Using Treebank, Dictionaries and GLARF to Improve NomBank Annotation

Adam Meyers
New York University
New York, NY
meyers@cs.nyu.edu

Abstract
In the field of corpus annotation, it is common to annotate text multiple times and then adjudicate the results. The resulting annotation is generally regarded as more consistent and more accurate than the results of a single pass. However, it is also very expensive to annotate in this way. Given text corpora that are annotated by many different research groups, another source of comparison is available: annotation of other linguistic information on the same corpora. By exploiting violations of expected relationships between the two annotation schemes, likely errors can be detected. This paper describes such an effort involving the NomBank annotation of noun arguments in the Wall Street Journal Corpus. These techniques made it possible to complete NomBank annotation efficiently and accurately.

1. Introduction
As with many annotation projects, NomBank took longer to finish than the creators initially expected. It eventually became necessary to find a way to complete the annotation in a way that minimized expenses, while maintaining high quality. In many projects involving the manual annotation of corpora with linguistic features, each text is annotated by two different annotators and the differences between their output are adjudicated. The resulting annotation is more consistent than singly annotated corpora and this increased consistency is usually assumed to indicate a corresponding improvement in accuracy. Due to practical constraints, this was not an option for NomBank.

Fortunately, the NomBank project was annotating a text corpus for which there was already previous annotation (in particular, Penn Treebank annotation). We established several expected relationships between the NomBank and the Penn Treebank annotation schemes. When any of these expected relationship did not hold, there were three possibilities: (1) there was an error in NomBank; (2) there was an error in the Penn Treebank; or (3) the expected relationship did not hold for this instance. Given these possibilities, annotation that violated an expected relationship was more likely to contain a NomBank error, than randomly selected annotation.

In addition, some parts of NomBank annotation had expected relationships with syntactic dictionaries, both ones created during the NomBank project (ADJADV, NOMLEX-PLUS) and existing ones (NOMLEX and COMLEX Syntax). By examining cases where these expected relationships were violated, we could predict likely NomBank (or dictionary) errors. As a result of these efforts, approximately 26% of NomBank manual annotation was predicted to contain likely errors and was examined and corrected by an expert annotator, a substantial savings in time and effort. Methods for evaluating the effectiveness of this effort are under consideration for future work.

NomBank annotators reviewed a total of 200,000 instances of nouns in the Penn Treebank corpus to produce 114,500 NomBank propositions. On average, they looked at about 20–25 noun instances per hour, working at a considerably slower pace than PropBank (Palmer et al., 2005)(less than one half the speed). This made double annotation of NomBank impractical. By comparing NomBank annotation to previous annotation, we were able to select approximately 30,000 propositions that were likely to contain errors and review those propositions in a focused way. This made for realistic and effective quality control.

We will now sketch an outline of the remainder of this paper. Section 2. provides an overview of NomBank annotation. Section 3. describes our approach to merging together various annotation schemes into a GLARF representation, which we use for our error detection system. Sections 4. through 7. describe the various constraints that we use to detect likely errors. Finally, Section 8. discusses ramifications and future research .

2. NomBank Annotation
NomBank.1.0 (Meyers et al., 2004a) provides a predicate argument structure representation of approximately 114,500 noun instances in the Wall Street Journal corpus. Like PropBank, this representation links particular word instances with words and phrases that are either arguments (ARG0, ARG1, . . .) or belong to one of the classes of nonarguments (ARGMs) defined in the specifications. For each word, there is a dictionary entry (its frame file) which defines the set of possible arguments. The set of markable ARGMs are essentially those that have counterparts in verbal argument structure, e.g., temporal, locative, manner, etc. In addition, we mark SUPPORT items, words that link arguments outside of the noun phrase to the nominal predicate. Some example sentences are provided below. The nominal predicate is underlined and the other parts of proposition are in bold. The labels following the arguments indicate the roles they play in the NomBank proposition. The set of support words in each of these examples forms a chain in that sentence connecting an argument outside the NP to the underlined predicate. For example the support chain, consisting of gave + dozens + of, links John to kisses – the chain should be viewed as filling a single SUPPORT slot in the NomBank proposition.

See the NomBank manual, available from nlp.cs.nyu.edu/meyers/NomBank.html, for more information.
1. Mary’s/ARG0 promise/ARG1-REF to John/ARG2
d2. The Press’s/ARG0 criticism of the candidate/ARG1
3. John/ARG0 gave/SUPPORT Mary/ARG2 dozens of/SUPPORT kisses
4. They/ARG0 accorded/SUPPORT minorities/ARG1 an opportunity for/SUPPORT representation.

Like PropBank, each word and phrase in NomBank is represented as a link to one or more nodes of Penn Treebank annotation (Marcus et al., 1994). This contrasts with most approaches to annotation such as: (a) inline annotation where the text is modified to include annotation features and (b) offset annotation which points to particular spans of text using another document (these text spans are usually referenced by byte offsets from the beginning of the target file). In this sense, NomBank is annotation of annotation, i.e., NomBank assigns features to units defined by pre-existing Penn Treebank annotation.

3. GLARFBANK

As part of the Unified Linguistic Annotation project (Pustejovsky et al., 2005), researchers at several United States universities are studying ways to merge together distinct annotation schemes. At New York University (NYU), we are taking an approach to merging that we call “aggressive” because we change incompatible aspects of the input annotation schemes so that they are compatible with each other, i.e., we change tokenization, phrase boundaries and text spans to maximize overlap between the input annotation schemes. In this respect, we are taking annotation created under different theoretical assumptions and converting them into a single-theory analysis. The output of the merging process is formalized as a Typed Feature Structure in the GLARF framework (Meyers et al., 2001a; Meyers et al., 2001b).

The current GLARF’d version of the Wall Street Journal data annotated for NomBank includes the following annotation schemes: Penn Treebank, PropBank, NomBank, Penn Discourse Treebank (overt relations)(Miltsakaki et al., 2004) and BBN Named Entity tags. Future merged GLARFBANKs will also include Brandeis’ TimeML (Pustejovsky et al., 2004) and University of Pittsburgh’s Opinion annotation (Wilson and Wiebe, 2003). The WSJ GLARFBANK also includes various automatically generated features based on both heuristic rules and lexical lookup (COMLEX Syntax, NOMLEX, ADJADV, and others). GLARF rules correct parts of speech, mark focused constituents, fill gaps not covered by Treebank annotation, assign grammatical roles to constituents, add semantic features, etc. A sample (simplified) GLARF representation is provided as Figure 1. It represents the merger of annotation for the sentence Meanwhile, they made three bids.

The GLARF representation essentially adds structure to the Penn Treebank and if you delete this additional structure, the result would be the original Penn Treebank (with minor changes). We will highlight two of these elaborations that are used in the current GLARFBANK.

OANC-1. Prior to the availability of hand annotation, automatically generated features are provided for PropBank, NomBank and the Penn Discourse Treebank. The author intends to make the Wall Street Journal GLARFBANK available either through the Linguistic Data Consortium, or by download should licensing restrictions on this corpus be relaxed.

5In the GLARF system the typed feature structure includes all the information in GLARF. A multi-level dependency representation is also available that is similar to the 2008 CONLL task representation (www.yr-bcn.es/conll2008/). In fact the latter is partially derived from the former.

3There are actually several different GLARF representations. The typed feature structure representation contains the most information and a dependency representation is the one that is most often used for Information Extraction and other applications.
here: (1) relational labels like HEAD, ADV, PRD, OBJ, that indicate relations between constituents, e.g., the constituent labeled SBJ is the subject of the sister constituent that is labeled PRD; and (2) Empty Categories that may or may not be part of the original Penn Treebank, e.g., the features prefixed with P- point to empty categories which bear PropBank, NomBank and PDTB relations with the HEAD constituent. These empty categories point to other GLARF constituents, e.g., the the NP they has an INDEX feature value of 2. The empty categories that are values of the P-ARG0 of made and bids both also have this index, representing that they is the PropBank ARG0 of made and the NomBank ARG0 of bids. The P-ARG2 of Meanwhile has a value of the entire sentence, which would appear to include itself. However, by convention, we assume that such arguments exclude what we call the SELF-PHRASE, the ancestor of the predicate (in this case Meanwhile) that is a child of the argument. This same rule is used for marking arguments of parenthetical predicates in PropBank and NomBank. Thus in the following two examples, the entire sentence can be marked as an argument of claimed and request because the self-phrases Mary claimed and at John’s request can easily be accounted for: Irving, Mary claimed, is ten feet tall, Mary, at John’s request, made ridiculous claims about Irving. The P-ARG1 of Meanwhile refers to the previous sentence (sentence 0, index 0).

Our system checks new NomBank data for its compatibility with other annotation frameworks, using the GLARF-BANK annotation as a way of incorporating the other annotation into a single representation. Following sections describe these compatibility tests and the subsequent adjudication.6

4. Structural Constraints on Internal Arguments of Nouns

We use the GLARF representation as a means to implement several types of constraints. First of all, by recognizing particular kinds of constituents, we can constrain how they appear in NomBank. Relative clauses typically are not markable in NomBank propositions. Thus, given a NomBank Proposition for a noun N, if one of the arguments (ARG0 . . . ARG9) or ARGMs is a relative clause, this is flagged as a likely error, e.g., the that relative in the banner that proclaims the renewal of socialism was detected as a likely error and then removed during adjudication. It is easy to identify relative clause arguments because relative clauses are labeled as such in the GLARF’d version of the Penn Treebank. The GLARF-generating program uses a combination of the representation in the original Penn Treebank (the appearance of empty categories in that-clauses following nouns, the POS markings on that, etc.) and whether or not a that-clause is a possible complement for the head noun (using COMLEX Syntax) to determine if a structure is a relative clause (if a that phrase follows a noun that can’t take that complements, the phrase is likely to be a relative clause).

NomBank annotators have the option of linking together constituents in the Penn Treebank to form a single NomBank argument. These combinations often correctly identify constituents not marked in the Penn Treebank, due to (for example) Penn Treebank’s tendency to underspecify pronominal structure, e.g., in a phrase like The ice cream man, ice and cream would probably be left as separate constituents. However, it turns out that some constituent combinations are unlikely to be correct. For example, given D and N two adjacent pronominal modifiers of some head H, if D is a determiner or possessive and N is a noun or adjective, it is unlikely that D and N form a constituent. For example, one annotator marked their financial as a single constituent (an ARG1) of the predicate viability in the phrase their financial viability. In the corrected version, their is marked as an ARG3 and financial is marked as an ARG1. The reason for this error is clear. ARG3 and ARG1 are similar roles for nouns like viability which belong to the ATTRIBUTE class and the annotator opted to combine the two rather than mark them separately. The ARG1/ARG3 split in NomBank reflects that viability is an attribute of the financialness and financial viability is an attribute of them. In this case, the ARG3 is a secondary-theme a type of argument that has this interpretation (as per the NomBank manual). Their financial viability is a phrase that represents the degree or VALUE of the viability trait and therefore viability is marked as its own ARG2. This error detection routine occasionally identifies non-errors. For example, the GLARF generating program incorrectly marked the numeral 1 as a determiner in the sentence CBS held the previous record for consecutive No. 1 victories. The annotator had correctly marked No. 1 as a single ARG1 – so this annotation was not changed during adjudication.

In a similar vein, annotations of discontinuous constituents are unlikely to be correct. Any series of constituents that form a NomBank argument are almost always consecutive. Nevertheless, NomBank annotators will occasionally mark discontinuous constituents, the most common reasons being: (1) one token is missed from a sequence, e.g., the comma was not included as part of the ARG1 stock, bond and foreign exchange in the initial marking of the phrase its stock, bond and foreign exchange trading; and (2) as in the determiner plus pronominal case above, the two arguments have similar relations to the head noun. For example, although one annotator marked a combination of conversion and on the stock as a single ARG1 of rights in the phrase conversion rights on the stock, the final version of NomBank makes conversion an ARG1 and on the stock an ARG3. The one consistent exception, discussed in Section 6., is where the entire sentence or NP is an argument of the noun (minus the self-phrase containing the nominal predicate). For example, in Mr. Nadeau said discussions are under way with potential purchasers of each of the units, the entire phrase minus under way is an ARG1. Apart from these carefully defined exceptions, there are also 10 cases involving the noun predicate age, where marking discontinuous constituents seemed unavoidable even though the examples did not fit into one of cases of external argu-

---

6Tests for compatibility between the structure of the GLARF-BANK and NomBank are mostly tests for compatibility between the Penn Treebank and NomBank. However, the GLARF/BANK actually incorporates structures from other annotation. So the relation is not one to one.
ments of nouns, e.g., we marked under 13 the ARG2 of age in the phrase 1,859 children under age 13.

5. A Constraint on Empty Categories

Empty categories (Penn Treebank’s way of representing gaps) are not typically noun arguments unless they are part of chains that link the empty category to a (pronounceable) word or phrase (the filler of the gap). Consider, for example, the NomBank annotation of veto in the following sentence: Mr. Bush and some other aides are strongly drawn to the idea of trying out a line-item veto. Mr. Bush and some other aides should be the ARG0 of veto as mediated by: (1) a number of empty categories in the Penn Treebank: the passive object of drawn and the subject of trying; and (2) the support verb trying. In the initial annotation, a NomBank annotator failed to make the final link from the passive object empty category to the lexical NP, but the error detection program predicted that this was a likely error. Exceptions do occur when an empty category represents an unfilled argument. For example, in the following definition of stock-index arbitrage, the ARG0 of trades should be the same as the empty subject of executing, which itself is unbound: Stock-index arbitrage – Buying or selling baskets of stocks while at the same time executing offsetting trades in stock-index futures or options.

The Penn Treebank resolves the referential properties of some, but not all empty categories. In the following example, a NomBank annotator needed to add the link between the possessive phrase Illinois Supreme Court’s and the empty subject of to institute: Illinois Supreme Court’s decision to institute the changes. Here institute acts as a support verb linking its subject to the ARG0 position of the noun changes, i.e. the Illinois Supreme Court is assumed to be the AGENT of the changes. Therefore, it turns out that only some of the cases where empty categories are not bound in the Penn Treebank need to remain so and it turns out that unbound empty categories are unlikely to be correct as NomBank arguments – their presence signals a likely error.

6. Structural Constraints on External Arguments of Nouns

NomBank specifications place restrictions on the markability of a given potential argument A of a noun N that lies outside of the NP headed by N. It turns out that, for the most part, these restrictions were codable in terms of GLARF’d representations of the sentence and therefore could be automatically checked. Although there are some outliers that the automatic system did not handle correctly, the automatic detection system tended to overpredict errors, rather than underpredict. This made it possible to accurately identify many cases that we needed to review more carefully and it resulted in corrections of many NomBank propositions.

There are three environments in which External arguments can be licensed: (a) support; (b) predication; and (c) PP constructions containing the nominal predicate. Each of these configurations make specific requirements on how the NP-external arguments are linked to the nominal predicate. Furthermore, the absence of any of these configurations means that an NP-external argument is unlicensed and thus tagged as a likely error.

6.1. Constraints on Support Structures

A NomBank external argument A is a legal argument of a nominal predicate P, by virtue of support, if there exists a support chain S linking A to P. To be well-formed, a support chain must meet the following criteria: (1) consist completely of lexical items (leaf nodes) in the Penn Treebank; (2) forms of be, auxiliaries, infinitival to and modals are skipped, i.e., for purposes of the support chain, we pretend that they do not exist and that the main verb, predicate adjective, or other predicative item is the main predicate of its clause; (3) at least one item in the support chain must have as its part of speech: noun, adjective, verb or determiner; (4) each link in the chain must be the head of the phrase containing it (after allowing for 11); (5) the first link in the chain must take A as its argument; (6) Each link N in the main chain must take the phrase headed by link N + 1 as its argument; (7) the last link in the main chain must take the phrase headed by P as its argument; and (8) the chain cannot cross any tensed clause phrasal boundaries. A schema of a support chain is provided as Figure 2. Some examples of legal support chains are provided as Figure 3. There are several ways which we use the constraints on support to verify the accuracy of NomBank annotation: (1) we verify that annotated support chains meet the criteria above; (2) we verify that there are external arguments that require support chains and propose the removal of annotated support chains that are extraneous; (3) we automatically generate a support chain and compare it to the one annotated. In each of these cases, we use the error detection procedures to identify potential errors. Should we determine that they are actual errors, we correct them.

Given a possible external argument A and a nominal predicate P, we assume that exactly one support chain is structurally possible. In simple cases, one can think of the typed feature structure as a labeled tree, although it is actually a rooted directed acyclic graph. In most cases, to find the support chain, one first must identify the path derived by going up the tree from A to the common ancestor of A and P, and then down the tree to P. The support chain is the
1. The real ARG-ADV battle is over who will control the market ARG2

2. This book is about his son ARG1

3. Trying to time the economy ARG1 is a mistake

4. They ARG1 are some ARG2 distance apart

Figure 4: Linking External Arguments to Nouns Via Predication

Support chains linked a given A with a given P. Special allowances are made so that conjoined predicates can both be part of the same support chain, e.g., in Mary gave and received lots of kisses, gave and received are assumed to be branches of the same support chain (gave + received + lots + of). It is as if the support chain splits in the middle and then merges together again because, for the purpose of a support chain, coordinate structures are assumed to have multiple heads.

6.2. Constraints on Predication

There are a number of instances in which predication licenses a connection between an argument and a noun predicate which we have determined are legitimate for marking NomBank arguments. We specifically avoid cases in which the argument can duplicate existing arguments, e.g., for argument nominalizations like teacher, we will always mark teacher as its own ARG0 and never NPs linked by predication, e.g., Mary is John’s teacher.

We recognize the following markable instances of linking external argument to nouns via predication: (1) when the noun predicate is the subject of the sentence and one of its arguments follows a copula, e.g., Examples, 1 and 2 in Figure 4; and (2) when the noun predicate P follows the copula and its argument precedes the copula and P is either a nominalization of an adjective, an ATTRIBUTE noun (a NomBank class) or in a preposition plus noun construction that has an adjective-like distribution, e.g., 3–4, in Figure 4. A subset of the nouns in COMLEX Syntax that are marked with the feature (COUNTABLE : PVAL) combine with the preposition to form adjective-like constituents, e.g., the entry of alert is marked (COUNTABLE : PVAL (“on”)). These entries can be used to identify instances of the aforementioned adjective-like PP construction.

Identifying these environments automatically is easy. One merely has to identify copulas, the subjects of those copulas (typically the NP or sentence immediately following the copula) and the underlying predicate (typically the phrase immediately following the predicate and often marked with the function tag -PRD). Other predicative environments, though rarer, are also easy to detect in the Penn Treebank: small clauses are S constituents consisting of an NP followed by another constituent marked with -PRD, as con-
1. Without/ARGM-NEG question, something intriguing is going on/ARG1 [PP Parenthetical]

2. Some last-minute phone calls that Mr. Bush made/ARG1 (at the behest of some conservative U.S. senators/ARG0) to enlist backing for the U.S. position/ARG1 [PP Parenthetical]

3. He/ARG1 was under consideration to succeed Joshua Lederberg/ARG2 [PP + Extraposition]

4. ABC’s baseball experience/ARG0 may be of interest to CBS Inc./ARG1 [PP + Extraposition]

5. they/ARG0 exercise for enjoyment [Subject-Oriented PP]

6. Garbage/ARG0 made its debut this fall with the promise to give consumers the straight scoop on the U.S. waste crisis/ARG1 [Subject-Oriented PP]

7. Participants/ARG0 in the meeting [Noun-Modifying PP]

8. the bitterness/ARGM-MNR of the battle [Noun-Modifying PP]

9. That/ARG1 was in addition to $34,000 in direct campaign donations/ARG2 [Discourse Connective]

10. That $130 million gives us some flexibility/ARG1 in case Temple raises its bid/ARG2. [Discourse Connective]

11. In important particulars, the Soviets are different from the Chinese/ARG1 [Discourse Adverbial]

12. In fact, they don’t take it seriously at all/ARG1

Figure 5: PP constructions that license External Arguments

Substituents begin with the word as, etc.

6.3 PP constructions and External Arguments

When the NP headed by a predicate noun is the object of a preposition, the argument taking properties of that noun may change. This subsection describes a set of argument-taking environments in which such PPs license external arguments according to NomBank guidelines. These environments include: (1) The PP-parenthetical construction; (2) The PP + Extraposition construction; (3) Subject Oriented PPs; (4) Noun modifying PPs; and (5) Other Adverbial PPs including discourse connectives. Examples are provided in Figure 5. Although we can automatically detect most of these environments, we have not implemented ways of detecting all of them. Thus our automatic procedures still flag many of these as instances of unlicensed external arguments. As a result, many of the rarer PP constructions are always revisited during the error detection phase of annotation.

The PP-Parenthetical (Examples 1 and 2) and extraposed PP constructions (Examples 3 and 4) are both licensed by COMLEX Syntax dictionary entries and, in the former case, is limited to a short list of prepositions. The configurations are easily defined in terms of syntactic trees (or graphs). The PP-Parenthetical cases are licensed by nouns that take clausal complements and this lexical information is readily available from a combination of COMLEX Syntax and/or Nomlex (or Nomlex-Plus). These PP phrases (Examples 1 and 2 in Figure 5) are like their verbal counterparts (e.g., the say phrase in Mary, John said, is an incredible botanist) in that they can precede, follow or infix their sentential argument. In addition to the lexical subcategorization of the nominal predicate, another restriction is that only a narrow set of prepositions seem to license this construction: (with, without, at, on in and possibly a few others). The PP is immediately dominated by the sentence that it takes as an argument (the PP is typically marked as a parenthetical in the Penn Treebank or offset by parentheses or commas). The Extraposition cases (Examples 3 and 4) are possible for a subset of nouns marked in COMLEX Syntax with the subcategorization features EXTRAP-P-CP-UPTO-NOUN-THAT-S. The COMLEX entry also specifies the preposition. For example, the COMLEX entry for interest includes the subcategorization feature (EXTRAP-P-CP-UPTO-NOUN-THAT-S :PVAL (“of“)). In the Penn Treebank, the nominal predicate is the rightward argument of the copula and the subject of the copula is one argument of the noun. Using a combination of these lexical clues and configurational data, it is easy to see how correctly licensed instances of these constructions can be automatically identified.

Subject oriented adverbial PPs containing a NomBank predicate (Figure 5, Examples 5 and 6) can be identified by the following characteristics: (1) the subject of the sentence is an argument of the NomBank predicate (hence the name subject-oriented); (2) the PP is either a child of the sentential node or a child of the VP; and (3) the preposition belongs to a defined set which includes mainly temporal prepositions (after, before, during), instrumental prepositions (with, without, through by) and several others. These PPs are similar to other subject-oriented adverbs like willingly, vengefully, etc., which typically select for an animate subject.

The fourth case (Figure 5, Examples 7 and 8) involves a noun A that is modified by a PP containing a nominal predicate P, such that P takes A as an argument. This is an easy to recognize configuration and is limited to approximately the same set of prepositions as the others. We have yet to fully figure out the distribution of the nominal predicates that can occur in this configuration, although it does seem that adjective nominalizations and ATTRIBUTE nouns are the most common.

Finally, there are some NomBank frame entries that classify particular nouns as being either a discourse-connective (Examples 9 and 10) or discourse-adverbial (Examples 11 and 12). Similar entries are found in the NOMADV dictionary giving them one of the COMLEX Syntax classes applied to similar adverbs, i.e., the various sub-types of the META-ADV class (the connectives belong to the (META-ADV :CONJ T) class). The discourse adverbials can take entire sentences as arguments, whereas the discourse connectives link two arguments in a similar manner to the discourse connectives in the Penn Discourse Treebank (PDTB). Nom-
After hours/SUPPORT+ARGM-TMP of/SUPPORT debate, the jury/ARG0 focuses on the facts.

John/ARG1 is 40/ARG2 pounds/ARG2+SUPPORT in/ARG2 weight.

Bank discourse connectives can link two sentences, two NPs or one NP and one sentence. This contrasts with PDTB connectives, which always link two sentences. The discourse adverbials, like the Parentheticals can follow or be embedded in the sentence it modifies. NomBank discourse connectives have a similar configurational distribution as the PDTB connectives: the connective forms a constituent with one argument (e.g., in case Temple raises its bid in Example 10 and the other argument is either the rest of the superordinate phrase (the subject and the verb) or the subject of the sentence (e.g., that in Example 10). However, unlike PDTB, NomBank does not link predicates in one sentence with arguments outside that sentence, e.g., NomBank does not mark the sentence preceding an example like no. 12 as an argument of fact.

In summary, there are a number of configurations in which a PP containing a NomBank predicate (as the head of the prepositional object) that license external arguments of that noun. The configurations are easy to define and additional lexical restrictions makes it possible to identify the markable cases in NomBank. As of this writing, we recognize a subset of the admissible cases automatically. The remainder we must verify manually.

6.4 Combining Support with Other Constraints

We end this section with the examples in Figure 6, which combine support with some of the other external argument licensing environments. Both cases involve transparent noun constructions, which are viewed as a type of Support in NomBank. After hours of debate is treated as if debate is the main predicate of this subject-oriented PP construction (the subject of the sentence is an argument of debate). The support chain hours + of makes this treatment possible. In a similar way, the support chain pounds + of makes it possible for weight to be connected to the subject of the sentence by predication. The support chains serve to bring the nominal predicate into the position required to link them via these other types of constructions.

7. Lexical Constraints on NomBank

We will now describe one of the main dictionary-based constraints that we used to correct NomBank. At the same time, we used this constraint to correct the dictionary ADJADV (Meyers et al., 2004b), which we made along side of NomBank.

Although ARG1 ... ARG9 features were applied according to frames for particular words, the distribution of the ARG feature was left to the annotator's interpretation of the NomBank specifications. Nevertheless, to a large extent the ARG features are also lexical in nature, but of a different sort. ARGs tend to be the same for particular modifiers (the value of the ARG itself), rather than nominal predicate. For example, the adjective recent is almost always marked ARG-TMP due to lexical properties of recent, not lexical properties of the noun it modifies. Thus recent should be marked ARG-TMP in the recent destruction of the documents, their recent marriage and the recent knowledge, regardless of what is in the frame entries of destruction, marriage and knowledge. We observed that the relevant information could not be found in the adjective entries of COMLEX Syntax, but could be found in related adverb entries. Specifically, recently, the adverb related to recent has the feature TEMPORAL-ADV. This motivated our construction of ADJADV. Some sample entries are given below in Figure 7. This dictionary was created in a semi-automatic way. For the most part, we simply found morphologically adjective adverb pairs and generated the entry based on the adverb. However, in some cases, e.g., big, we created an ADJADVLIKE entry based on a semantically related adverb.

Given the assumption that specific adjectives tended to be compatible with the same ARG function tags, we could automatically detect likely errors by comparing the ARGMs assigned adjective premodifiers in NomBank against the ADJADV dictionary entries for those adjectives.14 We assumed the table of compatibilities between function tags and COMLEX-SYNTAX features listed as Table 1. When an adjective was marked in a NomBank proposition in a way that was incompatible with the ADJADV entry, this would usually lead to either changing the NomBank annotation or changing the ADJADV lexical entry. In this way, we were able to simultaneously improve both NomBank and ADJADV.

8. Concluding Remarks

Above, we have outlined major ways in which we have improved NomBank by evaluating the compatibility of annotation with other resources. As a result of these and similar techniques, we have looked closely at over 30,000 of

14 Other premodifiers were handled in other ways, e.g., prefixes were specially classified; numbers between 1000 and 2100 were assumed to be potential time modifiers, etc. Also, with respect to hyphenated items, we identified one hyphenated segment (typically the last segment) as the head and looked up the ADJADV entry for that segment. We omit a full description due to space limitations.
the 114,500 NomBank instances. We believe that these measures caused us to focus our efforts on the most likely causes of error, improving both the accuracy and efficiency of quality control. Had we annotated NomBank twice and then adjudicated instead of using this methodology, it would clearly have been a more expensive undertaking. Furthermore our attention would not have been as directed as it was using the error detection program.\footnote{Of course, we fixed errors that we found that were not detected by the program as well.}

We have considered creating a degraded version of NomBank that consists of only pre-edited entries. We could then test to see if a automatic role labeling system (Jiang and Ng, 2006) trained on that version would not perform as accurately as a system trained on the final version. Better performance on the final system would confirm that we improved the system using our methods. However, this result would hardly be surprising because our technique does involve a selective second pass on the annotation by an expert annotator, methodology which is widely recognized to improve results. Clearer evaluation would require the annotation of additional data in a test setting in which dual annotation plus adjudication could be fairly compared with it would clearly have been a more expensive undertaking.

This paper provides examples of how constraints on a new annotation scheme can be formulated in terms of previous annotation in order to provide quality control. Researchers who would like to take advantage of this methodology should consider annotating corpora that has already been annotated by other members of the annotation community

<table>
<thead>
<tr>
<th>COMLEX Feature</th>
<th>ARGGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(META-ADV :CONJ T)</td>
<td>ARGGM-DIS</td>
</tr>
<tr>
<td>other META-ADV</td>
<td>ARGGM-ADV</td>
</tr>
<tr>
<td>MANNER-ADV</td>
<td>ARGGM-MNR</td>
</tr>
<tr>
<td>DEGREE-ADV</td>
<td>ARGGM-MNR</td>
</tr>
<tr>
<td>EVAL-ADV</td>
<td>ARGGM-MNR</td>
</tr>
<tr>
<td>LOC&amp;DIR-ADV</td>
<td>ARGGM-LOC, ARGGM-DIR</td>
</tr>
<tr>
<td>TEMPORAL-ADV</td>
<td>ARGGM-TMP</td>
</tr>
</tbody>
</table>

Table 1: ADJADV/ARGM Compatibility

Acknowledgments

This research was supported by the National Science Foundation, award CNS-0551615, entitled Towards a Comprehensive Linguistic Annotation of Language.

9 References


