

[14]*Factory* is very similar to this class, but we feel that it is markable. For each case of *An X Factory*, X is the thing manufactured in the factory. The same analysis is given to *An X machine*.

1. *some of those workers and managers*

2. *two of the world's most powerful financial companies*

3. *100 of the 2,809 people*

4. *the first of five small fields*

5. *the biggest of those factors*

6. *less of the liability.*

Figure 4: NPs without Head Nouns

1. *stock market*

2. *shopping mall*

3. *train museum*

4. *massage parlor*

5. *apple orchard*

Figure 5: Places of Activity

1. *ardent anti-abortionist*

2. *TV evangelist*

Figure 6: Ideologists

1. *The biggest factor of them all*

2. *The smallest sandwich in the world*

3. *A bigger sandwich than I can eat*

4. *A likely place to visit*

5. *An easy book to read*

6. *Too many books to read*

7. *Not enough books to read*

8. *So many books that I don't want to look at them*

Figure 7: Arguments of Modifiers

The second class of nouns are people who believe in some ideology – these nouns can be modified by: the medium in which they communicate about their beliefs, the firmness of their beliefs, etc. Our experience has been that annotators disagree on the markability of these cases. We list them here to create greater consistency.

3.3 Arguments of Modifiers are Unmarkable

Comparative, superlative and adjective argument structure also effects our analysis of NPs with noun heads, as do arguments of degree words like *too*. Interestingly, when these words left-modify head nouns or (in some cases) are part of left modifiers, their arguments can occur after the head noun. All such non-noun arguments are unmarkable. Examples are provided in figure 7.[15]

3.4 When are Noun Modifiers Unmarkable

As mentioned above, not all constituents of markable NPs are markable. In particular, most noun-specific constituents are unmarkable, such as relative clauses, reduced relatives, most determiners (negatives and some possessives are markable) and a subset of the other prenominal (adjective and noun) modifiers. The rationale for not marking these is simple. We are only annotating argument-taking and propositional nouns. Since noun-specific modifiers occur with all nouns, their clas-

---

[15]As these examples don't have any markable arguments of the head nouns, the NPs themselves would not be annotated.

sification would be part of a task involving annotation of all NPs, not just a subset. Thus it seems reasonable to relegate such annotation to a separate task.

There is thus a major difference between the NomBank task and the PropBank task where virtually all constituents of the sentence were annotated for a given matrix verb. The remaining subsections of section 3 select the types of noun modifiers with the largest gray areas of markable/unmarkable and provides criteria for identifying unmarkable phrases. There are a number of reasons for pointing out what is not markable, the most important being to provide clarification for the annotator so they do not try to shove each phrase into some category. Secondly, however, it keeps us honest. By outlining what sort of phrases we are assuming should not be marked, we are exposing our guidelines to a closer inspection and critique. We hope that this helps annotators and other readers find inconsistencies (if they exist) and help us eliminate them.

3.5 Forbidden *of* Phrases

This subsection gives examples of nouns containing *of* phrases that are not markable.[16]

The examples in Figure 8 are examples of what we call the equative *of*. Basically, these *of* constructions function similarly to appositives: *X of Y* entails that *Y* is *X* or *Y* is an instance of *X*. For example the NP *the city of Paris* is equated with *Paris*. While some of the classes in later sections allow an equative *of*, few do not also take other arguments. A subset of the partitives including *set* would appear to have this property, however, unlike the equative cases, the partitive words have less semantic content. In the equative, *the city of Paris*, *Paris* is an instance of a *city*, a word with a fixed semantic class. In contrast, partitives are "transparent" as discussed in section 2. Thus *set*'s semantic class depends on the object of the preposition *of*, e.g., *a set of people* has the same class as *people*, *a set of toasters* has the same class as *toasters*, etc. Partitives are also transparent to "support" as discussed in section 5.4, thus *John* is the subject of *attack* in both *John launched an attack* and *John launched a large set of attacks*.

All ISSUE (section 4.20) and FIELD (section 4.21) nouns may also be thought of as equative. However, it would seem to us that outside of the NPs *the issue of abortion* and *the science of sleep*, one would think of very different senses of the words *abortion* and *sleep*. These sort of NPs force their arguments to be viewed as subject matter or fields of interest. In contrast, the heads of the NPs in figure 8 do not force unusual interpretations on the objects of *of*.

---

[16]The purpose of subdividing these unmarkables is to provide a full specification. While a particular example may fit multiple unmarkable classes, once it has been identified with any unmarkable class, it can be safely eliminated without further study.

1. *the city of Paris*

2. *the island of Borneo*

3. *a game of chicken*

4. *the month of October*

5. *the hulk of Stalinism*

6. *the old standby of retribution*

Figure 8: Examples of Equative *of*

1. *the people of Utah* [LOCATIVE-like]

2. *the high-yield fund of IDS Financial Services* [PRODUCT of a COMPANY]

3. *the expensive clothes of a successful trader* [PERSONAL POSSESSION]

Figure 9: Examples of Ownership *of*

1. *people of all viewpoints*

2. *the TV of tomorrow*

3. *a person of integrity*

4. *wall of death*

Figure 10: Examples of Modification by *of* + Characteristic

1. *my (household) products* [UNMARKABLE]
   ACME's (household) products [MARKABLE]

2. *her attic* [UNMARKABLE]
   the house's attic [MARKABLE]

3. *Mary's adapter* [UNMARKABLE]
   The plug's adapter [MARKABLE]

4. *John's cage* [UNMARKABLE]
   the bird's cage [MARKABLE]

Figure 11: Comparison of unmarkable and markable Possessives

Figure 9 show a variety of (unmarkable) ownership *of* phrases, where "ownership" is defined very broadly. Some of these have a locative-like meaning, e.g., *the people of Utah* or *Utah's people* are those people who live in Utah or are citizens of Utah (they belong to Utah). In contrast, *The high-yield fund of IDS Financial Services* represents a product of a company and *the expensive clothes of a successful trader* represents clothing that a trader owns.

Another set of unmarkable *of* phrases include those of the form X plus *of* plus some attribute, as exemplified in Figure 10. In each case, *of* characterizes the head in terms of some attribute. As virtually any noun can be characterized with an attribute, these are unmarkable. Note that in examples (6) and (7) of figure 9, the *characteristic* is in subject position, whereas in figure 10 the *characteristic* is in the *of* phrase.

3.6   Forbidden Prenominal Modifiers

While possessives can be markable, figure 11 gives some clear examples of unmarkable possessives. Interestingly, in a small sample of argument taking nouns with possessives, the possessive is usually an argument. However, possessives are not markable when there is a mismatch. For example, *ACME's household products* is interpretable as a case where *ACME* is an ARG0 of *product* (ACME produces something). In contrast, in *My household products*, *my* is unlikely to be an ARG0, assuming that you know that *I* am not a manufacturer. Similarly, An *attic* is assumed to be a meronymic partitive when it cooccurs with a possessive like *house*, but not when it cooccurs with something that it cannot be a part of.

Some NPs with unmarkable noun and adjectival prenominal modifiers are given in figure 12. Making this determination may be deceptive at times. For example,

1. *one major accomplishment*

2. *black urban leaders such as Charles Rangel, Basil Paterson and Mr. Sutton*

3. *a tall, Scottish-born hitter*

Figure 12: Examples of Unmarkable Prenominals

1. *fast computer* [Modifies Characteristic]

2. *better fishermen* [Probably Modifies Characteristic]

3. *beautiful dancer* [Ambiguous]

Figure 13: Modifiers of Characteristics

major sounds a lot like a MEASURE ARGM-MNR (section 11.10). However, we claim that it does not actually modify the extent of the accomplishment, but rather its importance. It is unmarkable because there is no similar concept for the verb. Similarly, *black* and *urban* say something about the *leaders*' heritage, background and appearance in example (2), but say nothing about the argument taking part of the meaning of *leader*, i.e., they modify the "human" part of the sense of *leader*, not the *leading* part. This can be seen more clearly by comparing example (2) to the instances in figure 13. A *fast computer* is one that computes quickly; *better fishermen* are most likely those who fish in a more skilled manner; and *beautiful dancer* is a famous example [28] because the phrase can either refer to one who dances beautifully or one who is physically attractive. The modifiers in figure 13 are in fact markable. *Black* and/or *urban*, in figure 12 example (2), could be markable as ARG1s under the interpretation that these were leaders of black people or leaders of city people. While this is not our view, we point this out to illustrate how fuzzy this sort of phrase can get. In contrast *tall* and *Scottish-born* only have the background/appearance interpretation – as this is from a Sports article and not an article about crime, the ARG1 interpretation is ruled out (in which someone is hitting tall and Scottish-born people). Consider the following additional example: *the military assault of June 3–4*. We assume here that *military* is either an ARG0 (the attacker) or an ARGM-MNR (the instrument), choosing the latter after viewing it in the wider context (*Americans haven't forgiven China's leaders for the military assault of June 3-4*). However, this is also clearly a gray area – one could also say simply that the phrase refers to *an assault of the military variety*. Given two plausible interpretations, we favor the one in which a constituent is markable.

1. *house debate*
   REL = debate, ARG0 = house

2. *the parliamentary debate*
   REL = debate, ARG0 = parliamentary

3. *a lively debate within an industry*
   REL = debate, ARG0 = within an industry, ARGM-MNR = lively

4. *the growing debate in Washington*
   REL = debate, ARGM-LOC = in Washington, ARGM-MNR = growing

5. *the quality of debate in Washington*
   REL = debate, ARGM-LOC = in Washington

Figure 14: ARGM-LOC vs. ARG0

1. *forest fires*
   REL = fire, ARG1 = forest

2. *the nearby woods are on fire* REL = fire, ARG1 = the nearby woods

Figure 15: ARGM-LOC vs. ARG1

3.7 ARG0/ARG1 vs. Location

In the examples in figure 14, there is some ambiguity as to whether a particular argument is an ARG0 (in these cases, that means AGENT) or an ARGM-LOC (a physical or nonphysical location or environment). This ambiguity may be particularly common with INTRANS-RECIP nominalizations (a COMLEX Syntax class for verbs like *debate*, *agree*, *meet*, *kiss* and other words involving reciprocal actions. In these cases, the annotator must decide which interpretation is more informative. If the constituent in question can reasonably be interpreted as a set of participants, then ARG0 should be marked. Otherwise, ARGM-LOC should be marked. Our judgments are included for the examples in figure 14.

Similarly, for the predicate *fire*, it may be difficult to tell the difference between the location (ARGM-LOC) of the fire and the thing that is burning (ARG0). As shown in figure 15, we assume that the location-like element is an ARG1.[17]

---

[17]Note that in the second example *fire* is part of the PP *on fire* and is linked to the subject of the sentence by a copula. The markability of arguments in this environment is discussed in section 6.

3.8 Choosing Nodes

The following rules of thumb will help choose nodes to annotate when in doubt:

- Given a non-branching tree structure, always choose the highest node for marking an argument. For example, if an NP consists of a single noun, always mark the NP, not the noun.

- Always mark the actual word w as the REL, even in cases where an empty category bound to the w is closer to the other items in the proposition. Typically, this empty category is actually bound to a phrase including w and possibly other arguments of w.

4 Lexical Entries of Markable Nouns

We provide a number of dictionaries. First we will describe a simple morphological dictionary that is useful for morphological analysis, also known as stemming. The user should use this dictionary to identify the lemmatized form of the noun to be looked up in two other dictionaries created for NomBank: our frame lexicon and NOMLEX-PLUS, both of which are described in the remaining sections.

4.1 A Simple Morphological Dictionary for Nouns

We provide a morphological dictionary for nouns where each line consists of one base form (the singular form) followed by one or more possible derived forms. This dictionary currently has over 10,000 entries (including more nouns than currently covered by NomBank). Sample entries are as follows:

- accountant accountant ex-accountant accountants

- acquisition acquisition acquisitions acquistion

- beginning beginning beginnings begining beginings

- worker worker disabled-worker co-worker metal-worker disabled-workers co-workers metalworkers metal-workers workers

Like typical stemming dictionaries for English, we include plural forms of nouns. However, we also include hyphenated forms, compound forms, alternate spellings and misspellings (found in the Wall Street Journal corpus). Each line in this dictionary lists the variation of forms that can correspond to a NomBank

annotation lemmatized to the first item on that line. For example, given the entries below, *ex-accountant* would be lemmatized as *accountant* and therefore have the same roleset, as discussed in the next section. In NomBank, the misspelled word *acquistion* is lemmatized as *acquisition*; instances of *beginnings* (with 1 n) are lemmatized as *beginning* (with 2 n's); and the compound word *metalworkers* is lemmatized as *worker*.

4.2   The Form of Lexical Entries

We create two types of lexical entries for each (markable) noun N:

- For each class (nominalization, partitive, etc.), we create one NOMLEX-PLUS entry. This entry lists that class along with some information for mapping the phrases within the noun phrase to the arguments. Currently, most of these entries are created semi-automatically and, therefore, the mapping information is only approximate. However, this dictionary does include 1000 entries of the original NOMLEX [16], which are hand-coded.

- For each word, there is one FRAME entry, which lists each possible set of arguments or ROLESETs that can occur with N. Each ROLESET in a FRAME entry usually corresponds to a coarse grained semantic sense.

For example, *megawatt* has one NOMLEX-PLUS entry for its partitive use, e.g., *400 megawatts of power*:

```
(PARTITIVE   :ORTH "megawatt"
             :NOM-TYPE ((NOM-REL))
             :FEATURES ((TRANSPARENT))
             :OBJECT ((PP :PVAL ("of")))
             :SEMI-AUTOMATIC T)
```

The above entry marks that the one argument (:OBJECT) is typically realized as a prepositional phrase headed by *of* and that the head noun *megawatt* (and its plural) is transparent so that the whole phrase takes the semantic class of the ARG1 (*power* in the above example). The tag :SEMI-AUTOMATIC T indicates that this entry was generated at least partially by automatic means – the hand-coded entries (from the original NOMLEX) lack this tag.

Megawatt also has a corresponding frame entry, which includes a single roleset, which we will list below in xml format (although a lisp-like equivalent is also available):

```
<roleset id="megawatt.01" name="partitive-quant">
<roles>
<role descr="quantified" n="1"></role>
</roles>
<example>
<text>
400 megawatts of power
</text>
<rel> megawatts </rel>
<arg n="1"> of power </arg>
</example>
```
This frameset indicates that there is a single argument ARG1 which represents the argument that is quantified by *megawatt*. Note the correspondence between the :OBJECT argument in the NOMLEX-PLUS entry and the ARG1 (n=1) argument in the NomBank frame entry.

4.2.1 Some notes about NOMLEX-PLUS

The NOMLEX manual (available at http://nlp.cs.nyu.edu/nomlex/index.html) should explain most of the details about NOMLEX-PLUS.

The main differences between NOMLEX-PLUS and NOMLEX are: 1. There are some non-nominalization entries based on NOMBANK classes; 2. If a verb that is associated with a nominalization participates in an alternation, this is indicated with the features SUBJ-OBJ-ALT and SUBJ-IND-OBJ-ALT, depending on whether the object or the indirect object alternates with the verb; 3. There are a few additional NOM-TYPES including those associated with the new classes and P-OBJ for argument nominalizations that correspond to objects of prepositions; and 4. There is an attempt to add some selection restrictions to the various argument slots (beyond what was in the original NOMLEX).

The selection restrictions include the following (some of which are inherited from the original NOMLEX):

Selection Restrictions in NOMLEX-PLUS	
RESTRICTION	**Description**
ACTION	clauses, verb nominalizations, event nouns etc.
ADJCLASS	adjectives, adjective nominalizations, etc.
BENEFICIARY	humans, organizations, projects, etc.
COMMUNICATOR	humans, organizations, governments, etc.
COMPANY	organizations
DIRECTION	PPs, ADVPs, some NPs
	(forward, toward/to + NP, etc.)
FRACTION	fractions, percentages, proportions, etc.
INSTRUMENT	instrumental *with* phrases
LOCATION	locations
NHUMAN	human NPs
NTIME	time NPs
NUMBER	number
NUNIT	NPs headed by units of measure
PROPOSITION	clauses, nominalizations, etc.
RANK	ARG2s of ACTREL relational nouns
RECIPIENT	humans, organizations, etc.
REFLEXIVE	reflexive objects, e.g. perjure oneself,
	as applied to nominalizations
TIME-ADJ	temporal adjectives
TOPIC	*about* phrases, that clauses, etc.
NOT-COMMUNICATOR	negation of communicator
NOT-LOCATION	negation of location
NOT-NHUMAN	negation of nhuman
NOT-NTIME	negation of ntime
NOT-NUNIT	negation of nunit

For more information on NOMLEX-PLUS, see the accompanying documentation entitled *Those Other NomBank Dictionaries*.

4.3 Creating Lexical Entries

We create lexical entries in two ways:

- We base the lexical entry on a verb or adjective. This is typical for a nominalization that is morphologically related to that verb or adjective. However, sometimes we also base noun entries on verbs/adjectives that are either in the same semantic class (including antonyms) or have the same basic argument structure. We refer to entries that fit this latter mold as "cousins" of

nominalizations or NOMLIKEs. When the basis of our new lexical entry is a verb, we can modify existing PropBank frame entries for NomBank.

- We base the lexical entry on an adverbial function or a related adverb. In some cases, the noun combines with a preposition or other item to have an adverbial function, e.g., *in particular*, *in fact*, *for starters*. In other cases, there is a preceding and following preposition, e.g., *in place of*, *in addition to* or a more complex construction, e.g., *no matter*, *in the first place*, etc. While various adverbial functions are possible, discourse adverbials (similar to PDTB) are the most commonly marked cases. See section 4.8 for more details. Further details about these nouns are provided in NOMADV, one of the ancillary dictionaries created along with NomBank which is described further in *Those Other NomBank Dictionaries*.

- We choose from an inventory of noun classes and modify (if necessary) default lexical entries created for these classes. The noun classes are detailed in section 4.9.

The senses defined by the rolesets (when there are multiple rolesets) are coarse-grained because we do not distinguish closely related senses of a noun unless they take incompatible sets of argument roles. The correspondence between sense and roleset breaks down for certain nominalizations that can occur both as normal verbal nominalizations and argument nominalizations. For example, *help* is a regular verbal nominalization in *She gave me some help*, but a subject nominalization in *The sales help*. The difference in meaning between these two senses reflect whether the noun refers to an action or a participant in the action, but the meaning of the arguments is the same. Therefore, these two senses can share a single roleset in which ARG0 is the HELPER, the ARG1 is the project and the ARG2 is the entity being helped. Thus *she* is the ARG0 and *me* is the ARG2 in *she gave me some help* and *sales* is the ARG1 and *help* is the ARG0 in *the sales help*.

In other instances, the same sense of a noun seems to participate with different sets of roles. For example, the canonical sense of the noun *game* refers to both an activity (the first NomBank roleset for *game*) and an art form (the third NomBank roleset for *game*). In the art form usage, a game has an author (it belongs to the work-of-art class described in section 4.15), e.g., *the Parker Brothers game about capitalism* where *Parker Brothers* is the ARG0 and *about capitalism* is the ARG1. On the other hand, it is also an activity, e.g., *her monopoly game with her brother*, where *her* is an ARG0 and *her brother* is an ARG2. These rolesets do not really correspond to distinct senses, even though they do correspond to distinct sets of arguments. In fact, these two usages can coexist in environments where the same instance of the a noun participates in two argument structures. For example, the

23

one of these usages of game is instantiated in the main clause and the other usage is instantiated in the relative clause of the following sentence: *They played the game, invented by Parker Brothers*.

It sometimes happens that multiple NOMLEX-PLUS entries correspond to a single roleset. For example, a single sense of *head* is a subject nominalization of the verb *head* and simultaneously, a relational noun. In such cases, we make attempts to maximize alignments both with other relational nouns and with the corresponding verbs. For *head* we assume an ARG0 (the noun itself), an ARG1 (the theme or subject matter) and ARG2 (the beneficiary or organization), e.g., in *the head of stock investments for Cigna Corp*, ARG0 = *head*, ARG1 = *of stock investments* and ARG2 = *for Cigna Corp*. This characterization aligns equally well with both the class of (ACTREL) relational nouns and the verb *head*. We can therefore use a single roleset as shown:

<roleset id="head.06" name="actrel/nom" source="verb-head.03">
<roles>
<role descr="job holder" n="0"></role>
<role descr="theme" n="1"></role>
<role descr="beneficiary" n="2"></role>
</roles>
<example>
<text> head of stock investments for Cigna Corp </text>
<rel> head </rel>
<arg n="0"> head </arg>
<arg n="1"> of stock investments </arg>
<arg n="2"> for Cigna Corp </arg>
</example>

Note that the xml attribute "source" on the roleset indicates if there is a corresponding verb. In addition, we may use this same "source" attribute on roles themselves if we need to use a different argument number to maintain alignment with both a NOMBANK class and a verbal roleset. For example, the ARG1 of the (DEFREL) relational noun *star* corresponds to the ARG0 of the corresponding verb and the ARG2 of the noun corresponds to the ARG1 of the verb. This is stated as follows:

<role descr="relation holder" n="0" source="1"> </role>
<role descr="relation receptor" n="1" source="2"></role>

4.4 Generalizing Over PropBank and NomBank Arguments

When designing rolesets, we make an attempt to assign argument numbers to semantic roles consistently across predicates. This is, however, a goal rather than a

constraint because figuring out what is consistent is not always easy. This section provides some details about how one may decide what is consistent.

In PropBank, arguments of verbs were classified in both verb-specific and verb-independent categories, the verb-independent categories approximating Relational Grammar (RG) of the 1970s and 1980s.[18] Arguments were numbered ARG0, ARG1, etc. In many cases (but not all), the numbered arguments are generalizations of sets of verb-specific categories, as shown in table 1. ARG0 corresponds to RG initial subject (or initial 1); ARG1 corresponds to RG initial object (or initial 2); and ARG2 includes both RG's indirect object (or initial 3) and some other categories. There are, some cases where the one to one correspondence does not hold. For example, an EXPERIENCER in PropBank is usually an ARG0 when it occurs in subject position (although the EXPERIENCER is ARG1 for the verb *experience*). In object position, an EXPERIENCER is generally marked ARG1 (*frighten*) and in prepositional object position, it is usually marked ARG2, e.g., *dawn*, *exhibit* (the latter assumes that PropBank's SEER role is a type of EXPERIENCER). Thus it is the surface position of the EXPERIENCER role that primarily predicts its ARG number.[19] In summary, the connections between RG relations and PropBank's ARG0 and ARG1 are fairly clear and straight-forward, but the connection between RG and ARG2 is less clear. For other rarer arguments (numbered ARG3 and ARG4), there is no corresponding relation in RG.[20]

Table 1 shows the degree to which conventional theta roles [6, 2, 7] generalize to both PropBank arguments and RG initial relations. While incomplete, this list should prove useful as a rule of thumb for identifying roles for a given predicate. Even the fact that INSTRUMENT, SOURCE and GOAL do not have consistent PropBank argument labels is useful to the annotator – the annotator must consult the frame for that word or a related word in such a case.[21] The degree to which it is

---

[18]Some of the ideas attributed to RG in this section occur in different forms in later theories. For example, in the Principles and Parameters family of linguistic frameworks, Williams' internal/external argument distinction [30], Larson's analysis of double object constructions [12] and Baker's UTAH principle [1] cover some of the same ground. As a package however, we believe that RG's representation of these ideas is closer to PropBank's than other theories of which we are aware.

[19]This is not surprising since both the CAUSER and the EXPERIENCER fall into the "typical subject" class. Due to the existence of verbs like *fear* (takes an EXPERIENCER subject and CAUSER object) and *frighten* (takes a CAUSE subject and EXPERIENCER object), it is hard to imagine standardizing EXPERIENCER into either ARG0 or ARG1 across verbs. Note that the object of *fear* is a CAUSE, whereas the subject of *frighten* is a CAUSER. The difference is that a CAUSER can be volitional, e.g., *John frightened Mary on purpose*.

[20]While ARG3 and ARG4 are undoubtedly less regular than ARG0, ARG1, and ARG2, there are some detectable patterns. For example, if not already covered by the first three arguments, START-POINT/SOURCE tend to be ARG3 and END-POINT/GOAL tend to be ARG4, e.g., *the price tumbled three points [ARG3 from a high of 14.5] [ARG4 to a low of 11.5]*.

[21]In Gruber's initial study [6], most of the verbs analyzed were verbs of motion. The THEME

PropBank	RG	Theta Roles
ARG0	Subject or 1	Causer, Agent, Actor
ARG1	Object or 2	Theme, Patient, Criss-Cross
ARG2	Indirect Object or 3	Recipient, Beneficiary
Varies	OBLIQUE	Instrument
Varies	OBLIQUE	Source
Varies	OBLIQUE	Goal

Table 1: A Comparison of Role Labels

possible to generalize a small set of argument types across verbs or even languages is the topic of of the Universal Alignment Hypothesis discussed in [23, 26].[22] The important point is not that a specific inventory of theta roles correspond to a particular ARG number, but rather that similar verbal/nominalization arguments tend to cluster together. Thus if a particular argument X of one predicate is very similar to a particular argument Y of another predicate, there is a high expectation that they bear the same ARG number.

Nominalizations (*cooperation, agreement, argument*, etc.) of INTRANS-RECIP verbs pose a special set of problems for the annotator, in part because the 2 (or more) participants in the activity of the predicate can be treated in a number of different ways in the PropBank frames: phrases that merge the participants into one phrase are (typically) labeled ARG0 (the subject of *John and Mary argued* would be labeled ARG0); for phrases where the participants are separated, the subject participant is marked ARG0 and the participant that follows the noun would be marked ARG2 (*John* is the ARG0 and *Mary* is the ARG2 in *John argued with Mary*). In NP structure, only one participant may appear and the annotator must still decide whether that participant should be assigned the ARG0 or ARG2 label. We assume the following convention: (1) if the sole participant is a possessive, it is an ARG0; (2) if the sole participant is plural and the object of the prepositions *between* or *among*, then it is an ARG0; (3) otherwise, if the sole participant is a

---

was the object in motion. The SOURCE was where the theme starts out and the GOAL is where it ends up, e.g., *The ball (THEME) floated from my hand (SOURCE) to the moon (GOAL).* Although the SOURCE and GOAL are separate arguments in some cases, the SOURCE and goal can simultaneously bear other roles, e.g., *John (SOURCE/AGENT) threw the ball (THEME) to Mary (GOAL/RECIPIENT)*; and *Mary (GOAL/AGENT) took the ball (THEME) from John (SOURCE).* Thus, when assigning PropBank argument types, we assume that the other roles (AGENT, RECIPIENT, etc.) may supersede SOURCE and GOAL. This is one of the reasons why SOURCE and GOAL labels may not be consistent.

[22]RG did not initially attempt to collapse the theta role categories into initial relations, but rather used other criteria (e.g., alternations) for distinguishing subjects, objects, indirect objects, etc. The Universal Alignment Hypothesis was proposed later.

1. *IBM's agreement*
   REL = agreement, ARG0 = IBM

2. *the agreement with IBM*
   REL = agreement, ARG2 = IBM

3. *John's agreement with IBM*
   REL = agreement, ARG0 = John, ARG2 = IBM

4. *the IBM agreement*
   REL = agreement, ARG2 = IBM

5. *IBM made an agreement*
   SUPPORT = made, REL = agreement, ARG0 = IBM

6. *the Apple-IBM agreement*
   REL = agreement, ARG0 = Apple-IBM

Figure 16: INTRANS-RECIP and ARG0/ARG2 conflicts

prepositional object or a prenominal noun modifier, it is an ARG2; (4) if the sole participant is outside the NP (e.g., connected by a support verb), we will only mark it ARG0 if some closer governing predicate makes the participant seem agent like; and (4) a pair of participants connected with a hyphen may occur in prenominal position – we will assume that the pair is marked as an ARG0 (the alternative would be to divide the hyphenated word into 2 constituents as discussed in section 13 and mark one as ARG0 and the other as ARG2.). Some examples are provided in figure 16.

Note that sometimes the above generalizations sometimes fail because they conflict with other generalizations. For example, *Beneficiary* roles are sometimes assigned to ARG3 for nouns which have both nominalization properties and AT-TRIBUTE noun properties (Section 4.18), because the latter entails a different ARG2 (labeled *VALUE*). For example, in *Mary's $2000 insurance coverage*, $2000 is the ARG2 (VALUE) of *coverage* and *Mary* is the ARG3 (BENEFICIARY).

4.5 Syntactic Clues for Argumenthood

Independent of roles or noun classes, certain post-nominal phrases tend to be complements. In arriving at our noun classes, we are doing our best to include all nouns that have these kinds of complements: (1) all sentential complements, including

full *that* clauses and WH clauses, *of* plus *-ing* clauses, etc.;[23] (2) PPs headed by *about*, *regarding*, *concerning*, and *on* when referring to the subject matter, e.g., *a book about chickens*, *a law concerning proper care of pets*, etc.; and (3) PPs headed by *to* or *from* indicating direction, e.g., *the road to Chicago*, *the path from Chicago*, etc..

4.6 Nominalizations and Their Cousins

For the annotation of NomBank, we begin with the rich set of verb frames created during PropBank and attempt to generalize these to nouns, sometimes with slight modifications. For nominalizations of verbs, this process is straight-forward. For example, consider *Motorola has engaged in fraudulent and inequitable conduct in the procurement of certain Motorola patents*. Prior to this example, only the *conduct an activity* sense of *conduct* was found (e.g., *she conducted a test*) in the PropBank framefile for *conduct*. Our new example, allows us to add a sense to our noun frame file corresponding to the verbal use *conduct oneself*, which does not seem to appear in the corpus. Furthermore, we can model the resulting additional roleset on an existing roleset for the verb *behave*. Thus the above instance would be annotated as follows:

Support = engaged + in, REL = conduct, ARG0 = Motorola,
ARG1 = fraudulent and inequitable

We can also add this sense of *conduct* to the verb frame file for examples like *Little Max conducted himself very well.*, assuming the analysis:

REL = conducted, ARG0 = Little Max, ARG1 = very well

and leaving *himself* unmarked.

Finding verbs in the same class can be achieved by searching through classes in COMLEX Syntax manual [31], WORDNET, www.thesaurus.com, books such as [13], etc. In this case, *behave* and *conduct* are synonyms in WordNet.

Our analysis of argument nominalizations is in the spirit of maintaining an alignment between noun and verbal (and adjectival) argument structure. For example, *the sandwich eater* would be given the analysis:

REL = eater, ARG0 = eater, ARG1 = sandwich

Although the ARG0 argument is "incorporated" into the noun itself, we still mark it. This has significance for future coreference projects. Because if for example, *John Smith* is coreferential with *the sandwich eater*, we can deduce that John Smith ate a sandwich. Other argument nominalizations are treated much the same way, e.g., the noun *nominee* would be marked ARG1, the noun *lessee* would be marked ARG2, etc. It depends on how the noun can be made compatible with the argument

---

[23]These clauses should, of course, be specific to the head noun, i.e., they should not be confused with relative clauses, purpose clauses, subordinate conjunction clauses or other adjunct clauses.

structure of the corresponding verb. However, by far, subject (or ARG0) nominal-
izations are the most common of the argument nominalizations. Furthermore, it
appears that there are some cases in which the argument nominalization status is
optional. For example, the ARG1 of *gift* may appear as an *of* phrase in *the gift of
a book* or alternatively, the whole phrase may act like the ARG1, e.g., in *The book
was Fred's gift.* or *Mary looked at Fred's gift. It was a book. John's gift* would
thus get the following analysis: REL = gift, ARG1-REF = gift, ARG0 = John. The
-REF suggests that *gift* fills the ARG1 role for referential purposes (coreference,
predication, etc.). However, it is still possible for the ARG1 to appear within the
NP headed by *gift*. The uses of -REF are discussed in detail in section 12.

We have found that in most cases, a noun with complex argument structure
has a similar roleset to some verb. This is true even of such nouns that are not
morphologically related to any verb. For example, the noun *consensus* has simi-
lar argument structure to a nominalization of *agree* (like *agreement*) and *bellwether*
acts like a subject nominalization of *predict* (like *predictor*). Thus, in the phrase *the
group's consensus*, *the group* is assigned an ARG0 (similar to *the group's agree-
ment*). In the phrase *the bellwether of privatization privatization* is assigned ARG1
and *bellwether* is assigned the role ARG0, a similar analysis as would be given to
the predictor of privatization. Like a subject nominalization, *bellwether* incorpo-
rates its subject argument (ARG0) – thus *bellwether* is treated as a predicate, as
well as one of the arguments of that predicate.

There are also cases where a particular noun, e.g., *complaint* has both a sense
related to the associated verb (*complain*) and a sense similar to an unrelated verb
(*sue*), as exemplified in figure 17. In this case, we must use rolesets based on both
verbs. *Complaint* in the latter sense can occur with the support verb *file*, which
PropBank has assigned a frame similar to that of *sue*, and which also takes as argu-
ments NPs headed by words like *complaint*, *claim*, etc. These same arguments can
occur without the support verb, although they typically don't all occur at the same
time, e.g., *the association's complaint against the pasta maker*, *the company's po-
lice complaint*, etc.

Of the nominalization cousins, there is one interesting set which we call ABLE-
NOM, e.g., *affordability*. These take mostly the same arguments as the correspond-
ing verb (*afford*), but have a slightly different meaning that can be paraphrased us-
ing the noun *ability*. Furthermore, these nouns usually include one of the following
suffixes: *-ability*, *ibility*, *able*, *ism* or *ness*. For example, in *reduced affordability of
homes for first-time buyers*, the phrases *for first-time buyers* and *homes* are respec-
tively realizations of ARG0 and ARG1. However, the NP relates to *the ability of
first-time buyers to afford homes*, rather than first-time buyers affording homes.

Lastly, there are a small set of environments in which the -ing form of a verb
is arguably used as a noun, e.g., *the eating of the grapes*, *John's eating of the*

29

1. *The Israeli Manufacturers' Association filed a police complaint against an Arab pasta maker for using the four colors of the outlawed Palestinian flag on spaghetti packages*
   Support = filed, REL = complaint, ARG0 = The Israeli Manufacturers' Association, ARG1 = for using . . ., ARG2 = police, ARG3 = against an Arab pasta maker

2. *the association's complaint against the pasta maker*
   REL = complaint, ARG0 = the association, ARG3 = against the pasta maker

3. *the company's police complaint* REL = complaint, ARG0 = the company, ARG2 = police

Figure 17: The *sue* sense of *complaint*

grapes, *Her eating grapes*. In these cases, we mostly rely on PTB's part of speech marking to determine when the -ing form is actually a noun. We treat these -ing forms as nominalizations of the corresponding verb (these are marked NOMING in NOMLEX-PLUS).[24]

Thus for many nouns, it is possible to: (1) use or extend current verb frames to number the arguments or (2) model a new set of frames based on existing ones.

4.7 Nominalized Adjectives

Some nouns are related to adjectives in much the same way as others are related to verbs. We classify these nominalized adjectives according to the guidelines discussed in this section. As we will discuss further, the classification of nominalized adjective in NomBank is influenced by other classes that that noun may belong to. For example, *envy* is simultaneously a nominalization of the verb *envy* and the adjective *envious* and *accuracy* is simultaneously a nominalization of the adjective *accurate* and the ATTRIBUTE noun class (cf. section 4.18). Ignoring such considerations, we employ the following heuristics to create rolesets for nominalized adjectives. We assume that adjectives in English have either one argument or two: the surface subject and possibly one surface complement. When the surface subject falls into one of the typical ARG0 theta role classes, it is marked as an ARG0. Otherwise it is typically an ARG1. The second argument, if it exists, is conditioned on the first: the second argument is ARG1 if the first is ARG0; the second is

---

[24]There are rare cases where the Treebank is clearly in error in which case we do not mark the instance, e.g., *thrashing* is marked as an NN that heads a VP in . . . *have been thrashing Columbia at the box office* (wsj_1634).

ARG2 if the first is ARG1. Clausal complements are typically ARG1 (just as they are with verbs). Thus (initially) there are 4 possible argument structures (rolesets) for adjectives (and thus for adjectival nominalizations): {ARG0, ARG1}, {ARG1, ARG2}, {ARG0} and {ARG1}. Some example nominalized adjectives are provided in figure 18. Observe that some nominalized adjectives incorporate one of the arguments, e.g., *conservative*, and *expert*. These nouns are very similar to argument nominalizations (*teacher*, *nominee*, etc.). They typically take one additional argument (ARG0 or ARG1, depending on which argument is incorporated). Also note that the adjective *ability* is the basis for the class of nouns described in section 4.14 and that there is a large set of nouns that are assumed to be "cousins" of the adjectival nominalization of *expert*, all of which incorporate the ARG0 (subject) role: *alchemist, buff, champ, champion, czar, expert, genius, guru, magnate, mogul, panjandrum, powerhouse, pro, prodigy, professional, pundit, purist, repository, strategist, titan, tycoon, whiz, wizard*.

Membership in other classes has an impact on the set of roles assumed as shown in figure 19. If a noun is simultaneously a nominalization of an adjective and a nominalization of verb, we make sure that the roles can be mapped to the corresponding PropBank entry. Figure 19, example 1 is an instance of *envy*, which is simultaneously a nominalization of the verb *envy* and of the adjective *envious*. The verb has three roles: ARG0 (the one who is envious), ARG1 (the one who they are envious of) and ARG2 (the reason for their envy). We repeat these roles in this example in which the ARG1 is incorporated into the word *envy* itself (which is then associated with the NP *the casino* by predication across the copula-like verb *became*). Other adjective nominalizations that are also verbal nominalizations include: *absence*, *anger*, *challenge* and *equivalent*. *Distance* is a nominalization of *distant* as well as a member of the ATTRIBUTE noun class (cf. section 4.18), the latter which includes a measurement or VALUE as an ARG2, in example 2, this ARG2 surfaces as the adjective *far*. As it turns out, a large number of adjective nominalizations also belong to the ATTRIBUTE class, e.g., *accuracy*, *allure*, *appropriateness*, *beauty*, *bitterness*, etc. In the third example of figure 19, *wizard* is a nomadjlike of *expert* as well as a relational noun (section 4.10) of the ACTREL variety, the latter classification adding an ARG2 representing the beneficiary or employer of the *wizard*.

4.8 Nominalizations of Adverbs and Nominal Anchors of Adverbial Constructions

NOMADV is one of the dictionaries described in detail in the accompanying document *Those Other NomBank Dictionaries*. It specifies relations between nouns and adverbs. There are two main types of entries: NOMADV entries which map

1. *her ability to produce higher student-test scores*
   REL = ability, ARG0 = her, ARG1 = to produce higher student-test scores

2. *the absence of patent lawyers on the court*
   REL = absence, ARG1 = of patent lawyers, ARGM-LOC = on the court

3. *order accuracy*
   REL = accuracy, ARG1 = order

4. *our long-term ambition of running a major entertainment company*
   REL = ambition, ARG0 = our, ARG1 = of running a major entertainment company, ARGM-TMP = long-term

5. *the beauty of a democracy*
   REL = beauty, ARG1 = democracy

6. *offensive capability*
   REL = capability, ARG1 = offensive

7. *flexibility in regulating pesticides*
   REL = flexibility, ARG1 = in regulating pesticides

8. *the vulnerability of many small communities to domineering judges*
   REL = vulnerability, ARG1 = of many small communities, ARG2 = to domineering judges

9. *An exorcism expert*
   REL = expert, ARG0 = expert, ARG1 = exorcism

10. *fiscal conservatives*
    REL = convervatives, ARG0 = conservative, ARG1 = fiscal

Figure 18: Sample Annotations of Nominalized Adjectives

1. *The casino became the envy of its competitors for its money-making ability*
   [NOMINALIZATION of verb *envy*]
   REL = envy, ARG1-REF = envy, ARG0 = of its competitors, ARG2 = for its
   money-making ability

2. *The Sun's far distance from Pluto* [ATTRIBUTE noun]
   REL = distance, ARG1 = The Sun's, ARG2 = far, ARG3 = from Pluto

3. *high-tech computer wizards at the major brokerage firms* [RELATIONAL
   noun]
   REL = wizards, ARG0 = wizards, ARG1 = high-tech computer, ARG2 = at
   major brokerage firms

Figure 19: Adjective Nominalizations that also belong to other classes

nouns to morphologically related adverbs and NOMADVLIKE entries which map
nouns to adverbs of approximately the same meanings. As with the ADJADV
dictionary, these entries include COMLEX Syntax adverb classes to classify the
type of relation between these nouns and their arguments (or modifiees). Some of
these entries correspond to frame entries with names that include any of the fol-
lowing substrings: *adverbial*, *discourse* or *NOMADV*.[25] Figure 20 provides exam-
ples where the noun is part of an adverbial construction (typically a prepositional
phrase) that: (1) modifies a clause as in examples 1 and 2[26]; (2) acts like a focus
modifier as in examples 3 and 4 (similar in function to the adverbs *just* and *only*)
– we assume that these modify the phrase or word immediately following the ad-
verbial expression, e.g., *sort of*[27]; or (3) joins two arguments together along the
same lines as the discourse connectives marked in PDTB. This third case is like the
PDTB approach in that it can link together two clauses within the same sentence
as in example 5. However, NomBank discourse adverbials differ from the PDTB
ones in two respects: arguments can be noun phrases as in example 6; and argu-
ment links with previous sentences are omitted as in example 7 (one can assume

---

[25]Some of the entries in NOMADV do not correspond to any frame entries at all, but rather corre-
spond to potential instances of adverbial modification discussed in section 10, e.g., in *the possibility
of an attack*, *possibility* is the ARGM-ADV of *attack*.

[26]We assume that these always modify clauses, ignoring the distinction between VP and sentential
modifiers assumed in many syntactic theories. However, the NOMADV entries can distinguish these
because META-ADV features correspond to sentential modification.

[27]When the focus modifiers modify verbs, we assume that the whole clause is the ARG1 (the
modifiee), i.e., we assume that it is an instance of case 1 above. For other parts of speech, we assume
that the following word or phrase is the argument.

1. *any fuel in particular*
   REL = particular, ARG1 = any fuel

2. *In particular, John looks like an emu*
   REL = particular, ARG1 = John looks like an emu

3. *It's kind of annoying*
   REL = kind, ARG1 = annoying

4. *It seems sort of draconian*
   REL = sort, ARG1 = draconian

5. *In case you have forgotten, his name was Rudolph Dirks.*
   REL = case, ARG1 = his name was Rudolph Dirks, ARG2 = you have forgotten

6. *That was in addition to $34,000 in direct campaign donations*
   REL = addition, ARG1 = That, ARG2 = to $34,000 in direct campaign donations

7. *By contrast, traditional investors are unlikely to generate sudden price moves*
   REL = contrast, ARG2 = traditional investors are unlikely to generate sudden price moves

Figure 20: Arguments of Nouns Anchoring Adverbial Constructions

that the underlying ARG1 would be *program trading* given the full context).

As with other NomBank classes, the same roleset may apply to more than one syntactic environment. For example, *contrast* uses the same roleset regardless of whether all the arguments are NP-internal (*the color of the desk's contrast with the color of the chair*) or the arguments are in a discourse context (*by contrast*) as in figure 20, example 7, or *in contrast with* in *He bought a Ferrari in contrast with his sister who bought a Rolls Royce*.

These type of constructions are discussed further in section 10.

4.9 Other Classes of Argument Taking Nouns

In this section, we define 16 classes of argument taking nouns that cover almost all the remaining cases. Figure 21 and 22 provide a sampling of each class (and subclass). Our general strategy for choosing argument numbers for each class of

nouns is as follows: (1) find a representative set of nouns that are members of a class as well as nominalizations; and (2) generalize the information from those frame files to the rolesets for that class. This strategy attempts to maximize the overlap between alternative frames for each noun. We need to do this in order to avoid situations in which different frames for (a single sense of) a noun assign different argument structures. This goal has been stated above in different forms and is repeated throughout these specifications.

4.10 RELATIONAL Nouns

A relational noun A is a noun that takes at least one argument B, such that there is an implied relation R between A and B. A convenient way of thinking about these is that they are like subject (ARG0) nominalizations with A playing the subject role, R being the underlying predicate and B being some other argument of R. In fact, many RELATIONAL nouns are also subject nominalizations, e.g., *leader*, *director*, *teacher*.

We further define R to include only such relations between combinations of two entities that are either people, organizations, government-bearing locations (countries, provinces, cities, etc.)[28], projects (movies, research efforts) or vessels including ships and airplanes. R can be in the domains of kinship, social, employment, representation, and many others.

Relational nouns typically can take their arguments in a number of different positions: as PPs (*a director of the finance department*), possessives (*the finance department's director*), noun noun modifier position (*the U.S. Army Pilot*, *a finance director*) and adjectivally *a financial director*.

As shown in the examples in figure 23, we recognize two subtypes of relational nouns: DEFinitional relational nouns (DefRel) and ACTion relational nouns (ActRel). The crucial differences are: (1) DefRel nouns (e.g., *father*, *capital*, *suburb*, *protagonist*) only take one argument (ARG1); and (2) ActRel nouns (e.g., *lawyer*, *president*, *director*) take at least one argument (ARG2), and in some cases an additional argument (ARG1). It may seem that there is an arbitrary distinction between the ARG2s of ActRels with only one argument and the ARG1s of the DefRels. However, we define ARG1 and ARG2 in terms of the typical ARG1s and ARG2s of section 4.4. In other words, a RELATIONAL noun is an ACTREL if it takes an ARG2 and a DEFREL otherwise. We assume that employers and beneficiaries are ARG2s and most other roles are ARG1 (e.g., patient, theme, subject matter, etc.).

Beyond the presence or absence of ARG2, there are tendencies of nouns in certain semantic classes to fit either the DefRel or ActRel categories. For DefRel

---

[28]This subset of locations corresponds to the Automatic Content Extraction (ACE) program's GPE class.

RELATIONAL (ACTREL): *the vice president of North America operations*
  REL = president, ARG0 = president, ARG2 = of North America operations, ARG3 = vice

RELATIONAL (DEFREL): *Mr. Engelken's sister*
  REL = sister, ARG0 = sister, ARG1 = Mr. Engelken's

JOB: *the new Treasury post of inspector general*
  REL = post, ARG0 = of inspector general, ARG2 = Treasury

HALLMARK: *the cornerstone of Phillips' chemicals operations*
  REL = cornerstone, ARG0 = cornerstone, ARG1 = Phillips' chemicals operations

PARTITIVE: *dozens of attorneys who received letters*
  REL = dozens, ARG1 = of attorneys who received letters

PARTITIVE/PIECE: *the back of your hand*
  REL = back, ARG1 = of your hand

PARTITIVE/MERONYM: *the president's head*
  REL = head, ARG1 = the president

SHARE: *each company's share of liability*
  REL = share, ARG0 = each company's, ARG1 = liability

GROUP: *an army of judicial activists*
  REL = army, ARG1 = of judicial activists

ENVIRONMENT: *a period of industry consolidation*
  REL = period, ARG1 = of industry consolidation

ABILITY: *the absolute right of everyone to disseminate materials*
  REL = right, ARG0 = of everyone, ARG1 = to disseminate materials, ARGM-MNR = absolute

Figure 21: Sample Annotation for Noun Classes 1 – 8

WORK-OF-ART: *Congress's idea of reform*
    REL = idea, ARG0 = Congress's, ARG1 = of reform

VERSION: *the House version of the deficit-cutting bill*
    REL = version, ARG0 = House, ARG1 = of the deficit-cutting bill

TYPE: *his unique brand of Christianity*
    REL = brand, ARG1 = of Christianity

ATTRIBUTE: *the breadth of inquiries*
    REL = breadth, ARG1 = of inquiries, ARG2-REF = breadth

ISSUE: *the subject of U.S. direct investment in Japan*
    REL = subject, ARG1 = of U.S. direct investment in Japan

FIELD: *the rapidly growing field of bio-analytical instrumentation*
    REL = field, ARG1 = of bio-analytical instrumentation REL = subject, ARG1 = of U.S. direct investment in Japan

CRISS-CROSS: *the victim of an assassination*
    REL = victim, ARG1 = of an assassination

EVENT: *the drought of 1988*
    REL = drought, ARGM-TMP = of 1988

Figure 22: Sample Annotation for Noun Classes 9 – 16

1. *Mindy Hymowitz's mother* [DEFREL]
   REL = mother, ARG0 = mother, ARG1 = Mindy Hymowitz's

2. *the nation's capital* [DEFREL]
   REL = capital, ARG0 = capital, ARG1 = nation's

3. *Mr. Noriega's friend* [DEFREL]
   REL = friend, ARG0 = friend, ARG2 = Mr. Noriega's

4. *a McDonald's customer* [ACTREL]
   REL = customer, ARG0 = customer, ARG2 = McDonald's

5. *movie director* [ACTREL]
   REL = director, ARG0 = director, ARG1 = movie

6. *a nonexecutive director of this British conglomerate* [ACTREL]
   REL = director, ARG0 = director, ARG2 = of this British conglomerate

Figure 23: Examples of Relational Nouns

nouns, R tends to hold between the ARG0 and ARG1 independently of any action undertaken by the ARG0, e.g., in *Mary's sister*, the relation between *Mary* and her sister exists even if they don't know that they are sisters. The name DefRel comes from the assumption that these relations are stative – they exist by definition. On the other hand for ActRel nouns, A typically serve some function or does some job in order for R to hold. The name ActRel comes from the idea that certain actions must be performed for the relation to hold. Differentiating ARG1s and ARG2s for ACTRELs is just like differentiating direct and indirect objects. Thus taking *teacher* as a model (which is simultaneously an ACTREL and a subject nominalization), the subject taught is the ARG1 and the person taught is the ARG2. It is easiest to identify ARG1s for ActRel nouns that are also nominalizations. However, we tentatively assume that ActRel nouns exert direct control over their ARG1 arguments. Thus a director has direct control over a movie, but merely works for his or her employer. Thus *MGM* would be an ARG2, whereas *movie* would be ARG1 in *A movie director for MGM*.

There are many types of relational nouns as exemplified in figure 23. In some cases, the relational noun represents a part of one of its arguments (e.g., a president of a club is usually a member), but it need not be (e.g., my friend is not a part of me). Furthermore, when an argument can be an organization, country or other entity made up of a set of people, a single member can play the role of argument.

Thus, one can refer to George W. Bush as both *your president* (ARG2 = *your*) and *president of the United States* (ARG2 = *of the United States*).

Relational nouns often overlap with argument nominalizations, e.g., *father*, *teacher*, *leader*, *director* and *captain* are all subject nominalizations, as well as relational nouns. It is possible that the role assignments dictated by the frames for the verb and the frames for relational nouns will conflict and we allow for that, by including mapping information in the lexical entry. For example, if the ARG2 of the relational noun entry corresponds to the ARG1 of the verb entry, we mark this argument as the ARG2 and indicate that this maps to the ARG1 of the corresponding verb using the "source" attribute, as discussed in section 4.3. For example, for the relevant sense of *the governor of Texas*, *of Texas* is marked ARG2 following the ACTREL type frame. However, the lexical entry also states that the verbal source is *govern* and the source of the ARG2 role is ARG1. Thus instances of *governor* will pattern both with other ACTREL nouns and with uses of the verb *govern*, although in the latter case one will have to associate the ARG2 relations of *governor* with the ARG1 relations of *govern*.

In other cases, we have determined that one possible classification overrides the other. For example, when an annotator encountered *ambassador*, she realized that it could either be viewed as a relational noun or a nominalization cousin of the subject variety, modeled on the verb *represent*. Ultimately, we decided that the roles should be classified in terms of the nominalization, because they are different than our standard ActRel frame. In particular, *ambassador* has an additional argument (the *to* phrase in *France's ambassador to China*) (cf. section 4.24).

For the DefRel noun *father*, the verb just happens to assign the ARG0 role to the father and the ARG1 role to the child – thus no conflict exists. Likewise the ACTREL entry and the nominalization entry for *teacher* lead to the same set of arguments. In *John's teacher*, *John's* is the indirect object (ARG2) of the verb *teach* as well as argument ARG2 for the ActRel relational noun entry. For an example like *Young McDuffie's first violin teacher*, *violin* is assigned the ARG1 role by the verbal entry and the other two roles are compatible with both the verb entry and the ActRel noun entry. For *director* and *captain* there are multiple senses: (1) the rank in a corporation or military organization; and (2) the one in charge of a project or vessel. The first sense does not correspond to the verbal forms at all – it suggests an employee. Thus the phrases *the U.S. army captain* and *the company's executive director* would be covered only by the relational noun interpretation and *U.S. army* and *the company's* would be ARG2s. In contrast, *the captain of the Santa Maria* and *the movie director* would be covered by both interpretations. Either way, *the Santa Maria* and *movie* would be assigned ARG1.

Most relational noun instances include exactly one argument (other than the relational noun itself). However, consider the following example: *Compaq's vice*

president of North America operations. Based on the preceding prose, it would be unclear whether to mark *Compaq* (the parent company) or *North American Operations* (a subdivision) as the ARG2. The answer is very simple: both should be marked ARG2. While this defies the one constituent/one argument number convention, we believe that a reasonable analysis would combine the two ARG2s together – a paraphrase might be *The vice president of Compaq's North American Operations*. Now let's look at a similar example: *Grand Met's first group finance director*. In this case, one might be tempted to mark three constituents as ARG2 (*Grand Met's*, *group*, and *finance*). However, *finance* refers to the area controlled by the director, rather than an actual finance department. On the analogy of *teacher*, we will mark such arguments ARG1 (the ARG1 of teacher is the subject taught just as this is the topic area that is presided over). Thus both *Grand Met's* and *group* should be ARG2, but *finance* would be ARG1. One unavoidable consequence is that *the director of the finance department* and *the finance director* will be marked differently, although they arguably have the same meaning. On the other hand, *finance directors* can be employed by companies too small to have finance departments.

Finally, I will discuss two additional arguments for some RELATIONAL nouns: ARG3 (Rank) and ARG4 (Employer), the latter limited to ACTREL nouns. There are some relational nouns that can be modified to convey slightly different relations, yet maintain their same argument structure. For example, the following table lists single words together with multiple word expressions that share the same argument structure:

president	*vice president, assistant vice president*
mayor	*deputy mayor, assistant deputy mayor*
cousin	*second cousin, second cousin once removed*
brother	*step brother, half brother, brother in law*

In each of the above cases, we mark the modifiers (*vice, assistant vice, second, once removed*, etc.) as ARG3 to represent the RANK of the relation. Thus *president* and *vice president* can share the same lexical entry, as can the other sets of items in the lists above.

The ARG4 or EMPLOYER argument occurs for a subset of ACTREL nouns, like *consultant*, *attorney*, etc. The beneficiary (ARG2) of most ACTREL nouns can be either an employer or someone else who benefits from the relation implied by the relational noun, e.g., the student of the teacher, the company of the president, etc. However, for consultant-type ACTREL nouns, there can be an employer distinct from the beneficiary (the client), e.g., *John's lawyer from Dewey, Cheatum and Howe*. Here *John* would be the ARG2 beneficiary and *Dewey, Cheatum and Howe* would be the ARG4, the EMPLOYER.

4.11 JOB nouns

JOB nouns (pronounced with a short *o*) are like relational nouns, except they do not have an incorporated ARG0. In other words, subject nominalizations are to relational nouns what regular nominalizations (non-argument nominalizations) are to JOB nouns.

For JOB nouns, the holder of the job is the ARG0, the type of job is the ARG1 and the employer is the ARG2. For example, in *the new Treasury post of inspector general*, *inspector general* is the ARG1 and *Treasury* is ARG2. In *John's new Vice President position*. *John* is the ARG0 and *Vice President* is the ARG1. Example JOB nouns include: {*job*, *position*, *post*, *role*, *seat*, *spot*, *title*, *presidency*, *mayoralty*}. There is a distinction between the first seven of these words and the last two.

Presidency and *mayoralty* specify the job titles for relation R. Thus in *John's club presidency*, *John's* is the ARG0 and *club* is ARG2 and R is a presidency relation. For these *JOB* nouns we assign the ARG1 relation to the predicate, i,e., *presidency* is its own ARG1. In contrast, in *the club title of president*, the ARG1 argument (*president*) identifies R. Other differences include *presidency* and *mayoralty* represent the relations that have time spans, e.g., *John's presidency ended last year*, whereas instances of the other five nouns represents the name of a job which can be obtained or held, e.g., *John held the club title of president*.

Other unusual *JOB* nouns include *seat* which does not have an ARG1 at all and *leadership*, which is ambiguous between an ACTREL relational noun and a *JOB* noun. Only one roleset is required since these types of nouns allow the same set of roles. In the ACTREL usage, *leadership* refers to the leaders (ARG0) themselves, (*The monster leadership made a decision*), whereas in the *JOB* usage (*Mary's leadership of the monsters*), the leader (ARG0) is represented by an NP (*Mary's*).

4.12 HALLMARK nouns

HALLMARK nouns are similar to relational nouns in that the noun itself A is assigned ARG0 and its argument B is assigned ARG1. A is a part of B that is particularly important: A is representative of B, is typical of B, represents the best of B, a particularly significant attribute of B, etc. Examples include: *backbone*, *centerpiece*, *cornerstone*, *hallmark*, *thread* and many others. There are no restrictions on the semantic class of B. We illustrate this class with the examples in figure 24. For these words, the whole sentences are necessary to put the words in context.

41

1. *Creative accounting is a hallmark of federal credit.*
   REL = hallmark, ARG0 = hallmark, ARG1 = of federal credit

2. *James Wright says homelessness is due to a complex array of problems, with the common thread of poverty*
   REL = thread, ARG0 = thread, ARG1 = poverty

3. *The idea was to let small investors, the backbone of the fund business, deal in the money market's high short-term interest rates.*
   REL = backbone, ARG0 = backbone, ARG1 = of the fund business

Figure 24: Sample HALLMARK Noun Annotation

1. *some of the TV stations* PTB analysis = (DT some)

2. *many of these funds* PTB analysis = (JJ many)

3. *enough of this kind of material* PTB analysis = (JJ enough)

4. *each of CVN's 20 million fully diluted shares*
   PTB analysis = (DT each)

5. *100 of the 2,809 people* PTB analysis = (CD 100)

6. *nearly half of Hong Kong consumers*
   PTB analysis = (QP (RB nearly) (NN half))

Figure 25: Unmarkable Partitive Constructions (headed by quantifiers)

4.13   PARTITIVE, SHARE, and GROUP Nouns

In this section we discuss PARTITIVE nouns and two of their cousins SHARE nouns and GROUP nouns. Nouns of each of these classes take a special argument B such that the whole noun phrase represents either a multiple of B, a fraction of B, a part of B, or any other possible quantification over an amount of B. B is assigned the role ARG1 on analogy of nouns that are simultaneously nominalizations on the one hand and PARTITIVE (*variety*, *cascade*), SHARE (*slice*, *share*) or GROUP nouns (*assembly*, *band*) on the other.

First we discuss the PARTITIVE class, which is itself somewhat of a hodgepodge and perhaps should be split apart at a future date. One reason not to break them apart just now is that this would introduce many gray areas that are irrele-

1. *dozens of fiber-end bunches* [PARTITIVE]
   REL = dozens, ARG1 = of fiber-end bunches

2. *tens of billions of dollars*[PARTITIVE]
   REL = tens, ARG1 = of billions of dollars

3. *jillions of dollars*[PARTITIVE]
   REL = jillions, ARG1 = of dollars

4. *a whole bunch of people*[PARTITIVE]
   REL = bunch, ARG1 = of people

5. *the first set of meetings*[PARTITIVE]
   REL = set, ARG1 = of meetings

6. *another instance of Washington's sticky fingers*[PARTITIVE]
   REL = instance, ARG1 = of Washington's sticky fingers

7. *a wide variety of crops*[PARTITIVE]
   REL = variety, ARG1 = of crops, ARGM-MNR = wide

8. *a cascade of genetic damage*[PARTITIVE]
   REL = cascade, ARG1 = of genetic damage

9. *a lot of harm*[PARTITIVE]
   REL = lot, ARG1 = of harm

10. *lots of internal debate*[PARTITIVE]
    REL = lots, ARG1 = of internal debate

Figure 26: Some Noun Partitive Examples

1. *an evil-looking cloud of black smoke*[PARTITIVE]
   REL = cloud, ARG1 = of black smoke

2. *big boxes of just-picked Red Delicious next to his barn*[PARTITIVE]
   REL = boxes, ARG1 = of just-picked Red Delicious next to his barn

3. *a pound of flesh*[PARTITIVE]
   REL = pound, ARG1 = of flesh

4. *a package of accelerated tariff cuts*[PARTITIVE]
   REL = package, ARG1 = of accelerated tariff cuts

5. *a part of the program*[PARTITIVE/PIECE]
   REL = part, ARG1 = of the program

6. *the largest single component of solid waste*[PARTITIVE/PIECE]
   REL = component, ARG1 = of solid waste

7. *the bottom of the ninth of the third game* [PARTITIVE/PIECE]
   REL = bottom, ARG1 = of the ninth of the third game

8. *the back of your hand*[PARTITIVE/PIECE]
   REL = back, ARG1 = of your hand

9. *a subsidiary of the Swiss company* [PARTITIVE/MERONYM]
   REL = subsidiary, ARG1 = of the Swiss company

10. *the brain of a Parkinson's patient* [PARTITIVE/MERONYM]
    REL = brain, ARG1 = of a Parkinson's patient

Figure 27: Some More Noun Partitive Examples

vant to the annotation task. NPs headed by PARTITIVE nouns are much like the partitive constructions that are headed by numbers and quantifiers (*several*, *one*, *most*, *more*, *many*, *less*, *few*, *enough*, *both*, *all*, *those*, *some*, *none*, *neither*, *either*, *each*, *any*, *another*)[29], although the number/quantifier partitives are not markable under our current annotation task (cf. Section 3). Due to the similarity between these two types of partitives, there is some gray area, which can be resolved by looking at PTB's part of speech assignment – we assume that a partitive phrase is markable if: (1) the argument phrase (usually an *of* phrase) is part of an NP that is headed by a noun (NN or NNS); or (2) the head is an NP and the ULTIMATE HEAD of *H* is a noun (NN or NNS), where the ULTIMATE HEAD is the lexical item obtained by taking the head of the head of the head . . ., given that the head of X may be a phrase, the head of the head of X may be a phrase . . ., e.g., *part* is the ultimate head of *(NP (NP (DT this) (NN part)) (PP (IN of) (NP (DT these) (NNS specifications))))*. We make this second provision to account for the fact that for some phenomena, heads of QPs can be nouns (figure 25 (6)) and heads of NPs can be other parts of speech (e.g., the head-less NPs discussed in cf. section 3.2). For example, number words occurring in the plural are marked NNS and are therefore markable, whereas the singular counterparts are usually not (typically such plurals represent sets, e.g., *thousands of books* means something like "multiples of a thousand books"). Example (unmarkable) quantifier partitives are given in figure 25 and a wide variety of different noun partitives are exemplified in figures 26 and 27.

Figure 26 and 27 present a variety of different types of partitives. The first several in figure 26 express quantities of whole units (*dozens*, *tens*, *jillions*, *bunch*, *set*, *instance*) with varying degrees of preciseness. *Variety* and *cascade* add other elements of meaning: variety suggests that the set contains elements that differ; *cascade* suggests that the elements appear one after the other. The next items *lot* and *lots* are compatible with both mass nouns and plurals. The first item in figure 27, *cloud* admits just a mass noun (and also conveys a certain image regarding the shape of the quantity). Next container and unit nouns are used to suggest quantities (*box*, *pound*, *packages*). The remaining cases all represent part of the ARG1, ranging from an unspecified part (*part*, *component subsidiary*), to more specific parts (*bottom*, *top*), to parts of the ARG1 that are idiosyncratic to particular ARG1s (*brain*). It is the last several examples, that perhaps raise the most eyebrows, especially *brain*. In other words, all meronym relationships expressed as NPs are viewed as partitives. Of all the types of partitives discussed in this paragraph, *brain* is the least like the quantifier examples. However, we believe that there is a gradual slope from the clearest partitives down to these, e.g., if *a person's brain*

---

[29]These are all possible non-noun heads of the partitive construction. They all have quantificational properties, although they may have different parts of speech.

is not a partitive, why should *a subsidiary of Exxon* be a partitive? It would seem necessary to include these in order to get the greatest consistency of annotation.[30]

As reported in section 2, an additional property of most partitives is that they take the semantic class of the ARG1, rather than the head. While this is a strong tendency, it is not generalize to all partitives. The generalization holds for all the examples in figure 26 and the first four examples of figure 27. For example, *tens of billions of dollars* is of type MONEY based on *dollars*, as derived by applying this property to the nested partitives – the whole phrase gets its semantic class from *billions of dollars*, which in turn derives its class from *dollars*. Similarly, *a whole bunch of people* is of class PERSON. In contrast, examples 5-8 in Figure 27 fail this test because it is often the case that a part has a different class than the whole. To distinguish these from the other partitives we will introduce the following two subclasses (PARTITIVE/PIECE)) and (PARTITIVE/MERONYM).[31] Words like *part*, *back* and *bottom* belong to the former class and words like *hand* and *brain* belong to the latter. The semantic class of the PIECE variety of meronym is based on the type of the ARG1, and depending on the ARG1, can even be the same type. For example, if ARG1 is a partitive, the NP is the same type as the partitive, e.g., *a part of a group of people* is of type PERSON. Although *a part of an airplane* is not of type AIRPLANE, its class is limited by the set of things that can be parts of airplanes, e.g., it cannot be of type PERSON. In contrast, members of the MERONYM class have fixed types, i.e., they identify a specific part of something, e.g., *the head of an animal*.[32] We will also point out the following additional points about this property of semantic transparency: (1) nouns are typically not transparent when ARG1 fills the possessive slot; (2) transparent nouns act like SUPPORT verbs as discussed in section 5; (3) transparency may be useful for distinguishing among senses of some partitives. For example, *chest* is an ordinary partitive in *chest of books*, but a meronym in the body part sense, e.g., *his chest*.

Some partitives may add a locative or temporal modifier to their ARG1, as exemplified by the following example: *7 years of bitter debate* (REL = debate, ARGM-MNR = bitter, ARGM-TMP = 7 years).
In this example, we are annotating an instance of *debate* when it occurs as the head of the ARG1 of the partitive *years*. Not only does *7 years* quantify over the amount of debate that takes place (its partitive function), but it also provides

---

[30]One consequence is that words like *city* and *province* are partitive when they occur in phrases like *a city of France* or *province of Canada*, but not in equative phrases like *the city of Paris* or *the province of Quebec*.

[31]We mark a subset of the (PARTITIVE/PIECE) nouns as belonging to the DIVISION subclass. The frame names are labeled PARTITIVE-PART/DIVISION. These nouns represent parts of organizations, e.g., *arm, bureau, chamber, chapter, district, province, region, section, ward*.

[32]*The head of an organization* is a RELATIONAL noun instance of *head*.

1. *40 acres of grapes in California*
   REL = acres, ARG1 = of grapes, ARG3 = in California

2. *5 minutes of fun on Tuesday*
   REL = minutes, ARG1 = of fun, ARG3 = on Tuesday

Figure 28: Partitives with a Secondary Theme Argument

the information that the debate takes at least 7 years (its temporal function). This (rare) feature of some partitives is similar to the main feature of criss-cross nouns (cf. section 4.22). Similarly, partitives representing units can modify prepositional phrases, e.g., in *two steps ahead of Marlowe*, the modifier *steps* takes *ahead of Marlowe* as its ARG1.

Partitives can sometimes take an additional "secondary theme" argument which we will label ARG3 – a similar argument occurs with ATTRIBUTE nouns and we would like to keep the argument number the same (thus we are skipping ARG2). Locative partitives allow a locative secondary theme argument and temporal partitives allow a temporal secondary theme argument. In both cases, the partitive phrase as a whole functions as a part of the secondary theme. Thus in the examples in figure 28, the acres are part of California and the minutes are part of Tuesday.

Group nouns[33] are different from partitives in that the group described by the group noun phrase has an identity independent of its members. Examples of group nouns include *academy*, *chorus*, *community*, *family*, *legion*, *herd* and *team*. Some examples are given in Figure 29. Note that the ARG1 of group noun predicates should be either a plural noun (members of the group) or a descriptor of the set of members (e.g., an adjective). In addition, group nouns can take an ARG2 (the employer, leader or other important individual that in some sense possesses the group) and an ARG3 (secondary theme, as discussed above).

Sample SHARE noun annotation is provided in figure 30. SHARE nouns presuppose that something is being shared or divided among more than one entity, each of which gets their portion. NPs headed by SHARE nouns (*portion*, *share*, *slice*, *stake*, *stock*), are like some partitives in that they take an argument B such that the whole NP represents only a piece of B. Unlike partitives, they admit an additional argument, an ARG0, the entity who was assigned that piece of B. For example, in *the exchange's share of the #2.5 billion marine market in London*, a number of entities are sharing the marine market and this noun phrase represents the portion alloted to *the exchange* (ARG0). Also unlike partitives, SHARE nouns can take a third argument, an ARG2, which indicates the portion (VALUE) as-

---

[33]This class corresponds approximately to the AGGREGATE class from COMLEX Syntax.

1. *a community of parents*
   REL = community, ARG1 = of parents

2. *a family of ceramic superconductors*
   REL = family, ARG1 = of ceramic superconductors

3. *a large herd of animals*
   REL = herd, ARG1 = of animals

4. *a team of researchers from the National Cancer Institute*
   REL = team, ARG1 = of researchers from the National Cancer Institute

5. *her publishing group*
   REL = group, ARG1 = publishing, ARG2 = her

6. *ACME's five-member board of directors*
   REL = board, ARG1 = directors, ARG2 = ACME's, ARG3 = five-member

Figure 29: A Sample of GROUP Noun Phrases

signed to the ARG0, e.g., *Nestle's share of 7%*.

It should be noted that both SHARE nouns and GROUP nouns can have the transparent property, just like PARTITIVE nouns, but they do not always. For example, *Each company's share of liability* is a transparent instance of a SHARE noun (the whole phrase is an instance of liability), but *his stock in a media company* is not transparent (the stock is not an instance of a company).

4.14   ABILITY Nouns

ABILITY nouns, exemplified in figure 31, can take *of* plus nominalization or -ing complements like ENVIRONMENT nouns, but can also take infinitival complements. These arguments are assigned the ARG1 role. ARG1 *of* phrases may also have prepositional object NPs that presuppose some action, e.g., *the art of the instant commercial*. A subset of these nouns (marked with a ":SUBJECT" feature in NOMLEX-PLUS) can take ARG0 arguments – this ARG0 is simultaneously the subject of the ARG1 complement. For example, *everyone* in figure 31 (4) is both the possessor of the right (the one who is capable on moral grounds) and the hypothetical disseminator. Examples of ABILITY nouns without subjects include *avenue*, *feasibility* and *gadget* those with subjects include: *ability*, *business*, *career*, *capacity* and *chance*. While some of these nouns are related to adjectives (e.g., *ability* and *feasibility*), they do not seem to be morphologically related to

48

1. *Each company's share of liability*
   REL = share, ARG0 = Each company's, ARG1 = of liability

2. *Nestle's share of 7%*
   REL = share, ARG0 = Nestle's, ARG2 = of 7%

3. *their slice of the profit pie*
   REL = slice, ARG0 = their, ARG1 = of the profit pie

4. *his stock in a media company*
   REL = stock, ARG0 = his, ARG1 = in a media company

Figure 30: SHARE Noun Examples

1. *the art of selling*
   REL = art, ARG1 = of selling

2. *a chance of recession*
   REL = chance, ARG1 = of recession

3. *a career of bribing federal, state and local public officials*
   REL = career, ARG1 = of bribing federal, state and local public officials

4. *the absolute right of everyone to disseminate materials*
   REL = right, ARG0 = of everyone, ARG1 = to disseminate materials

Figure 31: Sample Annotations of ABILITY Nouns

1. *Bob Woodward and Carl Bernstein's book about Watergate*
   REL = book, ARG0 = Bob Woodward and Carl Bernstein's, ARG1 = about Watergate

2. *a drawing of Pinocchio*
   REL = drawing, ARG1 = of Pinocchio

3. *Congress's idea of reform*
   REL = idea, ARG0 = Congress's, ARG1 = of reform

4. *the brainchild of Colorado real estate developer James Morley*
   REL = brainchild, ARG0 = of Colorado real estate developer James Morley

5. *the memory of October 1987*
   REL = memory, ARG1 = of October 1987

6. *his central point about private enforcement suits by environmental groups*
   REL = point, ARG0 = his, ARG1 = about private enforcement suits by environmental groups

Figure 32: Sample WORK-OF-ART Noun Annotation

verbs. We nevertheless base our analysis of argument structure on the canonical argument types, e.g., clauses/events tend to be ARG1s and the main participants in those clauses tend to be ARG0s.

4.15 WORK-OF-ART Nouns

This class basically corresponds to the so-called picture-noun class, discussed in the linguistics literature. Noun's of this class include the originator/transmitter/belief-holder (ARG0) of some X and the subject-matter (ARG1) of X, where X is a human abstraction of some sort (idea, music, etc.) or something capable of recording that abstraction. For example, in *Mary's picture of John*, *Mary's* is the ARG0 and *John* is the ARG1 and *picture* is X, under the interpretation that Mary painted (or photographed) the picture rather than merely owning it. Some examples are provided in Figure 32. As shown, these examples extend outside the reaches of what is normally referred to as a work of art, but rather to anything (e.g., *idea*, *memory*, *restatement*, *philosophy*) that can have some subject matter, and optionally, an originator. Using the verbal frames corresponding to nominalizations among these nouns, e.g., *drawing* is a nominalization of *draw*, would yield the same argument

1. *a metaphor for the U.S. economic system*
   REL = metaphor, ARG1 = for the U.S. economic system

2. *her enthusiastically awful rendition of the "Candy Man"*
   REL = rendition, ARG0 = her, ARG1 = of the "Candy Man", ARGM-MNR = enthusiastically awful

3. *a parody of his previous work*
   REL = parody, ARG1 = of his previous work

4. *the first draft of a screenplay for a "Flashdance"*
   REL = draft, ARG1 = of a screenplay for a "Flashdance"

5. *the Sept. 8 issue of the Journal of the American Medical Association*
   REL = issue, ARG1 = of the Journal of the American Medical Association, ARGM-TMP = Sept. 8

Figure 33: Sample VERSION Noun Annotation

roles (ARG0 and ARG1).[34]

Occasionally, work-of-art nouns can take additional arguments. For example, the work-of-art noun *curriculum* allows an ARG2 beneficiary argument as in *Ms. Crabtree's high school math curriculum* (ARG0 = Ms. Crabtree, ARG1 = math, ARG2 = high school). In many cases, such work-of-art nouns are simultaneously placed in other categories, e.g., *curriculum* is assumed to be a NOMLIKE of *teach*. Other work-of-art nouns that allow beneficiary arguments include *affidavit* and *drawing*.

Autobiography is an unusual work-of-noun in that it violates the one-role/one-phrase rule across the board, since the ARG0 must also be the ARG1, e.g., in *T. Boone Pickens 's autobiography*, *T. Boone Pickens 's* is labeled as the ARG0 (the agent) and the ARG1 (the THEME).

4.16 VERSION Nouns

VERSION nouns typically take one argument (ARG1), but can occasionally take an ARG0, in which case they overlap with the WORK-OF-ART nouns. Like the WORK-OF-ART nouns, the ARG0 argument, when it occurs, is the originator

---

[34]In principle, multiple ARG1s are possible for this group of nouns, e.g., *Ted Bates' Florio commercial about honesty in politics*, where the constituents *Florio* and *about honesty* would both be marked ARG1.

(e.g., author) of the version depicted by the NP. The ARG1 argument (typically an *of phrase*) represents either a model or a series of related objects, such that the NP headed by the VERSION noun represents either: (a) something that emulates the ARG1 (the degree of emulation ranges from indirect to a close copy), e.g., *metaphor*, *parody*, *rendition*, *replica*, *shadow*, *echo*, *gist*; or (b) one of the series of related objects, e.g., *draft*, *version*, *issue*, *edition*. Examples are provided in figure 33. Note that the the ARG0 and ARG1 of this class correspond to the ARG0 and ARG1 of the verb *copy* in PropBank.

4.17 TYPE Nouns

The following nouns are TYPE nouns: *brand*, *breed*, *category*, *class*, *form*, *kind*, *line* (*product line*), *sort*, *strain*, *type*, *variety* and *vein*. TYPE nouns typically take exactly one argument (ARG1), an *of* phrase or left modifier of a noun. As with the VERSION class, the ARG1 represents some model and the meaning of the entire NP is based on that model. In this case, however the NP represents a set of such items, with modifiers of the head potentially limiting the size of the set. Typical verbs that are morphologically related to these nouns (*sort*, *type*, *classify*, etc.) do not share their argument structure – they are causatives, e.g., *He sorted the data.* means something like, "organized the data into piles (sorts)". Thus, our choice of ARG1 is based on the similarity with the TYPE class and the VERSION class. Some examples are provided in figure 34.

Some *TYPE* nouns have an additional argument (ARG2), a specification of the type involved, e.g., in *the series A preferred stock*, *stock* is the ARG1 and *A* is the ARG2.[35] Many of these are also classified as ATTRIBUTE nouns (see next section).

4.18 ATTRIBUTE Nouns

A phrase P headed by an ATTRIBUTE noun A include some key argument B (ARG1) such that P represents an attribute of B. Intuitively, an attribute is some intangible thing about B. They are often described like possessions that things can gain and lose. For example, consider the attribute noun *glamour* in the following sentences:

If growth regains its glamour among investors . . .

People are taking the glamour out of the fur business

Attributes typically have values. In these cases, there is some ARG1 which the glamour is a property of (*its* and *out of the fur business*) and there is an implied

---

[35]Note that *type* in that example is a premodifier of its ARG1. See section 7 for a discussion of modifiers as NomBank predicates.

1. *another type of non-violent civil disobedience*
   REL = type, ARG1 = of non-violent civil disobedience

2. *the sort of nicknames normally associated with linebackers and heavyweight contenders*
   REL = sort, ARG1 = of nicknames normally associated with linebackers and heavyweight contenders

3. *East Germany's conservative vein of communism*
   REL = vein, ARG1 = of communism

4. *this new breed of no-nonsense administrator*
   REL = breed, ARG1 = of no-nonsense administrator

5. *her new line of work*
   REL = line, ARG1 = of work

6. *Sansui's product line*
   REL = line, ARG1 = product

Figure 34: Sample TYPE Noun Annotation

value (ARG2) to this glamour which is in flux. In the example: *My head's width of five inches*, *width* is an attribute of *my head* (ARG1) and the the value (ARG2) of that attribute is *five inches*. Although values of attributes vary widely, they typically are either: (a) some number of units of measure, e.g., *five inches*; (b) some relative measure, e.g., *the lesson's short length*;[36] (c) members of a finite set (e.g., *the color of the book is red*); or (d) something that is characterizable as a descriptive adjective, nominalized form of an adjective, attribute or ability, e.g., *The professor's bad mood*. The attribute value is marked ARG2. Some ATTRIBUTE nouns are also verbal nominalizations, e.g., *color* and *price*. In these cases, the ARG1 and ARG2 roles assigned under the verbal frames line up with those assigned under the AT-TRIBUTE frame. Some examples of ATTRIBUTE noun phrases are provided in figure 35.

When the ARG2 is not present in the NP headed by the ATTRIBUTE noun, the noun itself is assigned the role ARG2-REF. We assume that ATTRIBUTE nouns are sort of like argument nominalizations except when the argument is present. For example, in *its red color*, *red* is the ARG2, but in *its color*, *color* is assigned ARG2-REF. Thus in *its color is red*, the identity between *red* and the ARG2 slot is

---

[36] Other relative measures include: *long*, *a lot*, *a little*, *somewhat*.

achieved via predication – see section 12 for a discussion about similar examples.

ATTRIBUTE nouns can sometimes take a secondary-theme argument (ARG3). For ATTRIBUTE nouns, the ARG1 combines with the attribute to form a new attribute. This new attribute takes the ARG3 as an argument. Thus the ARG2 can be viewed as a value of either the ARG1 or the ARG3. Examples are provided in figure 36. According to our analysis, the number of lines in the first example (the underlying ARG2) refers to both the number of lines [of advertising] and the number of lines [of advertising in the New York Times].

Some attribute nouns take additional arguments. For example, *agility* takes an ARG0 (experiencer) argument. In *His incredible agility in doing the tango*, *His* is the ARG0, *in doing the tango* is the ARG1 and *incredible* is the ARG2; *backlog* has an ARG0 (owner) argument – in *It has a $80 billion backlog of orders*, *It* is the ARG0, linked to *backlog* by the support verb *has* (cf. section 5), $80 billion is the ARG2 and *of orders* is the ARG1.

Attribute nouns can have a CRISSCROSS property (see section 4.22) allowing them to be arguments of their arguments, as shown in figure 37. In most cases (examples 2-4), they are adverbial (ARGM-MNR or ARGM-ADV) modifiers, but they can also be regular numbered arguments, e.g., ARG2, as in the first example.[37]

As stated, the current definition for ATTRIBUTE may be viewed as overlapping the definitions for previously defined classes, particularly the classes: HALLMARK and the subclass of ABILITY that can take a subject. In addition, a large number of adjectival nominalizations also fit the ATTRIBUTE noun description. In all these cases, the nouns are multiply classified. The resulting lexical entries have the superset of the arguments of the classes assigned. Thus an ABILITY/ATTRIBUTE noun like *power* has an ARG0 (agent) , ARG1 (action), ARG2 (recipient) and ARG3 (value). The ARG0 is from the ABILITY classification, the ARG1 combines the ABILITY action and ATTRIBUTE theme role, the ARG2 recipient is specific to this word and the ARG3 value is the equivalent to the standard ARG2 of Attribute nouns.[38] Some examples of ADJNOM/ATTRIBUTE overlaps are provide in figure 38. In some cases, the nominalized adjective arguments are a subset of the ATTRIBUTE arguments, as in the first three examples. In others cases, the nominalized adjective entry adds arguments (usually just 1), e.g., examples 4 and 5.

---

[37]In the first example, both *amount* and *uncertainty* are ATTRIBUTE nouns. *Amount* also happens to be transparent. Thus the whole phrase represents an instance of *uncertainty*.

[38]Here we allow the standard that RECIPIENT arguments should be ARG2 trump the standard that VALUE arguments be labeled ARG2.

1. *the breadth of inquiries*
   REL = breadth, ARG2-REF = breadth, ARG1 = of inquiries

2. *the beauty of a democracy*
   REL = beauty, ARG2-REF = beauty, ARG1 = of a democracy

3. *annual capacity of 600,000 tons*
   REL = capacity, ARG2 = of 600,000 tons, ARGM-TMP = annual

4. *the humor of his uncombed appearance*
   REL = humor, ARG2-REF = humor, ARG1 = of his uncombed appearance

5. *an undercurrent of anger*
   REL = undercurrent, ARG2 = anger

6. *a nationwide mood of despondency*
   REL = mood, ARG1 = nationwide, ARG2 = despondency

7. *The conservative bent of the incumbent appointees*
   REL = bent, ARG1 = of the incumbent appointees, ARG2 = conservative

8. *the quality of the underlying home equity loans*
   REL = quality, ARG2-REF = quality, ARG1 = of the underlying home equity loan

9. *the shortcomings of the institution*
   REL = shortcomings, ARG2-REF = shortcomings, ARG1 = of the institution

10. *the subsidiary's status as a proposed discontinued operation*
    REL = status, ARG1 = the subsidiary's, ARG2 = as a proposed discontinued operation

11. *the title of his speech*
    REL = title, ARG2-REF = title, ARG1 = of his speech

Figure 35: Sample Annotations of ATTRIBUTE Nouns

1. *The advertising lineage in the New York Times*
   REL = lineage, ARG1 = advertising, ARG3 = in the New York Times

2. *the expiration date of its tender offer* REL = date, ARG1 = expiration, ARG3 = of its tender offer

Figure 36: Attribute examples with the secondary theme argument

1. *a tremendous amount of uncertainty*
   REL = amount, ARG2 = tremendous, ARG1 = of uncertainty
   REL = uncertainty, ARG2 = a tremendous amount

2. *the inevitability of the move to small machines that don't make compromises.*
   REL = inevitability, ARG2-REF = inevitability, ARG1 = of the move . . .
   REL = move, ARG2 = to small machines . . ., ARGM-ADV = inevitability

3. *the possibility of a conventional Soviet attack*
   REL = possibility, ARG2-REF = possibility, ARG1 = of a conventional Soviet attack
   REL = attack, ARG0 = Soviet, ARGM-MNR = conventional, ARGM-ADV = possibility

4. *the bitterness of the battle*
   REL = bitterness, ARG2-REF = bitterness, ARG1= of the battle
   REL = battle, ARGM-MNR = bitterness

Figure 37: Crisscrossing ARGM Attribute Nouns

1. *the bitterness of this battle*
   REL = bitterness, ARG2-REF = bitterness, ARG1 = of this battle

2. *the song's incredible beauty*
   REL = beauty, ARG1 = the song's, ARG2 = incredible

3. *glasnost's authenticity* REL = authenticity, ARG2-REF = authenticity, ARG1 = glasnost's

4. *the closeness of test preparatives to the fifth-grade CAT-score*
   REL = closeness, ARG2-REF = closeness, ARG1 = of the test preparatives, ARG3 = to the fifth-grade CAT-score

5. *The Sun's far DISTANCE from Pluto*
   REL = distance, ARG1 = The Sun's, ARG2 = far, ARG3 = from Pluto

Figure 38: ADJ-NOM vs. ATTRIBUTE Senses of Nouns

4.19   ENVIRONMENT Nouns

ENVIRONMENT nouns, exemplified in figure 39, depict either a setting or atmosphere for its argument. We view environment nouns essentially as attribute nouns with boolean values. In other words, an environment is like an attribute that either exists or does not exist for a particular ARG1 and canonically. While no ARG2 is necessary, ARG2s are possible (with the same interpretation as with ATTRIBUTE nouns), e.g., figure 39, example 8.

The setting depicted by an environment noun may be a situation, time period, philosophy, location or physical body (e.g., *the earth's atmosphere*). The ARG1 may be something physical (like *the earth*), an event (like *discussion* in *forum for discussion*), or set of events that occur relative to the setting (like *battling* in *years of battling*). The ARG1 typically consists of an *of* phrase (although other syntactic positions are possible). When the argument is an event, *of* takes a nominalization or -ing phrase object. ENVIRONMENT nouns include specialized words like *crucible* and *medium* as well as many nouns that simply name periods of time (*June*, *morning*, etc.) location nouns (*land*, *site*), nouns depicting philosophies or mechanisms that guide actions or processes (*politics*, *economics*) or nouns representing fields of energy or gas that surround the argument (*atmosphere*, *aura*). As ENVIRONMENT nouns do not seem to include nominalizations, we cannot base our choice of arguments on existing verb frames. Nevertheless, we mark these events as ARG1 because events canonically are marked ARG1 for verbs.

1. *the crucible of Panama's long history of conspirators and pirates*
   REL = crucible, ARG1 = of Panama's long history of conspirators and pirates

2. *a time of exceptionally high U.S. exports of dry milk*
   REL = time, ARG1 = of exceptionally high U.S. exports of dry milk

3. *a period of industry consolidation*
   REL = period, ARG1 = of industry consolidation

4. *a recent morning of working at home*
   REL = morning, ARG1 = of working at home, ARGM-TMP = recent

5. *the earth's atmosphere*
   REL = atmosphere, ARG1 = the earth's

6. *the arithmetic of deals*
   REL = arithmetic, ARG1 = of deals

7. *the economics of medicine*
   REL = economics, ARG1 = of medicine

8. *ballooning hour of 6 a.m.* [ATTRIBUTE or Sense 1] REL = hour, ARG1 = ballooning, ARG2 = of 6 a.m.

9. *trading hours* [ENVIRONMENT or Sense 1]

Figure 39: Sample Annotation of ENVIRONMENT Nouns

1. *the issue of abortion*
   REL = issue, ARG1 = of abortion

2. *the subject of U.S. direct investment in Japan*
   REL = subject, ARG1 = of U.S. direct investment in Japan

Figure 40: Sample Annotations of ISSUE Nouns

1. *the rapidly growing field of bio-analytical instrumentation*
   REL = field, ARG1 = of bio-analytical instrumentation

2. *the science of sleep*
   REL = science, ARG1 = of sleep

Figure 41: Sample Annotations of FIELD Nouns

4.20   ISSUE Nouns

ISSUE nouns including *issue*, *subject* and *topic* are heads of noun phrases which depict the general topic of discourse, specified by the only argument, an ARG1. Examples are provided in Figure 40.

4.21   FIELD Nouns

FIELD nouns (*field*, *science* and *study*) represent fields of study, as specified by the only argument an ARG1. Examples are provide in Figure 41.

4.22   CRISS-CROSS Nouns

A CRISS-CROSS noun is a noun that: (1) takes a PREDICATE as one of its arguments (ARG1); and (2) acts like an argument of this predicate, e.g., in *the victim of the attack*, *The victim* is an argument of *the attack*. Another argument (ARG0) is also possible which would simultaneously be the ARG0 of the predicate (the ARG1), e.g., in *John's experiment victim*, *John* would be the ARG0 of both *victim* and *experiment* (the ARG1). We have identified three subclasses of CRISS-CROSS nouns: the PATIENT variety, the TOPIC variety and the INTRANS variety. PATIENT-CRISS-CROSS nouns include: *victim* and *prey*. These CRISS-CROSS nouns are typically a patient of their ARG1s. TOPIC-CRISS-CROSS nouns include: *matter*, *area* and *subject*. These CRISS-CROSS nouns are typically topics of discussion. The word *object* may be a member of both these subclasses.

1. *victim of a tainted transfusion*
   REL = victim, ARG1 = of a tainted transfusion

2. *the victim of widespread fraud*
   REL = victim, ARG1 = of widespread fraud

3. *the prey of ambitious local politicians*
   REL = prey, ARG0 = of ambitious local politicians

4. *likely prey for corporate raiders*
   REL = prey, ARG0 = for corporate raiders, ARGM-ADV = likely

5. *his victim*
   REL = victim, ARG0 = his

6. *the standard object of suggestions for organizational and institutional changes*
   REL = object, ARG1 = of suggestions for organizational and institutional changes

7. *a major topic of post-game discussion*
   REL = topic, ARG1 = of post-game discussion

8. *another matter of contention*
   REL = matter, ARG1 = of contention

9. *a subject of speculation*
   REL = subject, ARG1 = of speculation

10. *specific areas of concern*
    REL = areas, ARG1 = of concern

11. *The two sides in the legal battle*
    REL = sides, ARG1 = in the legal battle

Figure 42: Sample Annotations of CRISS-CROSS Nouns

Although it is usually the case, it is not necessary for the CRISS-CROSS noun to be an argument of the actual ARG1 of TOPIC- and PATIENT- CRISS-CROSS nouns, provided that the ARG1 implies such a predicate. For example, in *a victim of Gramm-Rudman cuts*, the victim may be something that is affected by the Gramm-Rudman cuts even if the victim was not actually cut, e.g., the victim could be someone who worked for a program that was cut. Similarly, the ARG1 of the criss-cross noun does not always take an argument, e.g., a *scandal victim* is one who is effected by a scandal, even though the word *scandal* does not independently take such an argument. There is a third type of criss-cross noun (INTRANS-CRISS-CROSS) that has at least one member (*side*). When occurring with INTRANS-RECIP nominalizations, nominalizations that take a reciprocal subject, *side* plays the role of subject (ARG0), e.g., in *The two sides of the battle*, *the two sides* are the ones battling. Figure 42 provides examples. Note that the entries for the ARG1 nouns will include the CRISS-CROSS noun heads as arguments for all these examples, e.g., in the first example, *victim* will be listed as the ARG1 of *transfusion* in the proposition for that instance of *transfusion*.

When marking predicates that are arguments of CRISS-CROSS nouns, the criss-cross noun may act like a SUPPORT predicate (see section 5 below). In other words, the ARG0 of the CRISS-CROSS noun may be an argument of its ARG1. For example, consider, the example *John's assassination victim* would require two annotated propositions, one for *victim* and one for *assassination*:

1. REL = victim, ARG0 = John, ARG1 = assassination

2. REL = assassination, SUPPORT = victim, ARG0 = John, ARG1 = victim

The relation between criss-cross nouns and support becomes clearer when comparing the following two examples: (1) *Mary received a grant* and (2) *Mary is the recipient of a grant*. The criss-cross noun *recipient* is an ARG0 nominalization of the support verb *receive*. In both cases the ARG0 is an argument (an ARG2) of *grant*. This suggests that we can treat criss-cross nouns as argument nominalizations of support verbs even when there is no corresponding support verb.

4.23 EVENT nouns

Event nouns are nouns which depict an event, and either take no arguments or take a single argument – typically a place that is directly effected by the event, e.g., *San Francisco* in the case of *the San Francisco earthquake*. Many instances of EVENT nouns include no arguments, but rather just propositional modifiers,

1. *The earthquake struck San Francisco*
   REL = quake, Support = struck, ARG1 = San Francisco

2. *Last year's drought in the Midwest*
   REL = drought, ARGM-TMP = Last year, ARGM-LOC = in the Midwest

3. *a luncheon in London*
   REL = luncheon, ARGM-LOC = in London

4. *the oily, hour-long rubfests*
   REL = rubfests, ARGM-TMP = hour-long

Figure 43: EVENT noun Instances

usually temporal and/or locative. Examples are provided in figure 43.[39] This class was included because these NPs provide valuable propositional information even when no arguments occur– one might think of them as nominalization cousins that are always intransitive. Rolesets for EVENT nouns that take no arguments contain very little information, i.e., they do not list any roles.

4.24   Minor Adjustments of Above Classes

The above classes provide default frames for the members of those classes. However, what should an annotator do when a word nearly meets a class definition, but does not do so exactly. There are two possible alternatives: 1. put it in a different class; or 2. alter the frame slightly, making the word in question a minority exception to that class.

Consider the word *ambassador*, which nearly fits our RELATIONAL noun (ACTREL) class, but unfortunately has an extra argument. Our solution is to assume that it is a subject nominalization cousin (also known as NOMLIKE) modeled on the word "represent". Thus *France's ambassador to China* is analyzed something like "X represents France in China", where X is the ambassador (ARG0 = X, ARG1 = *France*, ARG2 = *in China*).

Consider *grader medalist* and *nationalist* which are sort of like object nominalizations of ATTRIBUTE nouns. For these, we can simply alter the default frame for ATTRIBUTE slightly to come up with appropriate frames – we assume that the ARG1 has been incorporated and that the ARG2 cannot be incorporated. For

---

[39] *Earthquake* and *rubfest* are assumed to be lemmas of *quake* and *fest* as per NomBank's morphology dictionary nombank-morph.dict.

example, *first grader* is analyzed as follows: REL = grader, ARG1 = grader, ARG2 = first.

It is also the case that some of the arguments used in the above classes pop up with nominalizations in addition to the nominal arguments that correspond to verbal arguments. In particular, it sometimes seem useful to add a VALUE argument when a typical nominalization behaves a little like an attribute noun or a SECONDARY-THEME argument when (like the attribute noun and partitive noun cases), an additional argument seems to double a theme or other typical ARG1. For example, in *Death notices for the factories*, we marked "Death" with the ARG1 role and "for the factories" with the ARG3 (or secondary theme) role. As with attribute nouns, we interpreted these arguments in a layered fashion: the notice was about *death* (ARG1) and the *death notice* was about the factories (ARG3).

5 Support Constructions

In the preceding sections, we mostly give examples in which all arguments of a noun N are assumed to be part of the NP headed by N. However, there are two notable exceptions: support verbs and transparent nouns (e.g., partitive/share constructions). In both cases, particular lexical items serve to connect head nouns with arguments that lie outside of the NP. We discuss support verbs first.[40]

5.1 Support Verbs

We are concerned here with arguments of nouns which occur in structures in which they are also arguments of a support verb. A support verb is a verb V that takes an argument-bearing NP A as one of its arguments, and at least one other argument B, such that A also takes B as an argument. With respect to this argument-sharing capacity, support verbs are much like the so-called raising and equi (control) verbs that have been studied so extensively over the years.[41] While support verbs have a smaller literature than raising/equi verbs, they have previously been analyzed in [4, 5, 17, 18, 3]. We have also previously undertaken a small support verb project at NYU [14] and created a pilot dictionary of 20 nominalizations and the

---

[40]Our paper "NP-External Arguments: A Study of Argument Sharing in English" ([19]) discusses the issues presented in the next few sessions in detail.

[41]For some researchers, "support verbs" must be semantically empty and are thus more like raising than equi predicates. We do not make this requirement because: (1) we wish to annotate all instances of argument sharing between a noun and the governing verb; and (2) the distinction between semantically empty and semantically contentful verbs is difficult to make consistently, e.g., which of the verbs in the following are semantically empty: "give a kiss", "make an attack", "hurl an accusation", "complete the invasion"?

co-occurring support verbs (a total of 432 verb/nominalization pairs). The project was called XMELLT.[42]

Figures 44 providesb some examples of support verb plus noun combinations. The remainder of this section is devoted to defining support verbs and giving guidance in annotating structures that contain them.

Maurice Gross [5] notes that support verb plus noun combinations may take different sets of complements than either the noun or verb would by themselves. For example, compare the well-formed *I had a dream that I could fly* with the ungrammatical **I described a dream that I could fly*. Notice that support verb plus noun combinations are somewhat idiom-like and we previously stated that most idioms are not markable. In fact, when considering a verb plus noun combination like *keep tabs*, one could alternatively claim that: (1) this is an idiom or (2) that the noun *tabs* takes a special set of arguments when it cooccurs with *keep*. In other words, in *Mary kept tabs on all her children* is *Mary* the subject of the idiom *keep tabs* or is *Mary* the shared subject of the verb *kept* and of (one sense of) the noun *tabs*. In principle, it would be difficult to distinguish these options. We therefore must provide guidelines for drawing the line between unmarkable idioms and allowable support verb plus noun combinations.

A support verb plus noun combination is always markable if the noun can take some of the same arguments in other environments. Thus *John took a walk to the store* is markable because NPs such as *a walk to the store* and *John's walk to the store* contain the same arguments even though the NP headed by *walk* is not the object of *take*. Secondly, if the noun takes at least one argument other than the subject of the support verb, we assume it is markable. Thus the *dream* example is allowed, as are the examples in figure 45. This, however, rules out sentences with idioms like *kicked the bucket* and *bought the farm* where the only argument that is not part of the idiom is the subject of the sentence. With these cases, we assume that the entire verb phrase (VP) is an idiom, the VP being the smallest constituent that dominates all instances of the noun/verb collocation. In contrast, for the markable cases there are variable items inside the VP. The *keep tabs* case takes both a subject and a PP headed by *on*. Therefore it is also markable according to these criteria in spite of the common assumption that this is actually an idiom. In fact for idioms that consist solely of a noun, verb and possibly closed class items (prepositions, particles), we make the task-centric assumption that these are instances of support.[43]

---

[42]The XMELLT project focused on combinations of support verbs and nominalizations. This annotation effort includes collocations between support verbs and nouns in general.

[43]We will end up marking some widely recognized idioms as support constructions. Nevertheless, the argument structure that we are annotating is easily adaptable should a user decide to assume a different analysis. (We would only be marking decomposable idioms that take arguments).

1. *The two sides in the legal battle have hurled accusations of duplicity at each other*
   SUPPORT = hurled, REL = accusations, ARG0 = The two sides in the legal battle, ARG1 = at each other, ARG2 = of duplicity

2. *it might take action to cure the default*
   SUPPORT = take, REL = action, ARG0 = it, ARG1 = to cure the default

3. *a hostile offer is being made by a foreign company for all of ESB's shares*
   SUPPORT = made, REL = offer, ARG0 = by a foreign company, ARG1 = for all of ESB's shares

4. *to give the department "maximum flexibility" to deal with the cuts*
   SUPPORT = give, REL = flexibility, ARG0 = the department, ARG1 = to deal with the cuts"

5. *the funds that [*T*] make a habit of taking out loans to buy extra junk*
   SUPPORT = make, REL = habit, ARG0 = [*T*]→ the funds, ARG1 = of taking out loans to buy extra junk

6. *The campaign takes advantage of the eye-catching photography.*
   SUPPORT = takes, REL = advantage, ARG0 = the campaign, ARG1 = of the eye-catching photography

7. *Many fly-by-night charities ride the coattails of the biggest, best-known and most reputable ones*
   SUPPORT = ride, REL = coattails, ARG0 = Many fly-by-night charities, ARG1 = of the biggest, best-known and most reputable ones

8. *A battle with Mr. Icahn would rattle even the most seasoned chief executive*
   SUPPORT = rattle, REL = battle, ARG0 = even the most seasoned chief executive, ARG1 = with Mr. Icahn

9. *[*] Take Comfort in Cotton*
   SUPPORT = take, REL = comfort, ARG0 = in Cotton, ARG1 = [*]

Figure 44: Sample Support Verb plus Noun Annotation

65

1. *Convenience store merchandise has not kept pace with current trends in consumer preferences* SUPPORT = kept, REL = pace, ARG0 = Convenience store merchandise, ARG1 = with current trends in consumer preferences,

2. *But the surprisingly durable seven-year economic expansion has made mincemeat of more than one forecast.*
   SUPPORT = made, REL = mincemeat, ARG0 = the surprisingly durable seven-year economic expansion, ARG1 = of more than one forecast

Figure 45: Idiom-like Support Verb plus Noun Annotation

We require that all support verb noun combinations be productive instances of lexical selection. In other words, it is not enough for an annotator to observe some causal connection between a verb and one of its arguments. Consider the examples in figure 46. In principle, one might assume that *determine* in example (1) is a support verb and the clause beginning with *whether the Gramm-Rudman ...* is an ARG1, i.e., *The White House Office of Management and Budget calculated whether ...*. However, this really appears to be a property of the verb *determine* and little to do with the collocation between *calculations* and *determine*. It would seem that it does not matter whether *calculations* or *John Smith* is the subject of *determine* – it is thus questionable whether the question of meeting the targets is actually being calculated. In contrast, the argument structure for real SUPPORT verb noun pairs is based on the interaction between the argument taking properties of both the noun and verb, not just one of the items. In example (2), any event can "lead to" another event. So it should be clear that the phrase *market conditions* is not the ARG0 of *cancellation*. What should make it even clearer, however, is the fact that *market conditions* violate the selection restrictions of *cancellation*, e.g., the following sentence would be ill-formed: *the market conditions canceled the planned exchange*. In example (3), there is a discourse relation between "proponents" and "cause" that is independent of the verb "say". For example, "proponents" can have this same sort of relation with a constituent in a previous sentence, e.g., *Gun control is the issue here. Proponents have contributed one billion dollars per year to the Democratic party.*[44] Finally, the PP *for a celebration* characterizes the reason for President Bush gathering. This PP could provide a reason for any semantically compatible action and would usually (but not always) imply that the actor is also one of the ones celebrating, e.g., compare *President Bush frolicked to town for the celebration*. Even though there is no support in this example,"President Bush"

---

[44]Such relationships are sometimes dealt with in the coreference literature. For example, [24] might characterize *the cause* as "a bridging description" rather than a true anaphor of "proponents".

1. *The White House Office of Management and Budget, whose calculations [*T*-2] determine whether the Gramm-Rudman targets are met*
   REL = calculations, ARG0 = whose → The White House Office of . . .

2. *market conditions led to the cancellation of the planned exchange*
   REL = cancellation, ARG1 = of the planned exchange

3. *Proponents say the cause is just*
   "cause" has no arguments present

4. *President Bush gathered for a celebration*
   "gathered" is not a support verb, although "President Bush" is the ARG0 of "celebration" (as per section 10)

Figure 46: Instances that do NOT have Support verbs

actually is the ARG0 of celebration due to the PP construction containing "celebration", as discussed in section 10. The examples in Figure 46 have in common that there is no support relationship between the main verb V and N, the argument of that verb under discussion. This is in spite of other factors (discourse factors, coreference, etc.) which may cause one to deduce that some other argument of V should fill some argument slot of N.

Another factor which is relevant to choosing support verbs has to do with such factors as modality and shared participation in an event. Observe how the main verbs in the following examples change the relation between the predicate "attack" and "John", the ARG0 or subject argument: "John made the attack", "John orchestrated the attack", "John participated in the attack", "John aided the attack", "John planned the attack", "John is considering the attack". In a very loose sense, "John" is the ARG0 of attack in each of these sentences. We have chosen this loose sense of argumenthood because we believe it is the most consistent and generally captures the argument structure of "attack". Mirroring PropBank's approach to arguments of equi verbs, we are interested in argument structure and not, in other aspects of meaning. These examples differ both in modality (whether or not the attack actually takes place) and in whether or not "John" is the only attacker. Nevertheless, in each example "John" is either the attacker, the potential attacker, someone controlling the attack, someone participating in the attack or an accomplice, and these distinction involve shades of gray. It would seem that the most consistent statement would be to say that "John" is an ARG0 in each case. For much the same sort of reasons, we ignore the effects of *belief* contexts, scope effects and negation, e.g., "John did not make the attack" would be annotated with

much the same argument structure as "John made the attack". Some borderline cases still remain. For example, consider "it was discussing technical cooperation with Saab", where "it" refers to "Fiat". Here, it is logically possible that Fiat is discussing someone else's cooperation with Saab, but highly unlikely. In the "control" (equi) literature, similar cases may be ones of arbitrary control, rather than subject control. We will nevertheless consider such cases markable, i.e., we assume that "it" is the subject (ARG0) of cooperation in this example.[45]

In a similar vein, the interpretation of nonsubjects can be effected by support verbs. For example, consider the phrase *the destruction to the Bay Area wrought by Tuesday's quake*. The support verb *wrought* changes the interpretation of the ARG1 relation for *destruction*. This phrase means something like *damage to the Bay area*, whereas instances of *destruction* without SUPPORT verbs typically imply that the ARG1 is in total ruin (without a support verb, the ARG1 of *destruction* does not occur as a *to* phrase). Once again, this is similar to the effect that raising and equi verbs have on their clausal arguments, e.g., *wanting to do something* is not the same thing as doing it.

In principle, any verb that takes an object or prepositional phrase complement can be a support verb. Some common support verbs include: *give*, *have*, *get*, *bring*, *carry*, *do*, *obtain*, *need*, *make*, *take* and *undergo*. Common subject control verbs (with both to-infinitive and other complements (-ING)) are often also support verbs, e.g., *Mary wants to be recognized/Mary wants recognition*, *Mary promised John to pay him/Mary promised John payment*, *John can't afford to pay taxes/John can't afford the payment of taxes*, *John abstained from eating/John abstained from all activity*, etc. For ease of annotation, a number of lists are provided in Appendix B and some supplementary files. Appendix B includes: a sampling of support verb plus noun combinations, all the support verbs found in XMELLT, lists of different types of subject control verbs from COMLEX Syntax. The XMELLT dictionary is also be available to annotators as a separate file (as well as other dictionaries extracted therefrom).

In principle, any noun can co-occur with a support verb. In practice, most support verb constructions involve nouns with complex argument structure (e.g., nominalizations). However, other argument-taking nouns may also occur with support verbs. Verbs like *rise, fall, increase, decrease* can be support verbs for partitive

---

[45]The examples discussed in this paragraph dealt with the agent (a subtype of ARG0) of the lower predicate, as controlled by the subject of the support verb. However, the same issues exist for other arguments as well, e.g., consider the subject of "bribery" in "John pleaded guilty to bribery" and "They convicted John of bribery". In these cases, the veracity of "John" being in a guilty state is at issue (the degree-of-argumenthood issue is not restricted to agents). In one of the cases, "John" is the subject of the support verb and in the other "John" is the object (i.e., this issue is not exclusive to subjects).

nouns. Thus, *The price increased 5 percent* would have the analysis: REL = *per-cent*, SUPPORT = *increased*, ARG1 = *the price*. Also, *have* can be a support verb for attribute nouns, e.g., *This desk has a height of 25 inches* (REL = height, Support = *has*, ARG1 *This desk* = ARG1, ARG2 = *of 25 inches*.

5.2 Some Clarifications about Support Verbs

Some modification at various higher up levels may have broader effects. For example, in *On Thursday, John did not take a walk*, the temporal and negative modification of the verb *take* effects the interpretation of *walk*. Nevertheless, the only structure marked for *walk* is: SUPPORT = take, REL = walk, ARG0 = John. The rationale includes: (1) that there are scope issues that we wish to avoid during this round of annotation; and (2) that the information was marked as part of the verbal argument structure during PropBank, i.e., it is recoverable for those users who wish to use it.

Given relative clauses, reduced relative clauses and gapping constructions, multiple instances of support verbs may occur with the same noun. The support verbs may introduce multiple events. For example, consider the use of *disorders* in the following sentence:

The disorders, which 20 years ago struck middle-age and older people , now strike people at the height of productivity

REL = disorders, Support = struck, ARG1 = middle-age and older people

REL = disorders, Support = strike, ARG1 = people at the height of productivity

For such cases of multiple events, we create multiple propositions, one for each support verb. Alternatively, multiple support verbs could fill in different arguments for the same event, e.g., *Rome suffered an attack, waged by Carthage*. In such cases, we have multiple SUPPORT slots for a single event as shown in Figure 47. There are currently 42 noun predicates (out of over 114K) that license multiple propositions. We have provided a list of these addresses with the NomBank release – the file name is: Addresses-of-Multiple-Propositions.

Passivized support verbs have some special properties: (1) If the shared argument is the underlying subject of the passive, this argument may be absent, e.g., *The bid for Nekoosa was made* – in this case, support is not marked; and (2) The argument taking noun may be separated from one of the arguments that typically occurs inside the NP, e.g., *A final modification was made to the five-point opening limit* – in this case, support is marked.

As shown in the Figure 48, in addition to passive verbs, adjectives and nouns may also play the SUPPORT role. For adjectives, as with passives, the copula is assumed not to be part of the SUPPORT chain.[46]

---

[46]Common support adjectives include: *confident* (*She was confident of victory*) and *subject* (*The*

1. *Commercial fishermen and fish processors filed suit in federal court in a claim that [*T*-1] has ballooned to more than $104.8 million*
SUPPORT = filed + suit + in, SUPPORT = ballooned, REL = claim, ARG0 = Commercial fishermen and fish processors, ARG3 = to more than $104.8 million

2. *In a separate complaint also filed in federal court, shareholder Max Grill charged Imperial with breach of fiduciary duty*
REL = complaint, Support = charged, Support = filed, ARG0 = shareholder Max Grill, ARG1 = with breach of fiduciary duty, ARG2 = in federal court, ARG3 = Imperial

Figure 47: Multiple Support Verbs for a Single Noun

1. *it is scheduled [*-1] for completion by Dec. 10*
SUPPORT = scheduled + for, REL = completion, ARG1 = it, ARGM-TMP = by Dec. 10
PASSIVE EXAMPLE

2. *It's so close to completion*
SUPPORT = close + to, REL = completion, ARG1 = it
ADJECTIVE EXAMPLE

3. *their responsibility for hard decisions.*
SUPPORT = responsibility, REL = decisions, ARG0 = their, ARGM-MNR = hard
NOUN EXAMPLE

Figure 48: Passive Support, Adjective Support and Noun Support

70

1. *I take advantage of this opportunity to make a plea to the millions of readers of this newspaper*
   SUPPORT = take+advantage, REL = opportunity, ARG0 = I, ARG1 = to make a plea to the millions of readers of this newspaper

2. *the students who took part in the demonstrations*
   SUPPORT = took+part+in, REL = demonstrations, ARG0 = the students

Figure 49: Some Examples of Chains of Support Verbs

5.3 Support Chains

In the world of noun argument structure, support verbs play the role that equi/raising verbs do in the world of verbal argument structure. The support verb shares an argument with the noun just as a raising or equi verb shares an argument with the lower predicate.[47] Thus it should not be surprising that just as there are chains of equi/raising structures (*John wants to try to leave.*, *Mary seems to be likely to want to leave.*), there can also be chains of support structures. Examples are provided in figure 49. In each case, the support verb is followed by an argument taking noun $N1$ (*advantage* and *part*) which shares its arguments with the head of its argument $N2$ (*opportunity* and *demonstrations*). Thus we must separately mark the argument structure of both $N1$ and $N2$. Figure 49 expresses the argument structure of $N2$.[48] In general, the value of SUPPORT is a list of heads $H_1, \ldots, H_N$ such that the "supported" argument of REL is a surface argument of H_1; H_N takes REL as its argument; and for each item value of $H_i, 1 <= i <= N - 1$, the phrase headed by H_{i-1} is an argument of H_i.[49] Figure 50 is a graphical representation of how support chains link nominals to their arguments.

Interestingly, support chains seem to share a constraint with the so-called raising and control phenomena – a support chain cannot cross a finite clause boundary. In other words, we are aware of no cases of support chains that link an argument of

---

rules were subject to change).

[47]The shared argument is always a surface argument of the upper verb, but raising and equi verbs differ as to whether the shared argument is also a logical or "deep" argument of the upper verb. We do not need to draw this distinction for support verbs in this project as we are only analyzing the noun's argument structure, i.e., the lower predicate. An argument's status with respect to the support verb (the upper predicate) was part of the earlier PropBank project.

[48]It should be noted that these instances of $N1$ may occur with these same support verbs with NP arguments that are not headed by argument-taking nouns. For example, contrast *John took advantage of Bill* and *John took advantage of the opportunity.*

[49]This puts a limitation on how far away an argument of a noun can be. It must, of course, be in the same sentence. Beyond that, there must be an unbroken chain of support words.
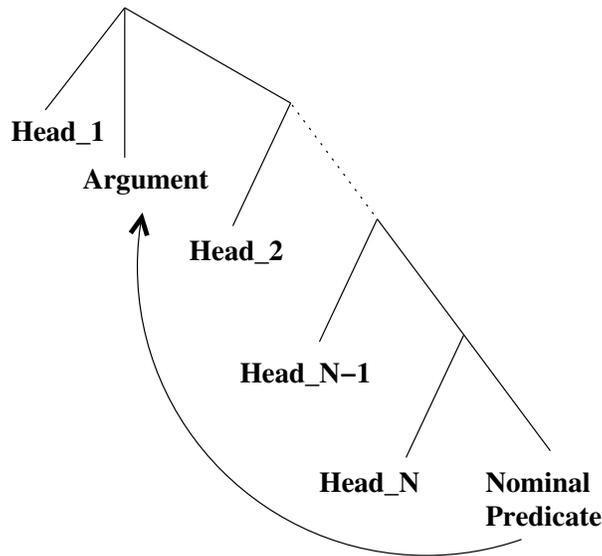
Figure 50: A Support Chain from a Nominal Predicate to its Argument

a matrix clause with a noun in a finite subordinate clause. For example, it is illegal to make a support chain linking *rivals* and *IBM* in a sentence like *IBM claims that rivals want its secrets*. As noted elsewhere the connection between *IBM* and *rivals* is an instance of bridging coreference, not a support relation – there is no lexical property of *claims* that licenses this relation. In fact, virtually any verb taking a sentential complement could occur in that position and the relation would be equally likely. This constraint is helpful in restraining our annotators from over-positing support.[50]

5.4   Partitive Support Chains

Some additional examples of chaining are provided in Figure 51. These examples would seem to be evidence that quite generally, all (transparent) partitives (PARTITIVE nouns or quantifiers in partitive constructions) and SHARE nouns link head nouns to their arguments, just like support verbs do (we use the same label as sup-

---

[50]Support across finite clausal boundaries is of course possible if other phenomena link a noun with a lower clause. For example, in *the walk that John took*, *walk* and *John* are arguments of *took* due to the properties of the relative clause construction. Thus support does not license this link and this is not a counter-example.

1. *a 15 percent share in the company*
   SUPPORT = share, REL = percent, ARG1 = in the company

2. *his share of accomplishments*
   SUPPORT = share+of, REL = accomplishments, ARG0 = his

3. *We had lots of internal debate about this one*
   SUPPORT = had+lots+of, REL = debate, ARG0 = We, ARG1 = about this one, ARGM-MNR = internal

4. *Some of West Germany's bluest chips took some of the biggest hits*
   SUPPORT = took+some+of, REL = hits, ARG1 = Some of West Germany's bluest chips

5. *the Big Board is considering a variety of actions to deal with program trading.*
   SUPPORT = considering+variety+of, REL = actions, ARG0 = the Big Board, ARG1 = to deal with program trading

Figure 51: Some Examples of Support Chains with Partitives

port verbs here).[51] Furthermore, given a sequence of support verb plus partitive, a support verb chain is often possible. The idea is that (most) partitive/share constructions are transparent (like determiners) and the object of the preposition *of* acts like the head with regard to the support verb or possessive. For example, *his share of accomplishments* is treated much the same as *his many accomplishments* – both *share of* and *many* can be passed over in order to find the ARG0 of *accomplishments*. Similarly, with support verbs, *had lots of debates* may be treated the same has *had many debates*, in both cases the surface subject of *had* is the ARG0 of *debates*.

Partitives occurring in support chains sometimes may fill an argument or adjunct position within the argument structure of the underlying predicate (like crisscross nouns). As is clear from the examples in figure 52, partitives can function as negation (ARGM-NEG), as arguments of ATTRIBUTE nouns (ARG2 in the example), and perhaps other positions.

---

[51]All transparent nouns (defined above) seem to have this linking property. PARTITIVE and SHARE nouns appear to be the two largest classes of transparent nouns.

1. *Hollywood has lots of clout*
   Support = has+lots+of, REL = clout, ARG1 = Hollywood, ARG2 = lots

2. *She took none of the opportunities* Support = took+none+of, REL = opportunities, ARG0 = she, ARGM-NEG = none

Figure 52: Partitives in both SUPPORT and argument/adjunct roles

1. *the 1906 San Francisco destruction included insured losses of $5.8 billion.*
   REL = destruction, ARGM-LOC = San Francisco, ARGM-TMP = 1906

2. *the car was tainted by false charges of sudden acceleration*
   REL = acceleration, ARGM-MNR = sudden

Figure 53: Metonymy raises questions about Support

5.5 Support Verbs and Metonymy

Metonymy is the ability of one noun M to stand in for another NP that it represents. For example, the city name *Chicago* can stand in for a sports team in *Chicago won the game* or a type of sandwich can stand in for a customer in a restaurant when the staff is talking about them, e.g., *Would you please get the ham sandwich another cup of coffee?*. *Metonymy*, unfortunately, seems to interact with SUPPORT. Consider the two examples in figure 53. For the first example, the question is whether or not *insured losses of $5.8 billion* qualifies as the ARG1 of *destruction*. Arguably, *include* is a support verb and its object (the losses) is part of what was destroyed. However, it would appear that *the 1906 San Francisco destruction* is standing in for the total amount of money lost as a result of that destruction. On this interpretation, metonymy is coercing *destruction* not to be an action at all, but rather an amount of money. We believe that this is the correct interpretation since the losses were not destroyed, nor even partially destroyed, as in the *destruction wrought by* case discussed at the end of section 5.1.

In the second case, *the car* is standing in for *the reputation of the car* by metonymy. Thus on the one hand, *the car* is arguably the ARG1 of *acceleration* and on the other its reputation is arguably the ARG1 of *taint*. The coercion of *the car* to *the reputation of the car* is a key factor in our analysis. First, we let's see how the SUPPORT analysis would go. Let's look at a clearer case with the same argument structure: *the charges of harassment tainted Bill*. It is clear that the ARG1 of *charges* is also the ARG0 of *harassment*, i.e., if X is charged with doing Y, X is an argument of Y (typically ARG0). Thus, the question is that if X is tainted by

1. *John's assassination victim*
   REL = assassination, SUPPORT = victim, ARG0 = John, ARG1 = victim

2. *Saab is looking for a partner for financial cooperation*
   SUPPORT = looking + for + partner, REL = cooperation, ARG0 = Saab,
   ARG2 = a partner, ARGM-MNR = financial

Figure 54: A CRISS-CROSS noun can interact with a Support chain

charges, does that entail that X is the ARG1 of *taint*? For some people the answer
is yes. However, even in this case, it is not *Bill* who is being tainted. Rather, *Bill* is
standing in for *Bill's reputation*. The fact that people stand in for their reputation
more often than cars do makes this case easier to accept than figure 53, example
(2). However, we will assume that all such examples are not markable instances of
SUPPORT. A key factor is that the ARG1 of *taint* and the ARG1 of *acceleration*
cannot be the same entity – reputations are *tainted* and things in motion *accelerate*.
Coercion is needed to represent them with the same string of words.

5.6 Criss-Cross Noun plus Support Chains

As noted in section 4.22, a criss-cross noun V takes an ARG1 A such that V is
an argument of A, e.g., in *the victim of an assassination*, *the victim* is the ARG1
of *assassination*. As shown in figure 54, criss-cross nouns can act like support
predicates by themselves. Additionally, they be part of support chains. Thus when
a CRISS-CROSS noun appears with both its ARG0 and its ARG1, the ARG0 is
typically an argument of the ARG1 (as in figure 54, example 1). In Example 2,
partner is an INTRANS-CRISS-CROSS noun occurring with an intrans-recip verb
(cooperation) and linking the ARG0 and ARG2. In this case, the ARG0 is linked to
"partner" via the SUPPORT verb "looking for". The chain then continues through
partner to reach the predicate *cooperation*.

5.7 Additional Examples of Support

There are a set of cases in which a noun is the same semantic class as a support
verb, e.g., figure 55. In the first two examples, a noun of communication is the
subject of a verb of communication, both which take a clausal argument or "about
PP" argument, both typically an ARG1. In these cases, the subject invariably is
the more informative predicate. In the third sentence, the support verb and its
nominalization object are "battle" predicates sharing a subject. These are possible
due to the redundancy between the semantic class of the noun and the semantic

1. *The complaint alleges that the price is unfair*
   SUPPORT = alleges, REL = complaint, ARG1 = that the price is unfair

2. *Harley-Davidson's complaint claims that the group violated securities laws*
   SUPPORT = claims, REL = complaint, ARG0 = Harley-Davidson, ARG1 = that the group violated securities laws

3. *Costume jewelry makers fought a losing battle*
   SUPPORT = fought, REL = battle, ARG0 = costume jewelry makers

Figure 55: Communication Nouns with Communication Support Verbs

class of the verb. This support is lexical because the argument sharing is due to the relation between these two lexical classes.

6 Arguments via Predication and Other Phenomena

Arguments of nouns can be linked to the noun via predication, as shown in figures 56, 57 and prep-noun-as-adj. Figure 56 only includes examples in which a copula is used for predication such that the nominal predicate (REL) precedes the copula and the argument follows the copula. Figure 57 and prep-noun-as-adj exemplify other possible forms of predication that can be used to link nouns to their arguments.[52]

When an argument-taking noun precedes a copula, its postnominal clausal arguments and PP arguments (when the head preposition is not *of*) can follow the copula as in figure 56. We assume that any such postcopular argument that is not an NP (or number phrase) is markable.[53] The copula is not a support verb because this situation is very general – most PP and clausal arguments of nouns may occur in this construction (predicate noun + copula + argument). This same analysis could be generalized to other predication environments. For example, in

the real battle will take place between center-stage players like Toshiba, Zenith and now Compaq [EXTRAPOSITION]

---

[52]Note that although we use instances of the verb *be* in our copula examples, we assume that other verbs are copulas as well, e.g., *seem*, *appear*, *remain*, etc.

[53]Postcopular NPs in in examples like *the teacher is John Smith* are not arguments. As *teacher* is a subject nominalization, the coreference-like relation will identify *John Smith* with the ARG0 (or subject) slot of *teacher*. This clarifies our treatment of cases in which the argument-nominalization status of the prenominal noun is ambiguous, e.g., *the gift of a book* vs. *The book was a gift*. In the latter case *gift* would be marked ARG1-REF as per section 12 and the association of *the book* with the ARG1 slot would also be due to with this coreference-like instance of predication.

1. *The real battle is over who will control the market*
   REL = battle, ARGM-ADV = real, ARG2 = over who will control the market

2. *This picture is about a middle-aged son who makes sure . . .*
   REL = picture, ARG1 = about a middle-aged son who makes sure . . .

3. *The theory is that Seymour is the chief designer of the Cray-3 . . .*
   REL = theory, ARG1 = that Seymour is the chief designer of the Cray-3 . . .

4. *Brooks Brothers' aim is for 20% of total sales to come from the women's department*
   REL = aim, ARG0 = Brooks Brothers', ARG1 = for 20% of total sales to come from the women's department

5. *the idea is to attack first, last and always*
   REL = idea, ARG1 = to attack first, last and always

6. *The primary purpose of a railing is to contain a vehicle*
   REL = purpose, ARG1 = of a railing, ARG2-PRD = to contain a vehicle, ARGM-ADV = primary

Figure 56: Arguments of Nouns that follow a Copula

1. *Trying to time the economy is a mistake* [Clausal Subject Argument]
   REL = mistake, ARG1 = trying to time the economy

2. *it's a mistake to put too much power in the hands of a single person* [Extraposed Clausal Subject Argument]
   REL = mistake, ARG1 = to put too much power . . .

3. They are some distance apart [ADJECTIVE NOUN CONSTRUCTION]
   REL = distance, ARG1 = they, ARG2 = some

4. John is 40 pounds in weight [ATTRIBUTE NOUN CONSTRUCTION]
   REL = weight, SUPPORT = pounds, ARG1 = John, ARG2 = 40 pounds

Figure 57: Arguments of Nouns that are matrix subjects of a Copula

77

REL = battle, ARG0 = between center-stage players . . ., ARGM-ADV = real
The extraposed PP (*between center-stage players . . .*) is linked to *battle* via a grammatical device (extraposition). In contrast, support verb/nominalization combinations are idiosyncratic. Additionally, predication cannot be part of a support chain, whereas support chains of support verbs and PARTITIVE/SHARE constructions are fairly common. The reason predication does not work this way because: (1) Predication links occur by splitting noun heads from their arguments; (2) Support constructions include: verbal support and PARTITIVE/SHARE constructions, but only the latter is a logical candidate for predication links as the former is verbal; and (3) PARTITIVE/SHARE nouns that allow support cannot be linked to their PP arguments by predication, perhaps because their PP arguments are of the *of* variety, e.g., the following sentence is bad: **The millions was of books*.

Note that only arguments should be linked by predication as in figure 56. There are some cases where PPs across copulas are not markable. For example, *about* phrases can be comments on what underlies situations *All life is about change* or *the heated fight over program trading is about much more than a volatile stock market*. In these cases, the *about* phrase does not appear to be an argument of (the head of) subject of the copula.

The examples in figure 57 are each instances of predication in which the argument of the nominal predicate (REL) is the subject of the sentence. These are different than the figure 56 cases in that they are limited by the type of predicate nominal. The first two examples involve nouns like *mistake* which license clausal subjects. These clauses can appear as matrix subjects as in figure 57, example (1) or then can be extraposed as in example (2). In addition, nominalizations of adjectives sometimes allow one of their arguments to occur as the subject of the copular construction (as with the corresponding adjectives). ATTRIBUTE nouns also allow their ARG1 to occur before the copula in the following construction:

ARG1 copula UNIT-NOUN-NP in ATTRIBUTE-NOUN
The unit noun (NUNIT in COMLEX) is transparent in this construction (hence it is marked as SUPPORT in the annotation). These adjective noun and attribute noun constructions take NPs as arguments, in contrast with the post-copular argument constructions (figure 56), where NP arguments are not permitted.

Finally, many preposition plus noun combinations take on adjectival characteristics, thus allowing the subjects to be arguments of nouns in the configuration: SUBJECT BE + PREPOSITION NOUN, as in figure 58. While many of these examples share the syntactic configuration of the subject oriented adverbials described later in section 10, there are some differences. For example, like the chosen examples (but by no means in all cases) these nouns can occur in the singular form without a determiner when they follow these prepositions, but not in other

78

1. *These individuals are under investigation*
   REL = investigation, ARG1 = These individuals

2. *Dozens of additional crews were on alert*
   REL = alert, ARG1 = dozens of additional crews

3. *Mr. Phelan is in control of the Big Board's factions*
   REL = control, ARG0= Mr. Phelan, ARG1 = of the Big Board's factions

Figure 58: Preposition plus Noun Combinations can take subjects

environments – this corresponds to the Prepnoun COMLEX Syntax class.[54]

7   Arguments of Non-head Nouns

As mentioned in Section 3, sometimes complements of modifiers can follow the head of a noun phrase. In the cases given in Figure 7, the modifiers in question were not nouns. However, it turns out that nouns in premodifying position can do this as well (e.g., *parent* in *the parent company of American Airlines*). In fact, it turns out that nouns in premodifying position can take most of the same phrases as arguments that the head noun can.

The examples in figure 59 are like the degree word cases in that the nominal predicate precedes the head noun of the phrase, yet it is the sole item that licenses arguments that follow the head. This is common for RELATIONAL and HALL-MARK nouns, but can occur with nominalizations as well. In the first three examples, the prenominal the relational noun *parent* and the HALLMARK nouns *theme* and *topic* license the PPs following the head nouns – the NPs would be ill-formed if these prenominals were eliminated, e.g., **the company of American Airlines*. The fifth example can be paraphrased as *a bill proposing appropriations for the office of the president*, i.e., it is the sense of bill that is submitted to a legal body before becoming a law, not a bill for services. In the sixth example, **the rates of patients* would be ill-formed without *recovery* – a paraphrase could be *the rate at which patients recover*.

More commonly however, a head noun can share modifiers with prenominals. In figure 60, arguments are shared by the head and a prenominal. In these cases, we mark the head as an instance of SUPPORT.

---

[54]In Comlex Syntax, prepnouns have the feature (COUNTABLE :PVAL list-of-prepositions). For example, *significance* is marked as (COUNTABLE :PVAL ("of")).

1. *the parent company of American Airlines*
   REL = parent, ARG0 = parent, ARG1 = of American Airlines

2. *the flagship banks of New York's Manufacturers Hanover Corp.*
   REL = flagship, ARG0 = flagship, ARG1 = of New York's Manufacturers Hanover Corp.

3. *The theme song for the 1980s*
   REL = theme, ARG0 = theme, ARG1 = for the 1980s

4. *the topic figure of '80s capitalism*
   REL = topic, ARG0 = topic, ARG1 = of '80s capitalism

5. *the appropriations bill for the office of the president*
   REL = appropriations, ARG1-REF = appropriations, ARG2 = for the office of the president

6. *the recovery rates of patients*
   REL = recovery, ARG1 = of patients

Figure 59: Non-head Nouns with Arguments

1. *a new incentive plan for advertisers*
   REL = incentive, ARG3-REF = incentive, ARG1 = for advertisers, Support = plan

2. *cooperation agreements with other companies*
   REL = cooperation, Support = agreements, ARG1 = with other companies

3. *greater buying interest for the precious metal*
   REL = buying, SUPPORT = interest, ARG1 = for the precious metal

4. *a veto threat from Bush*
   REL = veto, Support = threat, ARG0 = from Bush

5. *the fraud suit against Mr. Keating*
   REL = fraud, Support = suit, ARG0 = Mr. Keating

6. *a 83.4 percent interest in this energy company*
   REL = percent, SUPPORT = interest, ARG1 = in this energy company

Figure 60: Non-head Nouns with Shared Arguments

1. *trade groups* [NOM]
   REL = trade, ARG0 = groups

2. *control devices* [NOM]
   REL = control, ARG0 = devices

3. *appeals count* [NOM]
   REL = appeals, ARG2 = devices

4. *average life* [ATTRIBUTE/NOMADJ]
   REL = average, ARG1 = life

5. *the 500 million cubic feet a day capacity pipeline* [ATTRIBUTE]
   REL = capacity, ARG1 = pipeline, ARG2 = 500 million cubic feet a day

6. *age 50* [ATTRIBUTE]
   REL = age, ARG2 = 50

7. *a centerpiece issue* [NOMADJLIKE]
   REL = centerpiece, ARG1 = issue

8. *draft agreement* [VERSION]
   REL = draft, ARG1 = agreement

Figure 61: Non-head Nouns with the Head as an Argument

In addition to posthead phrases, prenominals can also take the head itself as an argument as in figure 61. The ability for a prenominal to take the head of the phrase appears to be idiosyncratic to the noun itself. When nominalizations are premodifiers, the most common role assigned to the head is ARG0 (*trade groups, control devices*), but other roles are possible (*court* is the ARG2 of *appeals* in the phrase *appeals court*). Attribute nouns, nominalized adjectives and related nouns appear to assign the ARG1 role to the head most frequently (*average life*, *the 500 million cubic feet a day capacity pipeline*), but ARG2 is also possible *age 50*.

As the examples above illustrate, arguments of a prenominal noun predicate can occur anywhere within the noun phrase where arguments of the head noun can occur. It turns out that it is also possible for prenominal nouns to participate in support constructions, as exemplified in figure 62.

1. *They placed buy orders for blue chip shares*
   REL = buy, Support = placed + orders, ARG0 = they, ARG1 = for blue chip shares

2. *The two countries are signing a trade agreement*
   REL = trade, Support = signing + agreement, ARG0 = the two countries

3. *South Korea registered a trade deficit*
   REL = trade, ARG1 = South Korea, Support = registered + deficit

4. *The Newspaper recorded circulation gains*
   REL = circulation, SUPPORT = recorded + gains, ARG1 = The Newspaper

Figure 62: Non-head Nouns in Support Constructions

1. *The computer system processes video images*
   REL = system, ARG1 = processes video images

2. *The rule forbids staffers to own competitors' cars*
   REL = rule, ARG1 = forbids staffers to own competitors' cars

Figure 63: Subjects that take their VPs as arguments

8   VP arguments of Sentential Subjects

In rare cases, it would seem that the matrix VP (or predicate) of a sentence should be an argument of its subject. Some examples are provided in figure 63. We have found this phenomenon to occur with certain nouns that can also take sentential complements, e.g., *a system to process images*, *the rule that staffers cannot own competitors' cars*. When the nouns occur as subjects of sentences, the VP plays the same role as these sentential complement.

9   Coordinate Conjunctions and Related Constructions

Treatment of (coordinate) conjoined phrases depends on the scope of the conjunction. When the conjoined phrase forms a single constituent which fills a slot in the noun's argument structure, nothing special needs to be done. For example, the bracketed constituent is marked ARGM-ADV (a sentence-like adjunct) in *The [unnecessary and inappropriate] use of the hospital*. When predicates are conjoined, each predicate licenses a separate noun structure, as shown in figure 64.

82

1. *anger and bitterness at the paper*
   REL = anger, ARG1 = at the paper
   REL = bitterness, ARG1 = at the paper

2. *breadth and depth of the discounting*
   REL = breadth, ARG1 = of the discounting
   REL = depth, ARG1 = of the discounting

3. *neither good working conditions nor good compensation packages*
   REL = conditions, ARG1 = working, ARGM-MNR = good, ARGM-NEG = neither
   REL = packages, ARG1 = compensation, ARGM-MNR = good, ARGM-NEG = nor

Figure 64: Sample Conjoined Predicates

Given a conjoined phrase M modifying a noun N, it is possible for each of the conjuncts of M to bear a different relation to N, i.e., the different conjuncts would need to be assigned a different label. For example, in the phrase *[feminist and civil] rights*, *feminist* bears an ARG0 relation, but *civil* does not bear any (currently markable) relation.

There are a number of other constructions that should be treated similarly to coordinate conjunctions, including:

- comparatives – in *More stocks increased five percent today than decreased ten percent yesterday*, the second instance of *percent* is assumed to take the support verb *increased* and the ARG1 *More stocks*. This treatment is basically the same as with coordinate conjunction.

- range constructions – in *The stocks increased from 5 percent to 10 percent*, each instance of *percent* would take *increased* plus a preposition (*from* or *to*) as a support chain and *The stocks* as an ARG1.

10 NP-External arguments of PP and Adverbial Constructions

This section describes a number of cases of PPs (*at John's request*) and other adverbial containing a noun (*a mile away*). In each case the adverbial contains a predicate noun (*request* or *mile*) that takes arguments external to the PP or adverbial. For example, in *Mary stole the secret recipe at John's request*, *Mary stole*

the secret recipe is the ARG1 of *request*. Similarly, in *A Texan can sniff a phony a mile away*, *A Texan* is an argument (ARG1) of *mile*. These are not instances of support. Rather the ability of nouns to take these external arguments derives from properties of PP and adverbial constructions.

These phenomena can occur when the adverbials modify either sentences (as above) or other nouns, e.g., *participants in the meeting*. When the adverbials modify sentences, the predicate nominal (prepositional object or adverbial noun) either: (i) takes the matrix sentence as an argument; (ii) takes the subject of the matrix sentence as an argument; or (iii) cuts up the sentence into more than one argument. When the PP or adverbial modifies an NP, that NP can be the argument of the predicate nominal. We are actually collapsing a bunch of distinct phenomena under one heading, in part because the distinctions are a little blurry at times.

COMLEX Syntax has a special class of nouns (EXTRAP-P-NOUN-THAT-S) that take sentential complements when they are objects of specific prepositions (usually without a determiner), e.g., *It is of concern that John cheats*. These phrases allow arguments of the noun to occur outside the NP without a support verb – examples are given in figure 65. Each sentence has the following structure:

SBJ COPULA PREPOSITION *NOUN* SENTENTIAL-COMPLEMENT
The subject of the sentence (SBJ) as well and the SENTENTIAL complement are either arguments of the noun (or part of the same argument).

Figure 66 provides example of a parenthetical-like PP construction, containing a noun N, such that the rest of the sentence comprises one or more arguments of N. The head preposition of these PPs are often: *with*, *without* or *at*. The nouns must take some sort of sentential complement. It appears that these are very similar to parenthetical constructions with verbs like *say* and *believe*. In fact, the sentence is even split into two parts in example 3 – we handle this the same way that PropBank handles parentheticals.[55]

In some cases, the sentence may actually comprise two arguments of the REL noun. For example, *with* or *without* plus *help* divide a sentence such that the subject of that sentence is the ARG2 of *help* and the VP is the ARG1. Some examples are provided in figure 67. In these examples, once an annotator identifies the sentence being modified, they must select the syntactic subject and VP as the ARG2 and ARG1 respectively. To avoid spurious ambiguity, we assume that the modified sentence is adjacent to the PP dominating *help*. Thus in example (2), we assume that the VP beginning with *will start* is the ARG1 – we avoid the interpretation that the lower verb beginning with *to climb* is the modified constituent. Similarly, in

---

[55] In our corpus, *behest* only occurs in this construction. This may lead us to believe that *at X's behest* and *at the behest of X* is an idiom. However, a web search reveals that other forms do rarely appear, e.g., *carrying out the behests of the Lord of the World* and *as behested by the client*.

1. *He was under consideration to succeed Joshua Lederberg . . .*
   REL = consideration, ARG1 = he, ARG2 = to succeed Joshua Lederberg . . .

2. *ABC's baseball experience may be of interest to CBS Inc.*
   REL = interest, ARG0 = ABC's baseball experience, ARG1 = to CBS Inc.

3. *how much money is at issue*
   REL = issue, ARG1 = how much money

Figure 65: COMLEX EXTRAP-P-NOUN-THAT-S Constructions

1. *Without question, something intriguing is going on . . .*
   REL = question, ARG1 = something intriguing is going on . . ., ARGM-NEG = without

2. *The court hearing began in early October at the request of Anthony Hazell*
   REL = request, ARG0 = of Anthony Hazell, ARG1 = The court hearing began in early October

3. *some last-minute phone calls that Mr. Bush made (at the behest of some conservative U.S. senators) to enlist backing for the U.S. position*
   REL = behest, ARG0 = of some conservative U.S. senators, ARG1 = some last-minute phone calls that Mr. Bush made + to enlist backing for the U.S. position

Figure 66: Parenthetical PP Constructions

1. They persuaded Mr. Trotter [*-1] to take it back and, with the help of the FBI, taped their conversation with him.
   ARG0 = of the FBI, REL = help, ARG2 = They, ARG1 = taped their conversation with him

2. Analysts insist that even without help from a shaky stock market, which provided a temporary boost for bonds during the Oct. 13 stock market plunge, bond prices will start to climb on the prospects that the Federal Reserve will allow interest rates to move lower in the coming weeks.
   REL = help, ARG0 = from a shaky stock market, ARG2 = bond prices, ARG1 = will start to climb on the prospects that the Federal Reserve will allow interest rates to move lower in the coming weeks., ARGM-NEG = without

3. Their recovery came surprisingly fast, and always with the help of neighbors.
   REL = help, ARG2 = their recovery, ARGM-TMP = always, ARG1 = came surprisingly fast, ARG0 = of neighbors

Figure 67: P + N Constructions with Two NP-external Arguments

example (3), we assume that *their recovery* is the ARG2, rather than assuming that *their* is the ARG2. In both cases, we have been offered alternative interpretations in which constituents that are further down in the tree would be viewed as arguments. We claim that such interpretations, if correct, are implied by reasonable interpretations of our initial analyses. In other words, if one helps X's recovery come, one is also helping X to recover. Similarly, if one helps X start to climb, then X has helped X climb. By positing this syntactic adjacency guideline, we are hopefully constraining annotation in a way that will increase consistency without loss of accuracy.

In some preposition plus noun constructions, the subject of the matrix sentence is an argument of the prepositional object, as exemplified in figure 68. In this sense, these PPs are like other subject-oriented adverbials, e.g., purpose clauses, adverbs like *willingly*, etc. Unlike the previously mentioned examples, these nouns do not take clausal arguments from the main sentence. (NB: The third example shows how the SUPPORT of a partitive noun may interact with this construction.)

These sort of PPs can also modify other nouns as exemplified in the first two examples in figure 69. They can occur inside the NP (the first example) or the argument noun may occur as the subject of a copula (the second example) with the predicate noun contained in a post-copular PP. In the examples we have found,

1. *they exercise both for health and enjoyment*
   REL = enjoyment, ARG0 = they

2. *President Bush gathered for a celebration*
   REL = celebration, ARG0 = President Bush

3. *After hours of conflict and debate, that jury focuses on the facts*
   SUPPORT = hours + of, REL = debate, ARG0 = that jury, ARGM-TMP = hours

4. *Garbage made its debut this fall with the promise to give consumers the straight scoop on the U.S. waste crisis*
   REL = promise, ARG0 = Garbage, ARG1 = to give consumers the straight scoop on the U.S. waste crisis

Figure 68: Subject Oriented P + N Constructions

1. Participants in the meeting
   REL = meeting, ARG0 = participants

2. She was in the race
   REL = race, ARG0 = she

3. *the possibility of a conventional Soviet attack*
   REL = attack, ARG0 = Soviet, ARGM-MNR = conventional, ARGM-ADV = possibility

4. *the bitterness of the battle*
   REL = battle, ARGM-MNR = bitterness

Figure 69: NP-external Arguments and Adjuncts in NPs containing PPs

87

1. *Earnings were in line with expectations.*
   REL = line, ARG1 = Earnings, ARG2 = with expectations

2. *Mitsubishi built the government's dream development in exchange for the official decision to locate Tokyo 's central railway station there*
   REL = exchange, ARG0 = Mitsubishi, ARG1 = built the government's dream development, ARG3 = for the official decision to locate Tokyo's central railway station there

3. *That $130 million gives us some flexibility in case Temple raises its bid.*
   REL = case, ARG1 = That $130 million gives us some flexibility, ARG2 = Temple raises its bid

Figure 70: Nouns in PPs that link NP-external arguments together

the predicate nominal is an event that has participants and the arguments are participants in the event. The third and fourth examples are adjuncts that can occur in these same positions (repeated from figure 37). These are discussed further in section 4.18.

There are a number of complex predicates, typically consisting of a noun, possibly a following preposition and possibly a preceding preposition which also license np-external arguments, e.g., *in line with*, *in exchange for*, etc. We will assume that these are not purely idiomatic because, the same sense of the nouns appears to be available both inside of these constructions and outside. Rather, we will assume that the arguments are markable. These connectives typically link two constituents: the PP following the noun and some constituent preceding the noun. The preceding argument can either be an NP, a VP or the rest of the sentence (similar to parentheticals). In particular, the two associated arguments should either be parallel to each other (like conjuncts of a coordinate conjunction) or related in a causal or temporal way (like arguments of subordinate conjunctions or discourse connectives).[56] Sometimes, the subject of the sentence is also an argument. Some examples are given in figure 70.[57]

Adverbial constructions that are not PPs can also contain nouns that take np-external arguments, as in figure 71. However, we have limited these to the subject-

---

[56]These are treated somewhat like arguments in the Penn Discourse Treebank [21]. The two main differences are: (a) we mark nonclausal arguments; and (b) we do not mark arguments outside the sentence containing the predicate.

[57]These and subsequent adverbial noun constructions were also discussed in section 4.8 in connection with creating dictionary entries for these nouns.

1. *Texans can sniff a phony a mile away*
   REL = mile, ARG1 = Texans, ARG2 = away, ARG3 = a phony

2. *Several phone calls and a visit to his broker's office later, the dentist found out that ...*
   REL = visit, ARG0 = the dentist, ARG1 = to his broker's office

3. *Residential construction was off 0.9% in September.*
   REL = %, ARG1 = Residential construction

Figure 71: Other adverbials with nouns that take NP-external arguments

oriented and connective varieties.[58] These other adverbials typically consist of a noun or NP and an adverb or particle. The types of nouns that typically occur in these constructions are temporal nouns and nouns indicating units of measure or fractions (e.g., percent).

In summary, we aim to mark arguments of a noun N which occur within the sentence containing N, even if they are not within the noun if N is part of an adverbial construction that allows N to take such arguments.

11   Function Tags of NomBank: Adjuncts and Related Issues

The PropBank manual [29] defines a number of function tags which can be added to argument labels (ARG1, ARG2, etc.) as well as the modifier label (ARGM). For argument labels, they are not so common, but for the modifier label they are rarely absent. For NomBank, we assume that function tags are required for ARGM labels.[59] Furthermore, the only function tags we allow for argument labels are -PRD, -REF and hyphen tags (-PRD is part of PropBank. -REF and hyphen tags are introduced below).

---

[58]In the sentence, *John went to the movies every day*, a reasonable stance might be to assume that *John went to the movies* is the ARG1 of *day*. This relation would be similar to the one between *day* and the *of* phrase in *A day of singing and dancing*. However, we believe that this level of redundancy with verbal argument structure is not productive. PropBank would already cover this relation by marking *every day* as an ARM-TMP of *went* and it is unclear what would be gained by marking the adverbial relation again in this other way.

[59]In PropBank, ARGM without a function tag is primarily intended for use with special phrases like postposed relative clauses (*The man arrived who was wearing a big yellow hat*) and other phrases "which are syntactically related to the verb, but have no bearing upon the event structure of that verb"[29],p.10. We assume that such phrases do not modify nouns.

TAG	Function	New to NomBank?
-DIR	Directional	No
-LOC	Locative	No
-MNR	Manner/Evaluative	No
-TMP	Temporal	No
-EXT	Extent	No
-PRD	Predicative	No
-PNC	Purpose	No
-CAU	Cause	No
-ADV	Sentence Adverbial/Focus Markers	No
-NEG	Negative	No
-DIS	Discourse	No
-REF	Outside Reference	Yes
-H0, H1, . . .	Hyphen Tags	Yes

Table 2: Function Tags Used in NomBank

In this section, we redefine each of the function labels used in PropBank with respect to their roles in noun argument structure, in the process defining the markable adjuncts. In sections 12 and 13 we define additional function tags that are specific to NomBank (-REF) and "hyphen" tags. All NomBank tags are listed in table 2.

First we discuss the function tags that were previously used in PropBank and how they should be viewed in NomBank. This entails: (1) elucidating how they are used in PropBank including a specification of any "gray areas"; and (2) a discussion of how they turn up in noun argument structure. Then we discuss some of the annotation issues relating to ARGMs, the class of arguments that are mostly marked with function tags.

11.1 -DIR

In the PropBank, arguments of verbs that specify end points of motion (or giving/receiving) were marked ARG3 and ARG4, without the -DIR function tag, e.g., the bracketed phrases in *the public didn't come [to the market]* and *All came [from Cray Research]* were marked ARG4 and ARG3, respectively. All other PPs and adverb phrases that indicated a direction (physical or otherwise) were marked with the -DIR function tag, e.g., *Father McKenna moves [through the house]*, *Workers dumped large burlap sacks of the imported material [into a huge bin]*, and *to help turn the automotive-parts manufacturer [around]*. Thus -DIR includes both part of the path of motion for motion verbs (other than the end points) and a more gen-

1. *its trip up north*
   REL = trip, ARG0 = its, ARGM-DIR = up north

2. *a trip East*
   REL = trip, ARGM-DIR = East

3. *the flight home*
   REL = flight, ARGM-DIR = home

4. *a return flight*
   REL = flight, ARGM-DIR = return

Figure 72: Some -DIR Constituents

eral path of motion that extends to other verbs. Whether a -DIR is an ARG1-DIR, ARG2-DIR, . . . ARG5-DIR or whether it is an ARGM-DIR depends on the frames for the verbal predicate. Lacking an explicit mention in a frame, ARGM-DIR is assumed.

To a large degree, noun argument mimics the verb argument structure. Nominalizations and their cousins which reflect motion/giving/sending can take source (ARG4) and goal (ARG3) arguments, e.g., *a free trip [from the Bronx, Wedtech's home,] [to Washington, D.C.]* and *[overseas] shipments* (the *to* phrase and *overseas* are ARG4s and the *from* phrase is an ARG3. Other constituents are marked as some sort of -DIR. Figure 72 gives some examples of -DIR constituents.

11.2   -LOC

Location adverbials that are not directional are marked -LOC. This includes places where something happens or where something is located. By "place", we mean both a physical location or a virtual location (*on television*, *in my mind*, etc.). In PropBank, this includes some arguments of verbs (as per their frame files) as well as ARGMs that represent all sorts of locations, both physical and nonphysical, e.g.,

- *Lorillard Inc. . . . stopped using crocidolite [in its Micronite cigarette filters]*

- *United Illuminating is based [in New Haven, Conn.]*

- *it stopped advertising its namesake cigarette brand [on television]*

Examples are provided in figure 73. Locatives can surface in noun argument structure in many forms. Postnominally, they take the same form as they do in verbal

91

1. *the daily television show*
   REL = show, ARGM-LOC television, ARGM-TMP = daily

2. *a flight on the U.S. carrier*
   REL = flight, ARGM-LOC = on the U.S. carrier

3. *our mission here*
   REL = mission, ARG0 = our, ARGM-LOC = here

4. *urban crime* REL = crime, ARGM-LOC = urban

5. *the crash near Detroit Metropolitan Airport*
   REL = crash, ARGM-LOC = near Detroit Metropolitan Airport

6. *Japan's power in the region*
   REL = power, ARG0 = Japan's, ARGM-LOC = in the region

7. *San Francisco's earthquake*
   REL = earthquake, ARGM-LOC = San Francisco

Figure 73: Some -LOC Constituents

argument structure: PPs, locative adverbs, subordinate conjunction phrases, etc. In addition, they can occur in a variety of prenominal positions.

Figure 74 is a set of phrases such that one of their constituents may be mistaken for a locative. In each case, there is some more specific relation or head noun type that marks this constituent as some sort of argument.[60] For each of the figure 74 examples, we have provided one category for the head noun that justifies the argument assignment. In cases where other classifications are possible, the same argument assignment should occur. For example, given that the *language* of a piece of text is an ATTRIBUTE, then the PP in *the language in the legislation* would be marked ARG1, not ARGM-LOC.

11.3   -TMP

Temporal adverbials are assigned -TMP. This includes specific points in time (*June 3, 2003*), time periods (*between June 3 and June 5*), durations (*for three hours*,

---

[60]Note that PropBank and NomBank's interpretation of the -LOC label should not be confused with the -LOC label currently used in the PTB. While PTB's -LOC label is a good indicator, PropBank/NomBank's -LOC is more narrow. Some of the examples in figure 74 are marked -LOC in the PTB.

1. *1992's European economic integration*
   REL = integration, ARG1 = European, ARGM-MNR = economic, ARGM-TMP = 1992's
   integration = nominalization

2. *strength in the defense capital goods sector*
   REL = strength, ARG1 = in the defense capital goods sector
   strength = ATTRIBUTE noun

3. *a deputy minister in the prime minister's office*
   REL = minister, ARG2 = in the prime minister's office
   minister = ActRel Relational noun

4. *the language in the legislation*
   REL = language, ARG1 = in the legislation
   language = a "cousin" of a nominalization for the verb *word*. **Thus this phrase means something like** *wording of the legislation*.

5. *in cases where a woman's life is threatened*
   REL = cases, ARG1 = where a woman's life is threatened
   cases **= PARTITIVE noun**
   cases, **like** *instances*, **is a unit of events.**

6. *a situation where a company pays a premium over market value to repurchase a stake . . .*
   REL = situation, ARG1 = where a company pays a premium over market value to repurchase a stake . . .
   situation = ENVIRONMENT noun

Figure 74: Some Constituents that are Not -LOC

during the meeting), approximate or vague times (*soon*, *previously*), relative times (*before I count to three*, *earlier*), frequencies (*often*, *every now and then*, *ever*, *still*), but not rates of speed (*quickly/quick*), which are assumed to be part of the -MNR class.

ARGM-TMP cases are numerous in PropBank, as exemplified by the bracketed phrases in Figure 75:[61]

Figure 76 provides a sampling of markable -TMP constituents of NPs. As expected, post-nominally these phrases are much like their PropBank counterparts, but prenominally there are some additional cases, e.g., possessives, prenominal noun modifiers and adjectives. It is also clear that care needs to be made in distinguishing temporal from locative constituents, as some of the same prepositions can signal either (*in*, *on*, etc.).

11.4 -EXT

Constituents that measure extent are labeled with the -EXT suffix. An extent is a measure of either an attribute underlying a proposition or a change in that attribute. For example, a journey can be measured in terms of the distance from beginning to end (*distance* is an underlying attribute); an increase in price can be measured by the fractional or additive difference from beginning price to ending price; (*price* is an attribute) etc. Note that durations could in principle be considered extents, we consider them to be temporal (-TMP) as per the previous section.

In PropBank, most -EXT constituents are headed by units, numbers or quantifiers/partitives of some sort, e.g.,

- *Commonwealth Edison closed [at $38.375]*

- *400 taxable funds grew [by $1.5 billion]*

- *Newsweek's ad pages totaled [1,620, a drop of 3.2% from last year]*

- *Magna recently cut its quarterly dividend [in half]*

- *Money Fund Report eased [a fraction of a percentage point]*

For non-numeric characterizations of amounts, there is a subtle distinction between -EXT and some -MNR (manner) constituents and -TMP (temporal) constituents. The reason is that -MNR includes the notions of degree and intensity (see section 11.10), and -TMP includes the notion of frequency. For adverbs that double

---

[61]Note that multiple temporal phrases may modify the same verb. Often these are different subtypes of temporal phrases.

1. *Pierre Vinken , 61 years old , will join the board as a nonexecutive director [Nov. 29]*

2. *[In July], the Environmental Protection Agency imposed a gradual ban on virtually all uses of asbestos*

3. *Compound yields assume reinvestment of dividends and that the current yield continues [for a year]*

4. *portfolio managers can capture higher rates [sooner]*

5. *Assets of the 400 taxable funds grew by $1.5 billion [*U*] [during the latest week]*

6. *The top money funds are [currently] yielding well over 9%*

7. *... to obtain regulatory approval and complete the transaction [by year-end]*

8. *... to act [until next week] [at the earliest]*

9. *the execs squeezed in a few meetings at the hotel [before [*-1] boarding the buses [again]]*

10. *South Korea has recorded a trade surplus of $71 million [so far] [this year]*

11. *However, none of the big three weeklies recorded circulation gains [recently]*

12. *[When we evaluated raising our bid [*T* ← when]], the risks seemed substantial and persistent [over the next five years], and the rewards seemed [a long way out].*

Figure 75: ARGM-TMP phrases from PropBank

1. *the principal fights during the major campaigns*
   REL = fights, ARGM-TMP = during the major campaigns

2. *a meeting with Premier Li Peng on Monday*
   REL = meeting, ARG1 = with Premier Li Peng, ARGM-TMP = on Monday

3. *a meeting after the market closed yesterday*
   REL = meeting, ARGM-TMP = after the market closed yesterday

4. *the early part of this century*
   REL = part, ARG1 = of this century, ARGM-TMP = early

5. *their introduction 10 years ago*
   REL = introduction, ARG1 = their, ARGM-TMP = 10 years ago

6. *yesterday's factory orders report*
   REL = report, ARG1 = factory orders, ARGM-TMP = yesterday's

7. *The Short Term Bond Fund would deliver a total return for one year of about 10.6%*
   SUPPORT = deliver, REL = return, ARG0 = The Short Term Bond Fund, ARG1 = of about 10.6%, ARGM-TMP = for one year

8. *The U.S. Congress and administration need frequent reminders of that responsibility*
   SUPPORT = need, REL = reminders, ARG1 = of that responsibility, ARG2 = The U.S. Congress and administration, ARGM-TMP = frequent

9. *The previous contract between Copperweld's Ohio Steel Tube division*
   REL = contract, ARG0 = between Copperweld's Ohio Steel Tube division, ARGM-TMP = previous

Figure 76: Examples of -TMP Constituents in Noun Argument Structure

1. *The six-mile trip to my airport hotel*
   REL = trip, ARG3 = to my airport hotel, ARGM-EXT = six-mile

2. *John gave me little help*
   REL = help, Support = gave, ARG0 = John, ARG2 = me, ARGM-EXT = little

3. *stronger advances in stock prices*
   REL = advances, ARG1 = in stock prices, ARGM-EXT = stronger

4. *a five cent change in actual earnings*
   REL = change, ARG1 = in actual earnings, ARGM-EXT = five cent

Figure 77: Some Sample -EXT constituents in Noun Argument Structure

as quantifiers and modifiers of quantities (*more*, *less*, *higher*, *lower*), context determines whether they represent *more/less often* (-TMP), *more/less intensely* (-MNR) or *more/less in number* (-EXT). Other non-numeric ways to express extent include: *further*, *a lot*, *a little*, *fractionally*, *normal* and *a little bit*.

Figure 77 includes some sample annotations for noun argument structure that include -EXT constituents. Figure 78 provides a sampling of NPs that contain constituents that are superficially like -EXT constituents, but are not considered so under our guidelines. We assume that -EXT only applies to propositional NPs (nominalizations and their cousins) and (as with the verbs) there is some underlying attribute that is being modified by the -EXT constituent, e.g., in *John gave me little help*, there is an underlying *helpfulness* attribute modified by the ARGM-EXT *little*. In most cases, this underlying attribute is in flux, e.g., *150-point* reflects the extent to which the price attribute changed in *a 150-point drop*.[62]

Post nominal -EXT phrases take many of the same forms as their verbal counterparts: mainly PPs and NPs headed by unit nouns, although in noun structure the PPs can be headed by *of* (not just *by*). Prenominally, -EXT phrases may occur as adjectives or nouns (including possessives). Determining whether a constituent is in fact an -EXT constituent and, not some other sort of adverbial (typically -TMP or -MNR) should be based on the following questions: (1) Does the NP represent an extent as per the above definition? If not, than it is not a -EXT constituent. If maybe, check the other questions; (2) Is the head of the phrase a number, a unit of measure (other than time), or a quantifier? (If yes, then -EXT)?; (3) If the con-

---

[62]We specifically exclude ATTRIBUTE senses of nouns from taking ARGM-EXTs. The ARG2 (value) of such nouns can be too similar to ARGM-EXT argument. For example, consider ARG2s of ATTRIBUTE nouns like *length*, *width*, *price* etc.

1. *[near record] earnings*
   No markable argument structure
   near record **is neither a numerical nor a simple high/low statement about the extent of the earnings**

2. *slower economic growth*
   REL = growth, ARGM-MNR = slower
   slower **is a statement about the speed of growth over time, not the extent of growth over one specific period**

3. *steep increases in foreign assistance and trade*
   REL = increases, ARG1 = in foreign assistance and trade, ARGM-MNR = steep
   steep **is a statement about how these increases compared to previous ones**

4. *a significant reduction of principal and interest*
   REL = reduction, ARG1 = of principal and interest, ARGM-MNR = significant
   significant **states how important the reduction was, not how large.**

5. *further declines in interest rates*
   REL = declines, ARG1 = in interest rates
   further **suggests that there are more declines - the total number of declines does not bear on the extent of any one decline**

Figure 78: Sample Constituents that should NOT be marked -EXT

stituent can quantify over a number of different qualities: time, intensity, speed, amount, etc., is the amount in question one of time (-TMP), intensity or speed (-MNR), or amount (-EXT)?; (4) If the constituent in question is an adverb/adjective, can it be construed as manner (-MNR)? (e.g., *extensive*, *sharp*, etc.) Please note that as, per section 11.10, -MNR is by far the most inclusive of the adverbial classes in PropBank. Therefore, it is very likely that a constituent that seems similar to -EXT, but is not -EXT, should be marked -MNR.

If PropBank annotation is any guide, one must be extremely cautious because the -MNR/-EXT border is precarious and it seems like erring on the side of -MNR is a good idea. Adverbs like *slightly*, *extensively*, *greatly*, *significantly* and *sharply* are predominantly marked as -MNR, not -EXT. A rule of thumb is that adverbials that in any way comment on the extent beyond saying that it is a lot or a little tend to be marked -MNR, as do adverbials that mark rates. Most of the verbs with any sort of -EXT modifier tend to be in the financial sublanguage.

It has turned out that -EXT vs -MNR is a difficult area for NomBank annotation as well. We have developed the following rules of thumb to help us differentiate.

1. There is a small list of non-numerical prenominal modifiers which are typically allowed to be marked -EXT. Currently, this list includes the following: *more, little, large, big, small, far, much, enormous, fractional, huge, infinitesimal, high, further, great, heavy, normal, limited, strong, weak, fat, fatter, fattest*.

2. When possible, an ATTRIBUTE (section 4.18 sense should be added for that noun. Typically, the apparent -EXT modifier will turn out to be a ARG2 of the ATTRIBUTE noun. For example, in *Pickles are in short supply*, *supply* is an attribute noun, *Pickles* is the ARG1 and *short* is the ARG2.

3. It is still possible to mark a prenominal modifier as ARGM-EXT, even if it is not numeric and not on the above list. However, the word must clearly paraphrase *a lot* or *a little* degree. It cannot include other meanings. For example, *widespread panic* suggests that the panic that takes place happens over a wide area – this goes beyond simple extent and would therefore be marked ARGM-MNR.

11.5   -PRD

A constituent X is labeled with the -PRD suffix if: (1) X is a predicate lacking tense; (2) X is not an *as* phrase; and (3) it does not fit the definition of a -PNC constituent as per section 11.6. In PropBank, -PRD includes virtually all infiniti-

val, -ing, copular, and so-called small clause complements of verbs, such as the constituents after the main verbs in the following examples:

- *They considered shorter maturities a sign of rising rates*

- *it expects to obtain regulatory approval*

- *They are still trying to lure back small investors*

In noun argument structure, -PRD is applied exclusively to infinitival and -ing arguments of nominalizations and their cousins. Some examples are provided in figure 79. Argument assignment is based primarily on frame files. Lacking a frame file, we assume that -PRD arguments of nominalizations of verbs lacking NP objects should be ARG1-PRD and those with NP objects should be marked ARG2-PRD. (This follows the pattern of the PropBank verbs.) It is interesting to note that the subject of the -PRD complement in many of these cases can be filled by an argument of the head noun, although such marking is the domain of PropBank. For example, in *John's promise to leave*, PropBank marks *John's* as the subject of *leave*.

11.6 -PNC

Constituents labeled -PNC include: (1) so-called purpose and rationale clauses, types of *to* infinitive (and infinitives beginning with *in order to*); (2) PPs and P-ing phrases beginning with *for* that express purpose; and (3) rarely, other forms. -PNC phrases should not be confused with -CAU phrases representing causes (section 11.7).[63]

Most purpose phrases in PropBank are adjuncts (ARGM-PNC), as in figure 80. However, there are also rare instances of complements of this type. For example, the bracketed phrase in the following example is an ARG2-PNC: *a form of asbestos, once used [to make Kent cigarette filters].*[64] Some instances of purpose phrases that are neither infinitival nor *for* PPs are the bracketed phrases in the following examples:[65]

- *Men are still combing the beach with shovels and hand brushes, [searching for that unusual glint].*

- *to withhold their crops from the marketplace [in the hope of higher prices].*

---

[63]In some future versions of PropBank and NomBank, this tag may change to PRP, which closely abbreviates the word "purpose".

[64]Verbs that have PNC arguments include: *design, exploit, file, incur, refocus, rush, structure, tap, use*.

[65]These examples were provided by Paul Kingsbury.

1. *Prosecutors need court permission to obtain the tax returns of an individual or a business*
   SUPPORT = need, REL = permission, ARG0 = court, ARG2 = prosecutors, ARG1-PRD = to obtain the tax returns of an individual or a business

2. *his promise to make the trains run on time*
   REL = promise, ARG0 = his, ARG2-PRD = to make the trains run on time

3. *The company earlier this year adopted a shareholder-rights plan to ward off unwanted suitors*
   SUPPORT = adopted, REL = plan, ARG0 = the company, ARG1-PRD = to ward off unwanted suitors

4. *Philip Morris's corporate campaign runs little risk of getting yanked off the tube.*
   SUPPORT = runs, REL = risk, ARG0 = Philip Morris's corporate campaign, ARG1-PRD = of getting yanked off the tube.

5. *their only hope of keeping viewers from defecting to cable*
   REL = hope, ARG0 = their, ARG1-PRD = of keeping viewers from defecting to cable

6. *another small encouragement for the Federal Reserve to lower interest rates in coming weeks*
   REL = encouragement, ARG1 = for the Federal Reserve, ARG2-PRD = to lower interest rates in coming weeks

Figure 79: Noun Structure with -PRD Arguments

1. *They devised a 69-point scale . . . [to rate the closeness of test preparatives to the fifth-grade CAT]*

2. *The researchers also pulled off a second genetic engineering trick [in order to get male-sterile plants in large enough numbers to produce a commercial hybrid seed crop.]*

3. *Mr. Cray, who couldn't be reached [for comment] . . .*

Figure 80: -PNC constituents in PropBank

1. *a so-called road show to market the package around the world*
   REL = show, ARGM-PNC = to market the package around the world

2. *a one-day trip to inspect earthquake damage*
   REL = trip, ARGM-TMP = one-day, ARGM-PNC = to inspect earthquake damage

3. *the flow of dollars to the U.S. to fund the deficits*
   REL = flow, ARG1 = of dollars, ARG2 = to the U.S., ARGM-PNC = to fund the deficits

4. *shifts of existing savings by taxpayers in order to cut their tax bill.*
   REL = shifts, ARG1 = of existing savings, ARG0 = by taxpayers, ARGM-PNC = in order to cut their tax bill

5. *the use of its card for retail sales*
   REL = use, ARG1 = of its card, ARG2-PNC = for retail sales

6. *an ambitious schedule to pare its massive debt load*
   REL = schedule, ARGM-MNR = ambitious, ARG2-PNC = to pare its massive debt load

7. *adjustment for inflation*
   REL = adjustment, ARG4-PNC = for inflation

Figure 81: Sample Purpose Phrases in Noun Structure

For NomBank, we restrict the -PNC tag to arguments and adjuncts of propositional nouns. Some examples are found in figure 81, which are patterned after their verbal counterparts. The argument/adjunct distinction is made primarily based on the verbal frame files.

11.7   -CAU

Constituents labeled -CAU include PPs and subordinate conjunction phrases and various participles that have a causal meaning. These typically begin with *as a result of*, *because*, *because of*, *due to*, *as*, *for*, *based on*, *by*, *since*, *under*, among others. In PropBank these are mostly adjuncts (ARGM-CAU), e.g.:

- *Shorter maturities are considered a sign of rising rates [because portfolio managers can capture higher rates sooner]*

1. *modest improvement in discretionary cash flow because of planned asset sales*
   REL = improvement, ARG1 = in discretionary cash flow, ARGM-MNR = modest, ARGM-CAU = because of planned asset sales

2. *Brisk domestic demand due to increasing capital investment*
   REL = demand, ARG0 = domestic, ARGM-MNR = brisk, ARGM-CAU = due to increasing capital investment

3. *distributions based on Swiss laws*
   REL = distributions, ARGM-CAU = based on Swiss laws

4. *transaction costs from its 1988 recapitalization as a result of a $160 million restructuring of its bank debt*
   REL = costs, ARG1 = from its 1988 recapitalization, ARGM-CAU = as a result of a $160 million restructuring of its bank debt

5. *execution under any circumstances*
   REL = execution, ARGM-CAU = under any circumstances

6. *strains in Sino-U.S. ties caused by China's crackdown against pro-democracy protesters in June*
   REL = strains, ARG1 = in Sino-U.S. ties, ARGM-CAU = caused by China's crackdown against pro-democracy protesters in June

7. *The bill's managers face criticism, too, for the unusual number of conditions openly imposed on where funds will be spent*
   SUPPORT = face, REL = criticism, ARG1 = the bill's managers, ARG2-CAU = for the unusual number of conditions openly imposed on where funds will be spent

Figure 82: Sample Cause Phrases in Noun Structure

- *[Due to an editing error], a letter to the editor in yesterday 's edition from Frederick H. Hallett mistakenly identified the NRDC.*

However, PropBank -CAU constituents can also be arguments. Verbs that have CAU arguments include *arise*, *commend*, *condemn*, *criticize* and *fare*. An example follows:

- *A Poquet spokesman, for example, condemns the Atari Portfolio [because it requires three batteries while the Poquet needs only two.]*

We limit -CAU constituents to argument structures for propositional nouns, ATTRIBUTE nouns and nominalized adjectives and we base annotation on practice in PropBank. Some examples are provided in Figure 82.

11.8   -ADV and -DIS

For PropBank, constituents labeled -ADV include various adverbial phrases that fall into two general categories: (1) sentential adverbials other than discourse adverbials; and (2) focus particles (*only, just* in sentences like *He only/just left recently*). The -DIS tag refers to those sentential adverbials that function as discourse adverbials.

Drawing on the COMLEX Syntax manual [31], we provide more precise definitions. As we show, these constituents show up in noun structure as well as sentence structure, although adverbs are often replaced by morphologically related adjectives, e.g., the adverb *probably* corresponds to the adjective *probable*. Many of the more complex adverbial constructions that are marked -ADV in PropBank may not show up in noun structure, e.g., parenthetical clauses of *say* type verbs and phrases consisting of an attitudinal adverb plus "speaking" (*strictly speaking*). Similarly, although sentence adverbial PPs are possible (*in actuality*, *in theory*, etc.), they are quite rare. We do, however, have some examples of PPs beginning with *according to* that modify nouns. Consequently, most of the noun examples of -ADV consist of adjectives that are related to adverbs and focus particles (which are consistently marked as adverbs in the PTB). In addition, the -DIS adjuncts seem fairly rare for noun phrases, e.g., *concurrent*, *consequent*, *subsequent*.

11.8.1   -ADV Sentence Adverbials and their Nominal Counterparts

Sentence adverbs have a number of different definitions in the literature. However, we focus here on essentially the COMLEX Syntax META-ADV class, a semantically rather than a syntactically defined class. Sentence adverbials are those that

modify the proposition rather than the verb or verb phrase. Following Sadock's dissertation [27], a simple intuitive way of looking at the distinction is that sentence adverbs "modify the hypersentence" rather than the event/state represented by the sentence, where if an author Mary produces a sentence like *John is a man*, there is an underlying hypersentence, *Mary wrote that John is a man.*. In a sentence like *Fortunately, John is a man* authored by Mary, the adverb *Fortunately* is really modifying the hypersentence. In other words, sentence adverbs say something about the opinions, attitudes and/or comments of the speaker/writer (Mary in the hypothetical examples) of the sentence, as opposed to something about the event/state described by the sentence. In quoted sentences, these adverbials reflect the speaker of the sentence (the one quoted) rather than the author doing the quoting. Sentence adverbials have the following characteristics in common:

- They do not select for the verb. Thus the sentence adverb *probably* can cooccur with any verb. In contrast, other types of adverbs/adverbials have cooccurrence restrictions. For example, manner adverbs like *clumsily* are ill-formed with stative verbs, e.g., **Mary knows a lot clumsily*. Temporal adverbials must be aspectually compatibility with the verbs they modify, hence the ungrammaticality of **John knows a lot frequently*.

- It is often possible to paraphrase the sentence adverbial using a related adjective or a noun like *fact* modified by a related adjective and putting the rest of the sentence as a sentential complement. For example, *John probably left* and *Unfortunately, John left* mean the same as *It is probable that John left* and *It is unfortunate that John left*. *Legally, John has three parents* could be paraphrased as *It is a legal fact that John has three parents*. This sort of paraphrase is not available for other types of adverbs.[66] The fact that you can pull out the adverbial in this way while maintaining the meaning suggests that the adverb is modifying the whole proposition represented by the sentence.

Figures 83–85 provide some examples of NPs modified by sentence adverbials or related adjectives which belong to the -ADV class. They fall into the following subclasses: epistemic, attitudinal and viewpoint, corresponding to the COMLEX Syntax subclasses of META-ADV.[67]

The examples in figure 83 are epistemics classified in COMLEX Syntax as (META-ADV :EPISTEMIC T). These adverbials comment on the likelihood or

---

[66]Some manner adverbs are also sentential adverbs, e.g., *happily*. However, only the sentential sense of the adverb allows this sort of paraphrase. Thus *It is a happy fact that John laughed* does not mean the same as *John laughed happily* (in a happy manner).

[67]The COMLEX distinctions draw on [8, 9], among others.

1. *probable criminal activity*
   REL = activity, ARG1 = criminal, ARGM-ADV = probable
   (ABILITY noun)

2. *a probable market capitalization between #300 million ($473 million) and #400 million*
   REL = capitalization, ARG1 = market, ARG2 = between #300 million ($473 million) and #400 million ARGM-ADV = probable

3. *a possible U.S. troop reduction in South Korea*
   REL = reduction, ARG1 = U.S. troop, ARGM-LOC = in South Korea, ARGM-ADV = possible

4. *possible repayment by Moscow of $188 million in pre-Communist Russian debts owed to the U.S. government*
   REL = repayment, ARG0 = by Moscow, ARG1 = of $188 million in pre-Communist Russian debts . . ., ARGM-ADV = possible

Figure 83: Epistemic Phrases in Noun Structure

1. *unnecessary burdens for the industry*
   REL = burdens, ARG1 = for the industry, ARGM-ADV = unnecessary

2. *The unnecessary and inappropriate use of the hospital*
   REL = use, ARG1 = of the hospital, ARGM-ADV = unnecessary and inappropriate

3. *the inescapable fact that the transplants are adding capacity*
   REL = fact, ARG1 = that the transplants are adding capacity, ARGM-ADV = inescapable

4. *the extraordinary fact that he hadn't paid his income tax for the previous four years*
   REL = fact, ARG1 = that he hadn't paid his income tax for the previous four years, ARGM-ADV = extraordinary

Figure 84: Attitudinal Phrases in Noun Structure

106

1. *liquidation value assuming the sale of all UAL assets*
   REL = value, ARG1 = liquidation, ARGM-ADV = assuming the sale of all
   UAL assets

2. *alleged violations of industry rules*
   REL = violations, ARG1 = of industry rules, ARGM-ADV = alleged

3. *this large hypothetical seller*
   REL = seller, ARG0 = seller, ARGM-ADV hypothetical

4. *President Bush doesn't have the legal authority to exercise a line-item veto.*
   SUPPORT = have, REL = authority, ARG0 = President Bush, ARG1 = to
   exercise a line-item veto, ARGM-ADV = legal

5. *a psychological lift for the market*
   REL = lift, ARG1 = for the market, ARG-ADV = psychological

6. *(there are) technical, economic, political and psychological reasons for the*
   market's recent drubbing
   REL = reasons, ARG1 = for the market's recent drubbing, ARGM-ADV =
   technical, economic, political and psychological

Figure 85: Viewpoint Phrases in Noun Structure

possibility of truth of the event/state (verb example: *It's probably worth paying a premium for funds that invest in markets*). Care should be taken however, to make sure that it is the proposition that is being modified. In contrast, *probable* in the following example should not be marked: *a more probable inflation estimate*. What is probable here is not whether or not the estimate takes place, but rather whether or not the outcome is accurate.

The examples in figure 84 are ATTITUDINAL modifiers (META-ADV :AT-TITUDE T) – a sentential counterpart is *unfortunately* in *It [unfortunately] encourages others to engage in a highly dangerous and illegal activity that only a very few are doing now.*. These can be deceptively similar to manner adverbials/adjectives.[68] The key difference is that these words express the opinion of the author of the sentence rather than the manner of the event – contrast the manner case, *He danced happily* with the sentence-adverb case *Happily, he was a good dancer*. They may also appear similar to evaluative adverbials,[69] complements of verbs like *behave* and *treat*, e.g., *He treated her well*, *She behaved badly*. One can see that these are two distinct classes because, when they cooccur (*John, unfortunately, behaved badly*), neither effects the meaning of the other. In contrast, using two evaluatives together produces an odd result, e.g., *John behaved badly very well*.- If well-formed, the meaning of the two adverbs clearly interact – *very well* clearly modifies *behave badly* – it means that John did a good job at behaving badly. When these type of modifiers occur as adjectives in NPs, the attitudinal adjective must precede the evaluative – *his unfortunate bad behavior* is well-formed, but **his bad unfortunate behavior* is distinctly odd.

Figure 85 contains viewpoint adverbials (META-ADV :VIEWPOINT T). These express that a statement is relative to a particular way of thinking or a particular discipline (verb example: *paleontologically speaking, "it is, indeed, a wonderful life."*). The postnominal phrase in example 1 is straight-forward – the statement assumes *the sale of all UAL assets* is a necessary precursor. The use of *hypothetical* and *alleged* in 2–3 have a similar feeling to the epistemics. Rather than dealing with probability, these adverbs place events in a possibly-untrue context, similar to the adverb *technically* in *Well, in some ways it is different, but technically it is just the same.*. In the case of *this large hypothetical seller*, the underlying proposition that a sale took place is hypothetical.[70] The remaining examples (4–6) are

---

[68]A subtype of PropBank's -MNR class which includes both rate of speed (*slowly*) adverbials and adverbials that express the style in which an action is performed (*clumsily*, *happily*, *expertly*).

[69]In PropBank, these are sometimes a subtype of -MNR.

[70]There is some ambiguity here of the classical *beautiful dancer* variety, to be discussed further in section 11.10. Either the sale is hypothetical or the existence of the entity engaged in the sale is hypothetical. We believe that this ambiguity is spurious. While those who dance beautifully may in fact be unattractive, if an entity who sells does not exist, there can be no sale.

tricky because these same adjectives may modify nouns without being markable as -ADV. One test for markability is to see if they can in fact be replaced with the sentential adverb. For example, (4) could be paraphrased as *Legally, President Bush doesn't have the authority to exercise a line-item veto.* The word *authority* refers to an ability to do an action and *legal* modifies the way this ability may come about. This contrasts with other instances of the word *legal* that cannot be paraphrased this way, e.g., *the plaintiffs failed to cite any legal authority that would justify such an injunction* and *a vice president asked him to intervene in an unrelated legal dispute involving a trust account.* In these examples, *any legal authority* is something concrete that can be cited (a rule or an entity) and *an unrelated legal dispute* is a dispute about legal matters, i.e., it is not only defined as a dispute once you enter the legal domain. Example 5 can be paraphrased in much the same way – it assumes a sense of *psychological* that is something like *in the realm of the mind.* Example 6 can also be paraphrased by turning the adjectives into adverbs as follows: *Technically, economically, politically and psychologically, there are reasons for the market's recent drubbing.* It is worth noting that all these adjectives have unmarkable instances as well, e.g., *psychological counseling, dubious psychological and sociological theories*, etc.

We suspect that some subtypes of sentential adverbs are simply not possible with NPs. For example, we have not found any performatives (META-ADV :PER-FORMATIVE T) in the annotation, although it is not possible to prove their nonexistence. So for completeness, we are including a brief definition (in case an annotator finds one and wants to mark it.) These adverbs express the speaker's state of mind or manner of speaking, e.g., *Frankly speaking, the U.S. was involved too deeply in the turmoil and counterrevolutionary rebellion which occurred in Beijing not long ago..* In this example, the speaker is telling his audience that he is being frank.

11.8.2   FOCUS particles

FOCUS particles are marked ARGM-ADV in PropBank. They include instances of *even*, *just*, *only* and possibly a few other words. Their purpose is primarily to place emphasis and they can modify many different constituents including NPs and verbs. Examples of NP modification are provided in Figure 86. PropBank clausal examples include: *One official newspaper, Legal Daily, [even] directly criticized Mr. Nixon.* and *we [just] need to understand it.* In most cases, the same word is used in both the noun case and the clausal case. However, there are at least two adjectives (*mere, sheer*) that can play this role as well.

1. *The mere fact of a meeting*
   REL = fact, ARG1 = of a meeting, ARGM-ADV = mere

2. *even the 25% market share that Nissan expects in 1989*
   REL = share, ARG1 = market, ARGM-ADV = even

3. *just the first leg of an aggressive three-to-five-year direct marketing plan*
   REL = leg, ARG1 = of an aggressive three-to-five-year direct marketing plan, ARGM-ADV = just

4. *only the second president of Random House since it was founded in 1925*
   REL = president, ARG1 = of Random House, ARG-TMP = since it was founded in 1925. ARGM-ADV = only

Figure 86: Sample Focus Particles in Noun Structure

1. *concomitant threats to cease making new loans* REL = threats, ARG1-PRD = to cease making new loans, ARGM-DIS = concomitant

2. Their related transaction with Acme Inc. REL = transactions, ARG0 = their, ARG2 = with Acme Inc., ARGM-DIS = related

Figure 87: Sample -DIS Adjuncts in Noun Structure

11.8.3   -DIS Sentence Adverbials and their Nominal Counterparts

Of the various types of sentential adverbs, the ones that relate sentences to each other or to the discourse, e.g., *however*, *therefore*, etc. are marked -DIS in Prop-Bank. Like the -ADV adverbials they can cooccur with any verb. They are fairly rare with nominals, but we have found a few cases as shown in figure 87. In these examples, the adjective relates the predicate to some previous part or the discourse. Thus *concomitant* suggests there is some event that these threats occur with and *related* suggests that this transaction is related to some other event.

11.9   -NEG

In PropBank, ARGM-NEG marks adjunct instances of *neither*, *nor*, *no*, *n't* and *not*. For noun phrases, negation appears in the determiner slot as either *neither* and *no*, and can also be realized as a preposition, like *without*, or *out of*. It can also appear as a superordinate noun like *lack*. Examples are given in Figure 88. Please take note of the following: (1) For *neither/nor* conjoined constructions, *neither* is

1. *neither good working conditions nor good compensation packages*
REL = conditions, ARG1 = working, ARGM-MNR = good, ARGM-NEG = neither
REL = packages, ARG1 = compensation, ARGM-MNR = good, ARGM-NEG = nor

2. *neither side in the debate*
REL = side, ARG1 = in the debate, ARGM-NEG = neither

3. *no information on whether users are at risk*
REL = information, ARG1 = on whether users are at risk, ARGM-NEG = no

4. *no question that some of those workers and managers contracted asbestos-related diseases*
REL = question, ARG1 = that some of those workers and managers, ARGM-NEG = no

5. *It rained without anyone's help* REL = help, ARG0 = anyone's, ARG1 = it rained, ARGM-NEG = without

6. *Federal credit is out of control*
REL = control, ARG1 = federal credit, ARGM-NEG = out of

7. *the investors' lack of control*
REL = control, Support = lack of, ARGM-NEG = lack of, ARG0 = investors'

Figure 88: Examples of Negation in Noun Structure

assumed to negate the first conjunct and *nor* the subsequent ones; and 23) Negation inside the NP is marked as part of noun structure, but negation that is a sister to a support verb is not.

11.10 -MNR

In PropBank, ARGM-MNR is used to classify a wide variety of adjunct modifiers of verbal argument structure, in fact it includes all adjuncts that are not part of another adjunct class.[71] In this section, we break down the -MNR class into a few

---

[71]While the PropBank corpus includes some rare instances of ARG1-MNR, ARG2-MNR and ARG3-MNR, we assume that these are errors as none of the frames files assigns -MNR to any constituent.

subcases and describe how these subcases fit into noun structure. The significance of these descriptions is to enable the annotator to distinguish modifiers that belong to the -MNR class from those that do not – the goal of distinguishing the subtypes of -MNR from each other is unimportant for this annotation effort.

Identifying -MNR phrases is essential to insure compatibility between Nom-Bank and PropBank. As discussed in section 3.6, we do not mark noun-specific modifiers in NomBank, including most determiners and many adjectives. One of the areas that requires the utmost care on the part of these specifications regards the distinction between adjectives that should be left unmarked and those that should be marked -MNR.[72] Figure 89 exemplifies various verbal adjunct categories that are assumed to make up almost all instances of *ARGM-MNR*.[73] We relate each of these subclasses to noun argument structure.

Figure 90 contains examples of noun annotation with EVALUATIVE ARGM-MNR constituents. EVALUATIVE adverbials correspond to the COMLEX Syntax class EVAL-ADV. These include adverbs that can occur as complements of verbs like *behave*, *treat*, *work*, *act*, *react* and *run*. Evaluatives can also modify the degree of success or failure, as adjuncts of other verbs. When these phrases occur as complements in PropBank, they are usually marked as arguments without the -MNR label, e.g., they are marked the ARG1 of *behave*. COMLEX EVAL-ADV adverbs include *well*, *badly*, *poorly*, *unsuccessfully*, among others. The noun examples mostly use the adjectives that correspond to these adverbs. Argument roles take precedence over ARGM-MNR. Thus *beautiful* in *the beautiful look of wool* would be marked ARG1, not ARGM-MNR.

If a markable NP contains an evaluative adjective modifier, we mark it ARGM-MNR. However for non-propositional nouns, the presence of such a modifier is not sufficient for deeming that NP markable. For example, the NP *a perfect match* (figure 90 example 4) is markable because it has one markable ARGM and *match* is propositional – in this case a "matching event" is presupposed. In contrast, *this dreadful group* is not markable because it is not propositional – there is no presupposed event or state associated with this NP;[74]

Figure 91 includes some temporal-like adjectives that are classified as ARGM-MNR. We will call these MANNER-TMP. For verbs, these include adverbs and PPs expressing rate of speed (*slowly*, *at five miles per hour*); point of time (*suddenly*); or pace (*unceasingly*, *continuously*). In COMLEX Syntax this type of adverb is

---

[72]In contrast, PropBank annotators could default to -MNR if an adverbial did not seem to fit into a different category.

[73]We do not claim to be exhaustive. We believe that these subtypes make up over 95% of the instances.

[74]As previously stated, a propositional NP represents an event, relation or state. This includes argument nominalizations which also represent a participant in that event/relation/state.

1. *That arrangement apparently has worked [well]*
   EVALUATIVE

2. *to seek, [unsuccessfully], a seat in Canada's Parliament*
   EVALUATIVE

3. *the nation's industrial sector is now growing [very slowly]*
   MANNER-TMP

4. *their purchases are growing [at a rapid rate]*
   MANNER-TMP

5. *33 men who worked [closely] with the substance*
   MANNER

6. *On Wall Street men and women walk [with great purpose]*
   MANNER

7. *asbestos was used [in very modest amounts]*
   MEASURE

8. *Commonwealth Edison closed at $38.375, [down 12.5 cents]*
   MEASURE

9. *players marching abroad [with their business]*
   CONCOMITANT

10. *[with a police escort], busloads raced to the Indianapolis Motor Speedway*
    CONCOMITANT

11. *researchers produced flashes of light in the visual field [with magnets]*
    INSTRUMENTAL

12. *defendants who commit a pattern of crimes [by means of a criminal enterprise]*
    INSTRUMENTAL

Figure 89: ARGM-MNR Examples from PropBank

1. *neither good working conditions nor good compensation packages*
   REL = conditions, ARG1 = working, ARGM-MNR = good, ARGM-NEG = neither
   REL = packages, ARG1 = compensation, ARGM-MNR = good, ARGM-NEG = nor

2. *unsuccessful efforts to sell the venerable newspaper*
   REL = efforts, ARG1 = to sell the venerable newspaper, ARGM-MNR = unsuccessful

3. *poor performance of its LaSalle I nuclear plant*
   REL = performance, ARG0 = of its LaSalle I nuclear plant, ARGM-MNR = poor

4. *a perfect match*
   REL = match, ARG1 = match, ARGM-MNR = perfect

5. *a wonderful study of Victorian marriage*
   REL = study, ARG1 = of Victorian marriage, ARGM-MNR = wonderful

6. *20 or so lovely minutes of drifting above the Vosges*
   REL = minutes, ARG1 = of drifting above the Vosges, ARGM-MNR = lovely

7. *the worst trade performance*
   REL = performance, ARG0 = trade, ARGM-MNR = worst

Figure 90: Sample Evaluative Modification in Noun Structure

114

1. *continuous economic expansion*
   REL = expansion, ARG1 = economic, ARGM-MNR = continuous

2. *the sharp and rapid appreciation of the yen*
   REL = appreciation, ARG1 = of the yen, ARGM-MNR = sharp and rapid

3. *a sudden burst in the demand for sterling*
   REL = burst, ARG1 = in the demand for sterling, ARGM-MNR = sudden

4. *a crash test at 30 miles per hour*
   REL = test, ARGM-MNR = at 30 miles per hour

Figure 91: Sample MANNER-TMP instances of ARGM-MNR

marked (MANNER-ADV :TEMP T). In noun structure, adjectives and PPs are possible manifestations.

The next subclass, which we will call MANNER, corresponds to the COMLEX Syntax class (MANNER-ADV). This covers adverbials that answer the question *How does X VERB?*, but are neither EVALUATIVE nor MANNER-TMP cases, as described above. For example, the question *How does he ski?* may be answered with an EVALUATIVE adverb (answer = *very well*); a MANNER-TMP adverb (answer = *slowly*); or other adverbials including *in a very silly manner*, *like a clown* and *wiggly*. As modifiers of verbs, MANNER adverbials include a wide range of adverbs including *quietly*, *happily* and *clumsily*. Many MANNER adverbs impose limitations on the aspect of the verb, e.g., they don't tend to co-occur with stative verbs. Some MANNER adverbs place restrictions on the subject of the sentence, e.g., requiring it to be either sentient (a human, animal or organization), capable of motion, able to communicate an idea, etc. In NPs, MANNER tends to be expressed by adjectives (related to MANNER adverbs). Examples are provided in figure 92.[75] We have found at least one example of MANNER adjectives with the *beautiful dancer* ambiguity: *the flamboyant market seer*. In this phrase, it is unclear whether the seer's personal style is flamboyant or whether the seer makes predictions in a flamboyant manner. Nevertheless, we assume that *flamboyant* is markable as ARGM-MNR.

Figure 93 includes examples of adverbials which provide some sort of measurement, but do not qualify as ARGM-EXT. These include various phrases that answer the question *How much?* and measure some implied attribute. Some of these play a similar role that degree words do inside quantifier and adjectival phrases. The

[75]Example 3 contains a second ARGM-MNR constituent: *hand*, an instrumental.

115

1. *a careful reader of the Stolen Art Alert*
   REL = reader, ARG1 = of the Stolen Art Alert, ARGM-MNR = careful

2. *the quiet exodus down the ramps*
   REL = exodus, ARGM-DIR = down the ramps, ARGM-MNR = quiet

3. *his usual clumsy hand gestures*
   REL = gestures, ARGM-TMP = usual, ARGM-MNR = clumsy, ARGM-MNR = hand

4. *They are keeping a close watch on the yield on the S & P 500*
   SUPPORT = keeping, REL = watch, ARG0 = they, ARG1 = on the yield on the S & P 500, ARGM-MNR = close

Figure 92: Sample MANNER Instances of ARGM-MNR in Noun Structure

subtle distinction between these MEASURE modifiers (labeled ARGM-MNR) and ARGM-EXT constituents is discussed in section 11.4.

In Figure 94, are instances of noun structures containing CONCOMITANT adjuncts of nouns. During an event, the CONCOMITANT phrase identifies what or who accompanies the event's main participant (typically ARG0). These seem to only modify events.[76]

Figure 95 contains INSTRUMENTAL ARGM-MNR constituents. INSTRUMENTAL phrases indicate tools or means for an agent to do some action. Hence these modifiers occur in propositional NPs, but not in other NPs. It is sometimes difficult to determine if an NP contains an INSTRUMENTAL phrase and no agent or if the reverse is true. We know that figure 95, example 1, contains an INSTRUMENTAL because we can find the agent in a wider context: *Thus armed for massive matching of documents by computer, they single out high-income groups, looking primarily for people who haven't filed New York income-tax returns.* Without additional context, it is really hard to tell due to the "Instrument Subject" alternation (cf. [13]).[77] Most INSTRUMENTAL modifiers of nouns seem to be *by* phrases and prenominal noun modifiers.

In the beginning of this subsection, we mentioned that we were concerned about identifying the different subcases of ARGM-MNR, but we were not so interested in distinguishing between subcases. This stance actually promotes greater consistency as it is not always easy to divide the subcases mentioned above. For

---

[76]In principle, the main participant need not be expressed, e.g., the phrase *frequent journeys with friends* is well-formed even if the traveler is unexpressed.

[77]In example 4, *repeated* is of type MANNER-TMP.

1. *a full membership on the exchange*
   REL = membership, ARG1 = on the exchange, ARGM-MNR = full

2. *Healthcare hasn't made complete rent and mortgage payments since July*
   SUPPORT = made, REL = payments, ARG0 = Healthcare, ARG1 = rent and mortgage, ARGM-MNR = complete, ARGM-TMP = since July

3. *Michael Carpenter will take complete control of Kidder*
   SUPPORT = take, REL = control, ARG0 = Michael Carpenter, ARG1 = of Kidder, ARGM-MNR = complete

4. *slight differences in the way human and animal insulins drive down blood sugar*
   REL = differences, ARG1 = in the way human and animal insulins drive down blood sugar, ARGM-MNR = slight

5. *the strongest dividend growth*
   REL = growth, ARG1 = dividend, ARGM-MNR = strongest

6. *minor property damage*
   REL = damage, ARG1 = property, ARGM-MNR = minor

Figure 93: MEASURE instances of ARGM-MNR Constituents in Noun Structure

1. *Nomura started a credit-card venture with American Express Co.*
   SUPPORT = started, REL = venture, ARG0 = Nomura, ARGM-MNR = with American Express Co.

2. *he took frequent junkets with friends to exotic locales*
   SUPPORT = took, REL = junkets, ARG0 = he, ARG4 = to exotic locales, ARGM-MNR = with friends

Figure 94: Concomitant Phrases in NP Structure

117

1. *massive matching of documents by computer*
   REL = matching, ARG1 = doc, ARGM-MNR = by computer

2. *an "inside" adjuster, who *T* settles minor claims and does a lot of work by phone*
   SUPPORT = does+lot, REL = work, ARG0 *T* ← who ← an "inside" adjuster, ARGM-MNR = by phone

3. *program traders*
   REL = traders, ARG0 = traders, ARGM-MNR = program

4. *repeated gun robberies*
   REL = robberies, ARGM-MNR = repeated, ARGM-MNR = gun

Figure 95: INSTRUMENT Phrases in Noun Structure

1. *the current and rather confused policies of perestroika*
   REL = *policies*, ARGM-TMP = current, ARGM-MNR = rather confused, ARG0 = of perestroika

2. *little or no integration*
   REL = integration, ARGM-EXT = little, ARGM-NEG = no

3. *long and costly effort* REL = effort, ARGM-TMP = long
   NB: *costly* is not assigned any role

Figure 96: When Coordinates take Different ARGM Roles

example, it appears that the adjectives in the following examples are both EVAL-UATIVE and MEASURE instances: *a silly price for a jaguar*, *decent bids* and *a more thoughtful, complete and competitive proposal*. Blurring this distinction results in an easier task and a more consistent result.

11.11   ARGMs and Coordination

When coordinated constituents modify nouns, the entire phrase is typically assigned a single NomBank role. For example, the *of* phrase would get an ARG1 in the phrase *The discovery [of bones and pottery*. In contrast, ARGM modifiers can take different roles. In addition, there are cases where only a subset of the coordinates take any (markable) NomBank role. Figure 96 provides some examples.

11.12 Adjunct Dictionaries

The adjunct dictionaries ADJADV and NOMADV are included with the Nom-Bank package. These dictionaries are semi-automatically derived from COMLEX Syntax entries of other parts of speech. These are intended to help determine the likely ARGM status of adjective (ADJADV) and noun (NOMADV) left modifiers of nouns. In fact, these dictionaries are being developed along side the annotated corpus and checking compatibility with these dictionaries is one of our quality control measures. ADJADV consists of adjectives with a number of different types of lexical entries, but mostly ADJADV and ADJADVLIKE entries. NOMADV consists of mostly NOMADV and NOMADVLIKE entries. ADJADV, ADJADV-LIKE, NOMADV and NOMADVLIKE entries include both word correspondence information and semantic class information.

ADJADV/ADJADVLIKE entries includes a correspondence between the adjective (:ORTH) and a related adverb. NOMADV/NOMADVLIKE includes a similar correspondence between a premodifying noun and a corresponding adverb. In the case of ADJADV and NOMADV entries, the connection with the adverb is a morphological one (*bad ↔ badly*, *amateur ↔ amateurishly*). In the case of ADJADVLIKE and NOMADVLIKE entries, the correspondence is based on distribution. For example, the adjective *hefty* in *the hefty improvement* has the same (degree type of) ARGM-MNR relation to *improvement* as *immensely* does in the sentence *It improved immensely*. Similarly, *blanket* maps to *broadly*, so that *blanket coverage* involves something covering something else *broadly*. While these correspondences usually involve near-synonyms, the minimum requirement is that they be in the same class. Thus users should not assume that these are synonymous, e.g., antonyms often belong to the same semantic class.

The semantic class information characterizes the adverbial relation in terms of COMLEX Syntax adverbial classes (cf. [31]). The following correspondences (approximately) hold:

COMLEX Adverbial Features in PropBank/NomBank	
Feature	**ARGM**
(META-ADV :CONJ T)	ARGM-DIS
other META-ADV	ARGM-ADV
MANNER-ADV, DEGREE-ADV, EVAL-ADV	ARGM-MNR
LOC&DIR-ADV	ARGM-LOC, ARGM-DIR
TEMPORAL-ADV	ARGM-TMP

It should also be noted that the combination of these dictionary entries and the ARGM tags can be used to help with many of the above-described ambiguities in annotating ARGM-EXT, ARGM-MNR, etc. For example, an ARGM-MNR tag for

119

an adjective marked as DEGREE-ADV in the dictionary will usually correspond to an extent-like meaning of that adjective. Observe that the ARGM-MNR tag for this adjective will correspond to this DEGREE-ADV meaning. Thus if one wanted to group ARGM-MNR/DEGREE-ADV annotation with ARGM-EXT annotations, this would be possible. Similarly, there is a (MANNER :TEMP T) subclass of the MANNER-ADV feature which corresponds to manner adverbs with some temporal meaning.

We have also included some nonderivational classes in ADJADV. These are listed below, along with the typical NomBank role assigned to them, or in the case of argument roles, typical descriptors associated with those roles (cf. section 4).

More ADJADV Classes			
Class	**Probable Tag/ Role Descriptor**	**Description**	**Example**
ADJVERB	ARGM-PNC	Related to Verb	*acquisitive* (verb = *acquire*)
ADJVERBLIKE	ARGM-PNC	Like ADJVERB	*mercenary* (verb = *hire*)
FOCUSADJ	ARGM-ADV	Focus adjectives	*only*
LOCADJ	ARGM-LOC	Locative Adjective	*rural*
NATIONALITY	ARGM-LOC	Nationality Adjective	*Uruguayan*
ORGADJ	ARG0, TOPIC, RECIPIENT, BENEFICIARY	organization adjective	*democratic*
STYLEADJ	ARGM-MNR	style-of adjectives	*orwellian*
TOPICADJ		ARGM-MNR, TOPIC	*financial*

12 A New Function Tag: -REF

In NomBank, we use one function tag that is not currently being used in PROP-BANK: -REF. ARG0-REF, ARG1-REF or ARG2-REF are sometimes assigned to the head of the NP. -REF has a meaning something like "this phrase functions as argument X for referential purposes only". In other words, -REF argument slots effect the interpretation of the NP referentially, but does not effect argument assignment within the NP. This section discusses our usage of -REF and provides examples.

12.0.1　The Defaults

The coreference/predication properties of argument taking NPs are somewhat predictable. Argument nominalizations "act" like the incorporated argument with respect to coreference. For example, if we assign ARG0 and ARG1 to *teacher* and *nominee* respectively, the whole NP is identified with this role as well with respect to coreference and predication. Consider the sentence *She was Fred's nominee*. In NomBank, we indicate that *Fred* is the ARG0 (the nominator) and *nominee* is both the REL (the predicate) and the ARG1 (the person nominated). The copular construction indicates that *She* is identified with the phrase *Fred's nominee*. The link between *She* and the ARG1 of *nominee* can be inferred by the processes used to interpret predication.[78] We treat RELATIONAL and HALLMARK nouns essentially as subject (ARG0) nominalizations (sections 4.10 and 4.12). Therefore, ARG0s cannot occur with these nouns as well. For example, in *John is the president of the United States*, *president* is both the REL and the ARG0, not *John*.

　For non-argument nominalizations, e.g., *destruction*, the noun phrase acts either like the state/event represented by the nominalization or some other object that is the result of the event represented by the nominalization, but is not an argument, e.g., *John's examination* may be the piece of paper with the answers on it that resulted from John being examined. As these referential properties (proposition vs. result) are disjoint from the set of arguments that the noun or related predicate takes, we do not attempt to distinguish between them in this project. It is our understanding that future annotation efforts involving coreference may attempt to characterize these and other distinctions relevant to coreference (e.g., aspect).

12.0.2　When we use -REF

There are some cases where, for reference purposes only, an argument slot of a noun is filled by the entire NP, even if that slot can also be filled internal to the NP. For example, NPs headed by *donation*, *gift* and *accomplishment* allow the ARG1 slot to be filled by *of* phrases or prenominal modifiers. Nevertheless such NPs act like the ARG1 of *donate*, *give* and *accomplish* with respect to coreference and predication. For example, consider *land donations* and *gift* in the following larger context:

> *For now, Goodwill's Mr. Wadsworth lectures other charities about accepting land*
> *donations. "It's supposed to be a gift," he says. "But it can turn out to be a*

---

[78]Other predication constructions include apposition, extraposition, clefting, ... Processes that interpret predications may have to make various distinctions that are orthogonal to NomBank, e.g., the distinction between identity relations (*She is the president of the United States.*) and "X is an instance of Y" cases (*She is a movie director*).

1. *His recent speech, provocatively titled ["Blacks? Animals? Homosexuals? What is a Minority?"] caused an uproar when* **its title** *leaked out.*
   REL = title, ARG1 = its, ARG2-REF = title

2. **One troubling aspect of DEC's results**, *analysts said, was [its performance in Europe].*
   REL = aspect, ARG1 = of DEC's results, ARG2-REF = aspect

Figure 97: Examples of -REF Defaults

liability."

Land donations includes *land*, an ARG1. However, the two instances of *it* and the one instance of *a gift* are coreferential with this phrase.[79] This coreference is possible because *Land donations* externally has an ARG1 meaning – it refers to the land that was donated. Even without an actual ARG1 *gift* also has this meaning and hence would be marked ARG1-REF (to show that the word includes the meaning of the thing given). Thus, when the argument in question occurs inside the NP, as in *land donations* or *gift of a book*, then no -REF argument need be assigned. When the argument is not present, the -REF argument is assigned.

Although -REF features may occur for any markable NP (nominalization, nominalized adjective, etc.), we have found one generalization – ATTRIBUTE nouns and SHARE nouns tend to be ARG2-REF – this is a tendency, not a defining attribute of these classes. In figure 97, the bracketed phrases are identified with the ARG2-REF argument of the bold-faced NP by coreference or predication. Other noun types do not seem to have a -REF characterization as a default.[80] While predication/coreference properties are perhaps the most observable instantiations of what we are trying to represent with the -REF tag, there are other instantiations. For example, it is clear that what is being raised in the following sentence is the ARG2 of *capacity*: *they raised the electrical current-carrying capacity of new superconductor crystals by a factor of 100*, i.e., the value is increasing.

Some clarifications are needed regarding links by predication, as exemplified by the examples in figure 98. Above, we assume that predication can be interpreted

---

[79]To be specific, this is type coreference, rather than token coreference. [22] calls this type of coreference identity of sense, in contrast with identity of reference (typical anaphora).

[80]Partitive nouns act like quantifiers. Like quantifiers, in some cases they represent one whole, e.g., *The amount of five dollars*. Some such cases may be marked ARG1-REF, e.g., *The amount was five dollars*, depending on their behavior. Many words in this category do not share this behavior. For example *dozens of bananas* is not classified as ARG1-REF as evidenced by the illformedness of *The dozens was bananas.*

1. *John's gift to Mary was a book*
   REL = gift, ARG1-REF = gift, ARG0 = John, ARG2 = to Mary

2. *John's behavior was awful*
   REL = behavior, ARG0 = John, ARG1 = awful

3. They are some distance apart [NOMINALIZED ADJECTIVE CONSTRUC-
   TION] REL = distance, ARG1 = they, ARG2 = some

Figure 98: Examples and Non-Examples of -REF predication

as coreference when an NP is predicated of another NP. However, we do not ex-
tend this interpretation of coreference to constituents other than NP. In figure 98
(1), *a book* is coreferential with the phrase *John's gift to Mary* (gift is assigned the
ARG1-REF role). This contrasts with (2), where the adjective *awful* is assigned the
ARG1 role directly as per the discussion in section 6. Section 6 further differenti-
ates examples like (3). *Distance* is an adjective nominalization occurring after the
copula. *they* is the ARG1 of *distance* due to the adjectival properties of *distance*,
i.e., adjectives normally occur in the copula construction selecting for the subject.
Thus this is also an instance of coreference-like predication.

This -REF feature interacts with our definition of support verbs. Without this
feature, we might be inclined to mark copulas as support verbs in cases like *The
book was a gift for Mary*. Under our analysis, *gift for Mary* is assumed to be an
ARG1-REF and is identified with *the book* as a result of predication. Without the
-REF feature, *be* would fit our definition of a support verb and we would have to
assume that *the book* was the ARG1 of *gift*. However, some predications (e.g.,
apposition) are not licensed by specific lexical items and these could not be treated
as instances of support. Our analysis has the advantage of treating all forms of
predication uniformly.

13 Incompatibilities with Treebank and Related Issues

This section discusses cases in which a satisfactory analysis requires that an anno-
tator rethink or reinterpret a PTB tree. We cover the following issues: (1) What
should an annotator do if he/she disagrees with one of PTB's attachment deci-
sions?; (2) What should one do if two constituents fill distinct slots in argument
structure, but PTB treats them as a single token because they are joined together
by a hyphen?; (3) How should the annotator treat sentential modifiers of apposi-
tion?; and (4) How should the annotator group conjoined prenominal modifiers in

123

1. *Heightened Japanese interest in American small business parallels an acceleration of [investments [giving Japanese companies control of large, highly visible U.S. corporations, . . .]]*
   REL = acceleration, ARG1 = of investments

2. *approximately 27,500 acres of timberland near Truckee, Calif.* REL = acres, ARG1 = of timberland, ARGM-LOC = near Truckee, Calif.

3. *English-Spanish translations*
   REL = translations, ARG3-H0 = English-Spanish, ARG2-H1 = English-Spanish

4. *government-controlled prices* REL = prices, Support = controlled, ARG0 = government

5. *Down's Syndrome, the leading cause of mental retardation, according to an NIH summary.*
   REL = cause, ARG1 = mental retardation

6. *neither the Contra nor the SDI cause*
   REL = cause, ARG2 = Contra + nor + SDI

Figure 99: Incompatibilities with Treebank

difficult cases?. All of these issues are represented by the examples in the figure 99.

In the PTB analysis of example (1), the bracketed -ing clause is attached as a sister of *investments* – this suggests that the -ing clause is a reduced relative clause that modifies *investments*. Suppose that an annotator believed that the PTB analysis entailed that each of the investments that is accelerating is giving Japanese companies control. Furthermore, suppose the annotator held that one of the following two analyses were correct:

- The -ing clause should be attached as a sentence adjunct with the interpretation that "the heightened interest" gives Japanese companies control.

- The reduced relative should be attached to the higher NP with the interpretation that the "acceleration" gives Japanese companies control.

Under either analysis, the -ing clause would not be part of any markable constituent. Thus for the annotator to follow his/her intuitions, they would have to tag the ARG1 as the concatenation of two consecutive PTB constituents: the terminal node *(IN of)* and the non-maximal NP node *(NP (NN acceleration))*. Similarly in

124

example (2), the locative PP *near Truckee* modifies the NP headed by *timberland* in PTB. Suppose an annotator thought that it should modify the NP headed by *acres*. The theory might hold: (1) that the plot of land represented by the partitive is near Truckee; and (2) that the PTB analysis implies that the partitive is a measure of an amount of "timberland near Truckee". To capture what the annotator feels is the correct interpretation, he/she must mechanically fill the ARG1 slot with a concatenation of the two smaller constituents represented by *(IN of)* and *(NP (NN timberland))*.

The second problem (examples 3 and 4) has to do with how PTB deals with hyphenation. In the relevant cases, PTB marks two constituents as one. While this did not matter very much for PTB's original goals, it does matter for NomBank because unless we make adjustments, two separable slots would be filled by the same constituent. To remedy this, we add a new set of function tags {H0, H1, H2, H3, H4}. H0 will refer to the first constituent in a hyphenated string, H1 will refer to the second, etc. Word compounds like "auto-maker" (REL-H1 = auto-maker, ARG0-H1 = auto-maker, ARG1-H0 = auto-maker) combine two words via a hyphen, and therefore are covered by this mechanism. However, word compounds can sometimes occur without the hyphen ("automaker") in which case we are left without such segmentation (REL = automaker, ARG0 = automaker, ARG1 = automaker). This same problem occurs with left slashes for tokens like *property/casualty insurance* where *property* is the ARG1-H0 (a thing insured) and *casualty* is the ARG4-H1 (a danger).[81]

The third problem has to do with how we interpret sentential modification of apposition. According to PTB, the second phrase in the appositive *the leading cause of mental retardation* in example 5 includes the modifier *according to an NIH summary*. However, we believe that this modifier actually modifies the entire apposition, i.e., this apposition construction has the same interpretation as the sentence *According to an NIH summary, Down's syndrome is the leading cause of mental retardation*. Since we are not annotating apposition, we will not mark this modifier.

The fourth problem reflects a general dilemma about how to concatenate fragments of structure that reflect a single slot filler, yet do not reflect a PTB-marked constituent. In this case, we have essentially an unmarked instance of right node raising. Assuming the following analysis where * represents an empty category coindexed with *cause*, the head noun: *[Neither [the Contra *] nor [the SDI *] cause]*. Under such an analysis, *Contra* and *SDI* would each be the ARG2 of a distinct empty category. I should point out that this is a very rare circumstance and

---

[81]We do not separate nouns from particles in for cases like *buy-out*. Rather we assume that these nominalizations incorporate the particle.

assuming an analysis like this would have many practical obstacles including: (1) it would be necessary to change the PTB structure for purposes of annotation; and (2) we would have to allow an empty category to be the value of REL, something which is currently not allowed. Instead, we have chosen to ignore the determiners and including the coordinate conjunction, which makes the parallel structure clear. Thus we mark *Contra nor SDI* as the ARG2.

The fourth problem and others like it motivated us to come up with the following conventions in the concatenation of strings as single constituents:

1. Use the largest chunks possible. In other words, if A consists of X, Y and Z, then it is preferable to choose A as part of an argument filler, than the concatenation of X, Y and Z.

2. Unless obviously incorrect, a segment should include all empty categories. Suppose A consists of X, Y and Z, where Z is an empty category. If there is no obvious reason to exclude Z from the constituent, then mark A rather than the concatenation of X and Y.

3. Only include punctuation as part of a constituent if it is between 2 segments of that constituent or if inclusion makes it easier to choose a larger constituent. Suppose Z is punctuation and A = [X Y Z], B = [X Z Y], C = [X Y Z Q], D = [X Y Q Z]. Then A and B would be usable constituents because the inclusion of the punctuation Z would make it possible to use a single constituent. On the other hand, for C and D, it would make sense to list the concatenation of X and Y (excluding Q) – there would be no point in adding Z to this concatenation.

Appendices

A   A list of Idioms

.

Please note that this list of idiomatic phrases includes both cases that are markable in NomBank and those that are not, according to the criteria discussed above.

in advance of	in the position of	on behalf of
inside of	instead of	outside of
line of credit	line of work	thing of the past
thing of beauty	rule of thumb of . . .	butt of a joke
can of worms	bedrock of society	a creature of the cold war
in the event of . . .	in/on the face of	a fact of life
the fact of the matter is . . .	against the grain of	with a grain of salt
one/a hell of . . .	for the hell of it	at the helm of
the lap of luxury	letter of intent	letter of credit
letter of reproval	in lieu of	in (the) light of
light of day	against the backdrop of	at the mercy of
in the midst of	in POSS midst	in the neighborhood of
part and parcel of	in place of	(= approximately)
a king's ransom	at the root of	for the sake of X
for POSS sake	the scheme of things	by the scruff of POSS neck
seal of approval	sigh of relief	sleight of hand
in spite of	on top of	at the top of POSS lungs
tug of war	to the tune of	in tune with
under the umbrella of	on the verge of	by virtue of
in the wake of	in POSS wake	in the way of
in POSS way	by way of	out of harm's way
the winds of change	in the face of	pillar of support
in escrow	first/second fiddle	fly in the ointment
fodder for X	a foot in the door	a foot in their shoes
hold feet to the fire	give + out	with/in/without regard to
in that regard	wind in the sails	tough sledding
in tow	out of whack	ground zero
out of existence	fly in the ointment	the better part of valor
apple of POSS eye	in contrast with	in concert with
uphill battle	set up shop	on first blush
under the hammer		

B   Lists Helpful in Identifying Support Verb Constructions

Sample Support Verb plus nominalization combinations including particles and/or prepositions as appropriate.

launch an attack on   accept blame   accuse attempt

launch an attack on	accept blame	accuse attempt
admit accusation	bring action	cancel offer
commit murder	complete takeover	deny accusation
disregard order	end participation	execute attack
face accusation	face criticism	file application
gain admission	give confession	grant admission
ignore offer	level blame	level accusation
make accusation	make attempt	mount attack
need alteration	provide accompaniment	put blame on
receive offer	reciprocate with offer	resort to murder
risk destruction	seek authorization	singled out for criticism
sign confession	stage attack	take action
take blame	undertake acquisition	violate order
wreak destruction		

Sample Support Verb plus non-nominalization combinations:

take advantage	play a role	have/set agenda
bear/feel/open to the brunt	catch the eye/fancy of	preach the gospel of
take inventory	make mincemeat of	take place
take the place of	bring shame	take shape
take side	raise the specter of	take step
take (X on) a tour	keep track of	set a trend
get wind of	change the face of	adopt an attitude
cut against the grain of X	scare the hell out of X	take my butt to X
ride the coattails of	give flexibility	give discretion
have the habit of		

Sample Support verbs taken from XMELLT (NB: Our definition of support verb is slightly different from XMELLT's):

abide+by	abort	absolve	accelerate
accept	accuse	acquit	act+on
address	adhere+to	admit	affix+on
agree+to	aim+at	allow	amend
answer	apply+for	appreciate	approach
approve	arraign	arrange	assign
assume	attach+to	attempt	attract
augment	avoid	balk+at	base+on
bear	begin	besiege	best
blurt+out	bombard	boost	botch
break+off	bring	brush+aside	bungle
calculate	call+for	cancel	carry
carry+out	cast+on	cause	cease
charge	choose	circumvent	clear
coerce	collaborate+in	come+back+with	come+forward+with
come+from	commit	compete+for	complete
comply+with	concentrate	confess+to	consider
consign	contest	continue	continue+with
convict	counter	counter+with	couple+with
credit	decide	decide+about	decline
defy	deliver	deluge	deluge+with
deny	design	dismiss	disobey
disprove	dispute	disregard	dissociate
do	draft	drop	ease+up+on
elicit	end	engage+in	engineer
enhance	entertain	escalate	escape
evade	exchange	execute	explore
extend	extract	fabricate	face
fail	fend+off	fight	file
firm+up	fling	flood	flood+with
flout	follow	follow+with	force
frame	fulfill	furnish	further
gain	get	get+away+with	get+in
get+out	give	give+up	go+along+with
go+over	grant	guarantee	have
have+in	heap+on	hear	heed
hit+with	hold+down	hurl	ignore

129

implicate	increase	induce	inflict
initiate	intend	intensify	intrigue
inundate	invite	issue	jack+up
join+in	jump+at	kowtow+to	laugh+off
launch	lay+on	lead	level
levy	levy+at	license	listen+to
lob	make	conduct	march+to
mark	match	miss	modify
mount	mull+over	need	negotiate
obey	obtain	offend+with	offer
orchestrate	organize	parcel+out	participate+in
pass	pass+up	pepper	perform
perpetrate	persevere+with	pin+on	place+on
plan	play	plot	ply
post	postpone	premeditate	prepare
prepare+for	present	press	proclaim
produce	propose	protect	provide
pursue	put+on	put+on+the+table	react+to
rebuff	rebut	recant	receive
reciprocate+with	reconsider	refuse	reject
renege+on	renew	replace	replicate
require	rescind	rescue	resist
resort+to	respond+to	restructure	retract
return+with	reverse	revive	revoke
risk	satisfy	save	scoff+at
seek	share	shepherd	shift+to
shoulder	sift+through	sign	sing
single+out	slate	smart+from	soften
spearhead	spread	spurn	stage
stand+for	start	step+up	study
subject	submit	succeed+in	suffer
suspend	swamp	sweeten	take
take+advantage+of	take+back	take+part+in	take+up
tender	thwart	toss+out	trade
turn+down	undergo	undertake	unleash
violate	vote+for	wage	weigh
welcome	win	withdraw	withhold
woo	wreak	yank	

Subject Control Verbs with objects in COMLEX (NP-ING-SC NP-P-ING-SC
NP-TO-INF-SC):

begin	close	comb	concentrate
count	counter	demand	derive
disgrace	earn	end	enthral
enthrall	excel	expend	help
impair	impede	injure	jeopardize
lend	lose	misspend	modify
promise	provide	reduce	risk
spend	tumble	utilize	waste

Subject Control Verbs without objects in COMLEX (P-ING-SC PART-ING-
SC TO-INF-SC PART-TO-INF-SC):

abstain	act	admit	adore
affect	afford	agitate	agree
aim	apply	arrange	ask
aspire	assay	assent	assist
attempt	baulk	bear	begin
bet	betake	bid	bluff
boast	bother	brace	bring
bristle	buckle	burn	calculate
campaign	care	cease	chafe
check	chip	choose	claim
come	commence	commit	conclude
condescend	confess	consent	consist
conspire	continue	contract	crusade
cut	dally	dare	decide
decline	defrost	deign	demand
deserve	desire	determine	ease
edge	egotrip	elect	emerge
end	endeavor	endeavour	endure
engage	err	exaggerate	excel
expect	experiment	fail	fare
fawn	fear	feel	figure
finish	fit	fixate	flinch
flock	fluctuate	focus	forget
forswear	freeze	function	fuss
get	give	go	grow

grudge	hasten	hate	have
(negated)	hesitate	holler	hope
improve	incline	indulge	intend
interfere	itch	jib	jump
keep	lament	land	lapse
learn	like	linger	live
loathe	lobby	long	love
manage	mean	mess	mind
motion	move	need	neglect
offer	omit	opt	participate
persist	petition	phone	pitch
plan	plead	pledge	plot
ponder	prattle	pray	prearrange
prefer	prepare	press	presume
pretend	proceed	profess	profit
promise	propose	purport	push
qualify	rally	re-elect	reciprocate
recover	recuperate	reelect	refrain
refuse	register	regret	remember
renege	reoffer	request	resolve
respond	resume	retire	return
ride	rule	ruminate	scheme
scorn	scramble	scream	scrounge
scruple	scurry	seek	seesaw
serve	set	settle	shift
shudder	shy	soar	spar
specialize	spring	stand	start
stick	strain	strive	struggle
study	subcontract	succeed	sum
swear	switch	take	tell
testify	think	thirst	threaten
throw	tremble	trim	trouble
try	turn	undertake	unite
venture	vie	volunteer	vote
vow	wait	want	whine
wind	wish	work	yearn

References

[1] Mark C. Baker. *Incorporation: A Theory of Grammatical Function Changing*. The University of Chicago Press, Chicago, 1988.

[2] Charles J. Fillmore. The Case for Case. In E. Bach and R. T. Harms, editors, *Universals in Linguistic Theory*, pages 1–88. Holt, Rinehart and Winston, New York, 1968.

[3] T. Fontenelle. *Turning a bilingual dictionary into a lexical-semantic database*. Lexicographica Series Maior 79. Max Niemeyer Verlag, Tübingen, 1997.

[4] M. Gross. Les bases empiriques de la notion de prédicat sémantique. In A. Guillet and C. Lecl'ere, editors, *Formes Syntaxiques et Prédicat Sémantiques*, volume 63 of *Langages*, pages 7–52. Larousse, Paris, 1981.

[5] M. Gross. Simple Sentences: Discussion of Fred W. Householder's Paper "Analysis, Synthesis and Improvisation". In *Text Processing. Text Analysis and Generation. Text Typology and Attribution. Proceedings of Nobel Symposium 51*, 1982.

[6] Jeffrey S. Gruber. *Studies in Lexical Relations*. PhD thesis, MIT, 1965. Reproduced by Indiana University Linguistics Club, January 1970.

[7] Ray Jackendoff. *Semantic Interpretation in Generative Grammar*. The MIT Press, Cambridge, 1972.

[8] Sven Jacobson. *Adverbial Positions in English*. PhD thesis, Uppsala, 1964. Published by Studentbok, Stockholm.

[9] F. W. Householder Jr., Wolfgang Wolck, P. H. Mathews, Jerome Tove, and James Wilson. *A Preliminary Classification of Adverbs in English*. The Indiana University Linguistics Club, Bloomington, 1965.

[10] P. Kingsbury and M. Palmer. From treebank to propbank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Spain, 2002.

[11] P. Kingsbury, M. Palmer, and Mitch Marcus. Adding semantic annotation to the penn treebank. In *Proceedings of the Human Language Technology Conference*, San Diego, California, 2002.

[12] Richard K. Larson. On the Double Object Construction. *Linguistic Inquiry*, 19:335–391, 1988.

[13] B. Levin. *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press, Chicago, 1993.

[14] C. Macleod. Lexical annotation for multi-word entries containing nominalizations. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Spain, 2002.

[15] C. Macleod, R. Grishman, and A. Meyers. COMLEX Syntax. *Computers and the Humanities*, 31:459–481, 1998.

[16] C. Macleod, R. Grishman, A. Meyers, L. Barrett, and R. Reeves. Nomlex: A lexicon of nominalizations. In *Proceedings of Euralex98*, 1998.

[17] I. A. Mel'čuk. *Dependency Syntax: Theory and Practice*. State University Press of New York, Albany, 1988.

[18] I. A. Mel'čuk. Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In *Lexical Functions in Lexicography and Natural Language Processing*. John Benjamins Publishing Company, Amsterdam, 1996.

[19] A. Meyers, R. Reeves, and Catherine Macleod. NP-External Arguments: A Study of Argument Sharing in English. In *The ACL 2004 Workshop on Multiword Expressions: Integrating Processing*, Barcelona, Spain, 2004.

[20] A. Meyers, R. Reeves, Catherine Macleod, Rachel Szekeley, Veronkia Zielinska, and Brian Young. The Cross-Breeding of Dictionaries. In *Proceedings of LREC-2004*, Lisbon, Portugal, 2004.

[21] E. Miltsakaki, A. Joshi, R. Prasad, and B. Webber. Annotating discourse connectives and their arguments. In A. Meyers, editor, *NAACL/HLT 2004 Workshop: Frontiers in Corpus Annotation*, pages 9–16, Boston, Massachusetts, USA, May 2 - May 7 2004. Association for Computational Linguistics.

[22] Ruslan Mitkov. *Anaphora Resolution*. Pearson Education Limited, Harlow, Great Britain, 2002.

[23] D. M. Perlmutter and P. M. Postal. The 1-Advancement Exclusiveness Law. In D. M. Perlmutter and C. G. Rosen, editors, *Studies in Relational Grammar 2*. The University of Chicago Press, Chicago, 1984.

[24] M. Poesio and R. Vieira. A Corpus-based Investigation of Definite Description Use. *Computational Linguistics*, 24:183–216, 1998.

[25] Randolph Quirk and Joan Mulholland. Complex Prepositions and Related Sequences. *English Studies*, 45:64–73, 1964.

[26] C. G. Rosen. The Interface between Semantic Roles and Initial Grammatical Relations. In D.. M. Perlmutter and C. G. Rosen, editors, *Studies in Relational Grammar 2*. The University of Chicago Press, Chicago, 1984.

[27] Jerry Saddock. *Hypersentences*. PhD thesis, University of Illinois, 1969.

[28] Muffy A. Siegel. *Capturing the Adjective*. PhD thesis, UMass, 1976. Published in 1980, New York: Garland Publishing, Inc.

[29] University of Pennsylvania. Annotation guidelines for PropBank. http://www.cis.upenn.edu/~ace/propbank-guidelines-feb02.pdf, 2002.

[30] Edwin Williams. Predication. *Linguistic Inquiry*, 11:203–238, 1980.

[31] S. R. Wolff, C. Macleod, and A. Meyers. *Comlex Word Classes Manual*. Proteus Project, New York University, 1998.