The NomBank Project: An Interim Report

Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely,
Veronika Zielinska, Brian Young and Ralph Grishman
New York University
meyers/reevesr/macleod/szekely/zielinsk/byoung/grishman@cs.nyu.edu

Abstract

This paper describes NomBank, a project that will provide argument structure for instances of common nouns in the Penn Treebank II corpus. NomBank is part of a larger effort to add additional layers of annotation to the Penn Treebank II corpus. The University of Pennsylvania's PropBank, NomBank and other annotation projects taken together should lead to the creation of better tools for the automatic analysis of text. This paper describes the NomBank project in detail including its specifications and the process involved in creating the resource.

1 Introduction

This paper introduces the NomBank project. When complete, NomBank will provide argument structure for instances of about 5000 common nouns in the Penn Treebank II corpus. NomBank is part of a larger effort to add layers of annotation to the Penn Treebank II corpus. PropBank (Kingsbury et al., 2002; Kingsbury and Palmer, 2002; University of Pennsylvania, 2002), NomBank and other annotation projects taken together should lead to the creation of better tools for the automatic analysis of text. These annotation projects may be viewed as part of what we think of as an *a la carte* strategy for corpus-based natural language processing. The fragile and inaccurate multistage parsers of a few decades were replaced by treebank-based parsers, which had better performance, but typically provided more shallow analyses.[1] As the same set of data is annotated with more and more levels of annotation, a new type of multistage processing becomes possible that could reintroduce this information, but in a more robust fashion. Each stage of processing is defined by a body of annotated data which provides a symbolic framework for that level of representation. Researchers are free to create and use programs that map between any two levels of representation, or which map from bare sentences to any level of representation.[2] Furthermore, users are free to shop around among the available programs to map from one stage to another. The hope is that the standardization imposed by the annotated data will insure that many researchers will be working within the same set of frameworks, so that one researcher's success will have a greater chance of benefiting the whole community.

Whether or not one adapts an *a la carte* approach, NomBank and PropBank projects provide users with data to recognize regularizations of lexically and syntactically related sentence structures. For example, suppose one has an Information Extraction System tuned to a hiring/firing scenario (MUC, 1995). One could use NomBank and PropBank to generalize patterns so that one pattern would do the work of several. Given a pattern stating that the object (ARG1) of *appoint* is *John* and the subject (ARG0) is *IBM*, a PropBank/NomBank enlightened system could detect that IBM hired John from the following strings: *IBM appointed John*, *John was appointed by IBM*, *IBM's appointment of John*, *the appointment of John by IBM* and *John is the current IBM appointee*. Systems that do not regularize across predicates would require separate patterns for each of these environments.

The NomBank project went through several stages before annotation could begin. We had to create specifications and various lexical resources to delineate the task. Once the task was set, we identified classes of words. We used these classes to approximate lexical entries, make time estimates and create automatic procedures to aid in

[1] A treebank-based parser output is defined by the treebank on which it is based. As these treebanks tend to be of a fairly shallow syntactic nature, the resulting parsers tend to be so also.

[2] Here, we use the term "level of representation" quite loosely to include individual components of what might conventionally be considered a single level.

1. *Her gift of a book to John* [NOM]
   REL = *gift*, ARG0 = *her*, ARG1 = *a book*, ARG2 = *to John*

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

3. *her husband* [DEFREL RELATIONAL NOUN]
   REL = *husband*, ARG0 = *husband*, ARG1 = *her*

4. *a set of tasks* [PARTITIVE NOUN]
   REL = *set*, ARG1 = *of tasks*

5. *The judge made demands on his staff* [NOM w/SUPPORT]
   REL = *demands*, SUPPORT = *made*, ARG0 = *The judge*, ARG2 = *on his staff*

6. *A savings institution needs your help* [NOM w/SUPPORT]
   REL = *help*, SUPPORT = *needs*, ARG0 = *your*, ARG2 = *A savings institution*

7. *12% growth in dividends next year* [NOM W/ARGMs]
   REL = growth, ARG1 = in dividends, ARG2-EXT = 12%, ARGM-TMP = next year

8. *a possible U.S. troop reduction in South Korea*[NOM W/ARGMs]
   REL = reduction, ARG1 = U.S. troop, ARGM-LOC = in South Korea, ARGM-ADV = possible

Figure 1: Sample NomBank Propositions

annotation. For the first nine months of the project, the NomBank staff consisted of one supervisor and one annotator. Once the specifications were nailed down, we hired additional annotators to complete the project. This paper provides an overview of the project including an abbreviated version of the specifications (the full version is obtainable upon request) and a chronicle of our progress.

2 The Specifications

Figure 1 lists some sample NomBank propositions along with the class of the noun predicate (NOM stands for nominalization, DEFREL is a type of relational noun). For each "markable" instance of a common noun in the Penn Treebank, annotators create a "proposition", a subset of the features {REL, SUPPORT, ARG0, ARG1, ARG2, ARG3, ARG4, ARGM} paired with pointers to phrases in Penn Treebank II trees. A noun instance is markable if it is accompanied by one of its arguments (ARG0, ARG1, ARG2, ARG3, ARG4) or if it is a nominalization (or similar word) and it is accompanied by one of the allowable types of adjuncts (ARGM-TMP, ARGM-LOC, ARGM-ADV, ARGM-EXT, etc.) – the same set of adjuncts used in PropBank.[3]

The basic idea is that each triple {REL, SENSE, ARGNUM} uniquely defines an argument, given a particular sense of a particular REL (or predicate), where ARGNUM is one of the numbered arguments (ARG0, ARG1, ARG2, ARG3, ARG4) and SENSE is one of the senses of that REL. The arguments are essentially the same as the initial relations of Relational Grammar (Perlmutter and Postal, 1984; Rosen, 1984). For example, agents tend to be classified as ARG0 (RG's initial subject), patients and themes tend to be classified as ARG1 (RG's initial object) and indirect objects of all kinds tend to be classified as ARG2.

The lexical entry or frame for each noun provides one inventory of argument labels for each sense of that word.[4] Each proposition (cf. figure 1) consists of an instance of an argument-taking noun (REL) plus arguments (ARG0, ARG1, ARG2, . . .), SUPPORT items and/or adjuncts (ARGM). SUPPORT items are words that link arguments that occur outside an NP to the nominal predicate that heads that NP, e.g., "made" SUPPORTS "We" as the ARG0 of *decision* in *We made a decision*. ARGMs are adjuncts of the noun. However, we only mark the sort of adjuncts that also occur in sentences: locations (ARGM-LOC), temporal (ARGM-TMP), sentence adverbial (ARGM-ADV) and various others.

3 Lexical Entries and Noun Classes

Before we could begin annotation, we needed to classify all the common nouns in the corpus. We needed to know which nouns were markable and make initial approximations of the inventories of senses and arguments for each noun. Toward this end, we pooled a number of resources: COMLEX Syntax (Macleod et al., 1998a), NOMLEX (Macleod et al., 1998b) and the verb classes from (Levin, 1993). We also used string matching techniques and hand classification in combination with programs that automatically merge crucial features of these resources. The result was NOMLEX-PLUS, a NOMLEX-style dictionary, which includes the original 1000 entries in NOMLEX plus 6000 additional entries (Meyers et al., 2004). The resulting noun classes include verbal nominalizations (e.g., *destruction*, *knowledge*, *believer*, *recipient*), adjectival nominalizations (*ability*, *bitterness*), and 16 other classes such as relational (*father*, *president*) and partitive nouns (*set*, *variety*). NOMLEX-PLUS helped us break down

[3]To make our examples more readable, we have replaced pointers to the corpus with the corresponding strings of words.

[4]For a particular noun instance, only a subset of these arguments may appear, e.g., the ARG2 (indirect object) *to Dorothy* can be left out of the phrase *Glinda's gift of the slippers*.

the nouns into classes, which in turn helped us gain an understanding of the difficulty of the task and the manpower needed to complete the task.

We used a combination of NOMLEX-PLUS and PropBank's lexical entries (or frames) to produce automatic approximations of noun frames for NomBank. These entries specify the inventory of argument roles for the annotators. For nominalizations of verbs that were covered in PropBank, we used straightforward procedures to convert existing PropBank lexical entries to nominal ones. However, other entries needed to be created by automatic means, by hand or by a combination of the two. Figure 2 compares the PropBank lexical entry for the verb *claim* with the NomBank entry for the noun *claim*. The noun *claim* and the verb *claim* share both the ASSERT sense and the SEIZE sense, permitting the same set of argument roles for those senses. However, only the ASSERT sense is actually attested in the sample PropBank corpus that was available when we began working on NomBank. Thus we added the SEIZE sense to both the noun and verb entries. The noun *claim* also has a LAWSUIT sense which bears an entry similar to the verb *sue*. Thus our initial entry for the noun *claim* was a copy of the verb entry at that time. An annotator edited the frames to reflect noun usage – she added the second and third senses to the noun frame and updated the verb frame to include the second sense.

In NOMLEX-PLUS, we marked *anniversary* and *advantage* as "cousins" of nominalizations indicating that their lexical entries should be modeled respectively on the verbs *commemorate* and *exploit*, although both entries needed to be modified in some respect. We use the term "cousins" of nominalizations to refer to those nouns which take argument structure similar to some verb (or adjective), but which are not morphologically related to that word. Examples are provided in Figure 3 and 4. For adjective nominalizations, we began with simple procedures which created frames based on NOMLEX-PLUS entries (which include whether the subject is +/-sentient). The entry for "accuracy" (the nominalization of the adjective *accurate*) plus a simple example is provided in figure 5 – the ATTRIBUTE-LIKE frame is one of the most common frames for adjective nominalizations. To cover the remaining nouns in the corpus, we created classes of lexical items and manually constructed one frame for each class. Each member of a class was was given the corresponding frame. Figure 6 provides a sample of these classes, along with descriptions of their frames. As with the nominalization cousins, annotators sometimes had to adjust these frames for particular words.

4 A Merged Representation

Beginning with the PropBank and NomBank propositions in Figure 7, it is straight-forward to derive the

1. ASSERT Sense

   Roles: ARG0 = AGENT, ARG1 = TOPIC

   Noun Example: *Her claim that Fred can fly*
   REL = claim, ARG0 = her, ARG1 = that Fred can fly

   Verb Example: *She claimed that Fred can fly*
   REL = claimed, ARG0 = She, ARG1 = that Fred can fly

2. SEIZE Sense

   Roles: ARG0 = CLAIMER, ARG1 = PROPERTY, ARG2 = BENEFICIARY

   Noun Example: *He laid claim to Mexico for Spain*
   REL = claim, SUPPORT = laid, ARG0 = He, ARG1 = to Mexico, ARG2 = for Spain

   Verb Example: *He claimed Mexico for Spain*
   REL = claim, ARG0 = He, ARG1 = Mexico, ARG2 = for Spain

3. SUE Sense

   Roles: ARG0 = CLAIMANT, ARG1 = PURPOSE, ARG2 = DEFENDANT, ARG3 = AWARD

   Noun Example: *His $1M abuse claim against Dan*
   ARG0 = His, ARG1 = abuse, ARG2 = against Dan, ARG3 = $1M

   Verb Example: NOT A VERB SENSE

Figure 2: Verb and Noun Senses of *claim*

1. HONOR (based on a sense of *commemorate*)

   Roles: ARG0 = agent, ARG1 = thing remembered, ARG2 = times celebrated

   Noun Example: Investors celebrated the second anniversary of Black Monday.
   REL = anniversary, SUPPORT = celebrated, ARG0 = Investors, ARG1 = of Black Monday, ARG2 = second

Figure 3: One sense for *anniversary*

1. EXPLOIT

   Roles: ARG0 = exploiter, ARG1 = entity exploited

   Noun Example: Investors took advantage of Tuesday 's stock rally.
   REL = advantage, SUPPORT = took, ARG0 = Investors, ARG1 = of Tuesday's stock rally

Figure 4: One sense for *advantage*

1. ATTRIBUTE-LIKE

   Roles: ARG1 = theme

   Noun Example: the accuracy of seasonal adjustments built into the employment data
   REL = accuracy, ARG1 = of seasonal adjustments built into . . .

Figure 5: One Sense for *accuracy*

ACTREL Relational Nouns with beneficiaries

Roles: ARG0 = JOB HOLDER, ARG1 = THEME, ARG2 = BENEFICIARY

Example: ACME will gain printing customers
REL = customers, SUPPORT = gain, ARG0 = customers, ARG1 = printing, ARG2 = ACME

DEFREL Relational Nouns for personal relationships

Roles: ARG0 = RELATION HOLDER, ARG1 = RELATION RECEPTOR

Example: public enemies REL = enemies, ARG0 = enemies, ARG1 = public

ATTRIBUTE Nouns representing attribute relations

Roles: ARG1 = THEME, ARG2 = VALUE

Example: a lower grade of gold
REL = grade, ARG1 = of gold, ARG2 = lower

ABILITY-WITH-AGENT Ability-like nouns

Roles: ARG0 = agent, ARG1 = action

Example: the electrical current-carrying capacity of new superconductor crystals
REL = capacity, ARG0 = of new superconductor crystals, ARG1 = electrical current-carrying

ENVIRONMENT **Roles:** ARG1 = THEME

Example: the circumstances of his departure
REL = circumstances, ARG1 = of his departure

Figure 6: Frames for Classes of Nouns

PropBank: REL = *gave*, ARG0 = *they*, ARG1 = *a standing ovation*, ARG2 = *the chefs*

NomBank: REL = *ovation*, ARG0 = *they*, ARG1 = *the chefs*, SUPPORT = *gave*

Figure 7: They gave the chefs a standing ovation

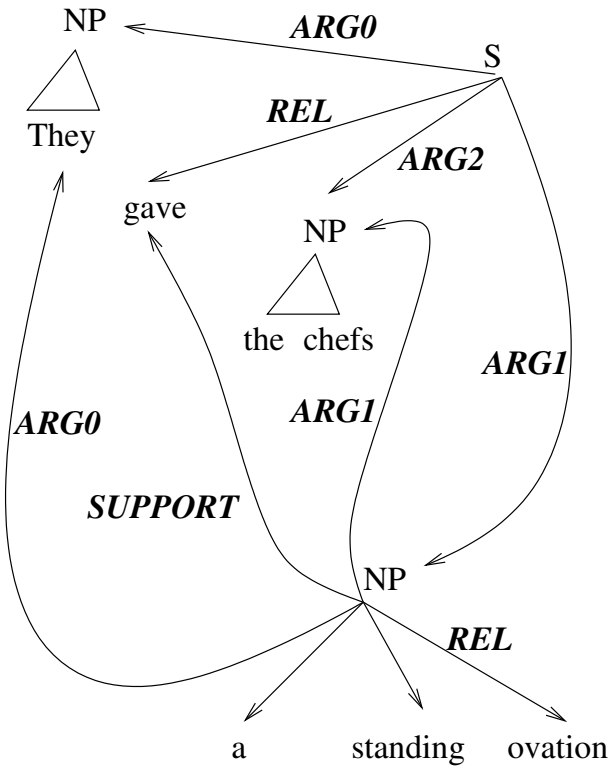

Figure 8: They gave the chefs a standing ovation

combined PropBank/NomBank graphical representation in Figure 8 in which each role corresponds to an arc label. For this example, think of the argument structure of the noun *ovation* as analogous to the verb *applaud*. According to our analysis, *they* are both the givers and the applauders and *the chefs* are both the recipients of something given and the ones who are applauded. *Gave* and *ovation* have two distinct directional relations: *a standing ovation* is something that is given and *gave* serves as a link between *ovation* and its two arguments. This diagram demonstrates how NomBank is being designed for easy integration with PropBank. We believe that this is the sort of predicate argument representation that will be needed to easily merge this work with other annotation efforts.

5 Analysis of the Task

As of this writing we have created the various lexicons associated with NomBank. This has allowed us to break down the task as follows:

- There are approximately 240,000 instances of common nouns in the PTB (approximately one out of every 5 words).
- At least 36,000 of these are nouns that cannot take arguments and therefore need not be looked at by an

annotator.

- There are approximately 99,000 instances of verbal nominalizations or related items (e.g., cousins)
- There are approximately 34,000 partitives (including 6,000 instances of the percent sign), 18,000 subject nominalizations, 14,000 environmental nouns, 14,000 relational nouns and fewer instances of the various other classes.
- Approximately 1/6 of the cases are instances of nouns which occur in multiple classes.[5]

The difficulty of the annotation runs the gamut from nominalization instances which include the most arguments, the most adjuncts and the most instances of support to the partitives, which have the simplest and most predictable structure.

6 Error Analysis and Error Detection

We have conducted some preliminary consistency tests for about 500 instances of verbal nominalizations during the training phases of NomBank. These tests yielded inter-annotator agreement rates of about 85% for argument roles and lower for adjunct roles. We are currently engaging in an effort to improve these results.[6]

We have identified certain main areas of disagreement including: disagreements concerning SUPPORT verbs and the shared arguments that go with them; disagreements about role assignment to prenominals; and differences between annotators caused by errors (typos, slips of the mouse, ill-formed output, etc.) In addition to improving our specifications and annotator help texts, we are beginning to employ some automatic means for error detection.

6.1 Support

For inconsistencies with SUPPORT, our main line of attack has been to outline problems and solutions in our specifications. We do not have any automatic system in effect yet, although we may in the near future.

SUPPORT verbs (Gross, 1981; Gross, 1982; Mel'ĉuk, 1988; Mel'ĉuk, 1996; Fontenelle, 1997) are verbs which connect nouns to one (or more) of their arguments via argument sharing. For example, in *John took a walk*, the verb *took* "shares" its subject with the noun *walk*. SUPPORT verbs can be problematic for a number of reasons. First of all the concept of argument sharing is not black and white. To illustrate these shades of gray, compare the relation of *Mary* to *attack* in: *Mary's attack against the alligator*, *Mary launched an attack against the alligator*, *Mary participated in an attack against the alligator*, *Mary planned an attack against the alligator* and *Mary considered an attack against the alligator*. In each subsequent example, Mary's "level of agency" decreases with respect to the noun *attack*. However, in each case *Mary* may still be viewed as some sort of potential attacker. It turned out that the most consistent position for us to take was to assume all degrees of argument-hood (in this case subject-hood) were valid. So, we would mark *Mary* as the ARG0 of *attack* in all these instances. This is consistent with the way control and raising structures are marked for verbs, e.g., *John* is the subject of *leave* and *do* in *John did not seem to leave* and *John helped do the project* under most accounts of verbal argument structure that take argument sharing (control, raising, etc.) into account.

Of course a liberal view of SUPPORT has the danger of overgeneration. Consider for example, *Market conditions led to the cancellation of the planned exchange*. The unwary annotator might assume that *market conditions* is the ARG0 (or subject) of *cancellation*. In fact, the combination *lead to* and *cancellation* do not have any of the typical features of SUPPORT described in figure 9. However, the final piece of evidence is that *market conditions* violate the selection restrictions of *cancellation*. Thus the following paraphrase is ill-formed **Market conditions canceled the planned exchange*. This suggests that *market conditions* is the subject of *lead* and not the subject of *cancellation*. Therefore, this is not an instance of support in spite of the apparent similarity.

We require that the SUPPORT relation be lexical. In other words, there must be something special about a SUPPORT verb or the combination of the SUPPORT verb and the noun to license the argument sharing relation. In addition to SUPPORT, we have cataloged several argument sharing phenomena which are markable. For example, consider the sentence, *President Bush arrived for a celebration*. Clearly, *President Bush* is the ARG0 of *celebration* (one of the people celebrating). However, *arrive* is not a SUPPORT verb. The phrase *for a celebration* is a subject-oriented adverbial, similar to adverbs like *willingly*, which takes the subject of the sentence as an argument. Thus *President Bush* could also be the subject of *celebration* in *President Bush waddled into town for the celebration* and many similar sentences that contain this PP.

Finally, there are cases where argument sharing may

[5] When a noun fits into multiple categories, those categories may predict multiple senses, but not necessarily. For example, *drive* has a nominalization sense (*He went for a drive*) and an attribute sense (*She has a lot of drive*). Thus the lexical entry for *drive* includes both senses. In constrast, *teacher* in the *math teacher* has the same analysis regardless of whether one thinks of it as the nominalization of *teach* or as a relational (ACTREL) noun.

[6] Consistency is the average precision and recall against a gold standard. The preliminary tests were conducted during training, and only on verbal nominalizations.

- Support verb/noun pairs can be idiosyncratically connected to the point that some researchers would call them idioms or phrasal verbs, e.g., *take a walk, keep tabs on*.
- The verb can be essentially "empty", e.g., *make an attack*, *have a visit*.
- The "verb/noun" combination may take a different set of arguments than either does alone, e.g., *take advantage of*.
- Some support verbs share the subject of almost any nominalization in a particular argument slot. For example *attempt* shares its subject with most following nominalizations, e.g., *He attempted an attack*. These are the a lot like raising/control predicates.
- In some cases, the support verb and noun are from similar semantic classes, making argument sharing very likely, e.g., fight a battle.

Figure 9: Possible Features of Support

be implied by discourse processes, but which we do not mark (as we are only handling sentence-level phenomena). For example, the words *proponent* and *rival* strongly imply that certain arguments appear in the discourse, but not necessarily in the same sentence. For example in *They didn't want the company to fall into the hands of a rival*, there is an implication that *the company* is an ARG1 of *rival*, i.e., *a rival* should be interpreted as *a rival of the company*.[7] The connection between *a rival* and *the company* is called a "bridging" relation (a process akin to coreference, cf. (Poesio and Vieira, 1998)) In other words, *fall into the hands of* does not link "rival" with *the company* by means of SUPPORT. The fact that a discourse relation is responsible for this connection becomes evident when you see that the link between *rival* and *company* can cross sentence boundaries, e.g., *The company was losing money. This was because a rival had come up with a really clever marketing strategy.*

6.2 Prenominal Adjectives and Error Detection

ARGM is the annotation tag used for nonarguments, also known as adjuncts. For nouns, it was decided to only tag such types of adjuncts as are also found with verbs, e.g., temporal, locative, manner, etc. The rationale for this included: (1) only the argument-taking common nouns are being annotated and other sorts of adjuncts occur with common nouns in general; (2) narrowing the list of potential labels helped keep the labeling consistent; and (3) this was the minimum set of adjuncts that would keep the noun annotation consistent with the verb annotation.

Unfortunately, it was not always clear whether a prenominal modifier (particularly an adjective) fell into one of our classes or not. If an annotator felt that a modifier was somehow "important", there was a temptation to push it into one of the modifier classes even if it was not a perfect fit. Furthermore, some annotators had a broader view than others as to the sorts of semantic relationships that fell within particular classes of adjuncts, particularly locative (LOC), manner (MNR) and extent (EXT). Unlike the SUPPORT verbs, which are often idiosyncratic to particular nominal predicates, adjunct prenominal modifiers usually behave the same way regardless of the noun with which they occur.

In order to identify these lexical properties of prenominals, we created a list of all time nouns from COMLEX Syntax (ntime1 and ntime2) and we created a specialized dictionary of adjectives with adverbial properties which we call ADJADV. The list of adjective/adverb pairs in ADJADV came from two sources: (1) a list of adjectives that are morphologically linked to -ly adverbs created using some string matching techniques; and (2) adjective/adverb pairs from CATVAR (Habash and Dorr, 2003). We pruned this list to only include adjectives found in the Penn Treebank and then edited out inappropriate word pairs. We completed the dictionary by transferring portions of the COMLEX Syntax adverb entries to the corresponding adjectives.

We now use ADJADV and our list of temporal nouns to evaluate NOMBANK annotation of modifiers. Each annotated left modifier is compared against our dictionaries. If a modifier is a temporal noun, it can bear the ARGM-TMP role (temporal adjunct role), e.g., the temporal noun *morning* can fill the ARGM-TMP slot in *the morning broadcast*. Most other common nouns are compatible with argument role slots (ARG0, ARG1, etc.), e.g., the noun *news* can fill the ARG1 slot in *the news broadcast*. Finally, roles associated with adjectives depend on their ADJADV entry, e.g., *possible* can be an ARGM-ADV in *possible broadcasts* due to the epistemic feature encoded in the lexical entry for *possible* (derived from the corresponding adjverb *possibly*). Discrepancies between these procedures and the annotator are resolved on a case by case basis. If the dictionary is wrong, the dictionary should be changed, e.g., *root*, as in *root cause* was added to the dictionary as a potential MNR adjective with a meaning like the adverb *basically*. However, if the annotator is wrong, the annotation should be changed, e.g., if an annotator marked "slow" as a ARGM-TMP, the program would let them know that it should be a ARGM-MNR. This process both helps with annotation accuracy and enriches our lexical database.

[7] The noun *rival* is a subject nominalization of the verb *rival*.

6.3 Other Automatically Detected Errors

We used other procedures to detect errors including:

Nom-type Argument nominalizations are nominalizations that play the role of one of the arguments in the ROLESET. Thus the word *acquirer* should be assigned the ARG0 role in the following example because *acquirer* is a subject nominalization:

a possible acquirer of Manville
REL = *acquirer*, ARG0 = *acquirer*, ARG1 = *of Manville*, ARGM-ADV = possible

A procedure can compare the NOMLEX-PLUS entry for each noun to each annotated instance of that noun to check for incompatibilities.

Illformedness Impossible instances are ruled out. Checks are made to make sure obligatory labels (REL) are present and illegal labels are not. Similarly, procedures make sure that infinitive arguments are marked with the -PRD function tag (a PropBank convention).

Probable Illformedness Certain configurations of role labels are possible, but very unlikely. For example, the same argument role should not appear more than once (the stratal uniqueness condition in Relational Grammar or the theta criterion in Principles and parameters, etc.). Furthermore, it is unlikely for the first word of a sentence to be an argument unless the main predicate is nearby (within three words) or unless there is a nearby support verb. Finally, it is unlikely that there is an empty category that is an argument of a predicate noun unless the empty category is linked to some real NP.[8]

WRONG-POS We use procedures that are part of our systems for generating GLARF, a predicate argument framework discussed in (Meyers et al., 2001a; Meyers et al., 2001b), to detect incorrect parts of speech in the Penn Treebank. If an instance is predicted to be a part of speech other than a common noun, but it is still tagged, that instance is flagged. For example, if a word tagged as a singular common noun is the first word in a VP, it is probably tagged with the wrong part of speech.

6.4 The Results of Error Detection

The processes described in the previous subsections are used to create a list of annotation instances to check along with short standardized descriptions of what was wrong, e.g., wrong-pos, non-functional (if there were two identical argument roles), etc. Annotators do a second pass on just these instances (currently about 5 to 10% of the total). We will conduct a formal evaluation of this procedure over the next month.

[8]Empty categories mark "invisible" constituents in the Treebank, e.g., the subject of *want* in *John*$_1$ *wanted* e_1 *to leave*.

PARTITIVE-QUANT

Roles: ARG1 = QUANTIFIED
Example: *lots of internal debate*
REL = lots, ARG1 = of internal debate

Figure 10: The entry for *lot*

7 Future Research: Automatic Annotation

We are just starting a new phase in this project: the creation of an automatic annotator. Using techniques similar to those described in (Meyers et al., 1998) in combination with our work on GLARF (Meyers et al., 2001a; Meyers et al., 2001b), we expect to build a hand-coded PROPBANKER a program designed to produce a PropBank/NomBank style analysis from Penn Treebank style input. Although the PropBanker should work with input in the form of either treebank annotation or treebank-based parser output, this project only requires application to the Penn Treebank itself. While previous programs with similar goals (Gildea and Jurafsky, 2002) were statistics-based, this tool will be based completely on hand-coded rules and lexical resources.

Depending on its accuracy, automatically produced annotation should be useful as either a preprocessor or as an error detector. We expect high precision for very simple frames, e.g., nouns like *lot* as in figure 10. Annotators will have the opportunity to judge whether particular automatic annotation is "good enough" to serve as a preprocessor. We hypothesize that a comparison of automatic annotation that fails this level of accuracy against the hand annotation will still be useful for detecting errors. Comparisons between the hand annotated data and the automatically annotated data will yield a set of instances that warrant further checking along the same lines as our previously described error checking mechanisms.

8 Summary

This paper outlines our current efforts to produce NomBank, annotation of the argument structure for most common nouns in the Penn Treebank II corpus. This is part of a larger effort to produce more detailed annotation of the Penn Treebank. Annotation for NomBank is progressing quickly. We began with a single annotator while we worked on setting the task and have ramped up to four annotators. We continue to work on various quality control procedures which we outline above. In the near future, we intend to create an automatic annotation program to

be used both as a preprocessor for manual annotation and as a supplement to error detection.

The argument structure of NPs has been less studied both in theoretical and computational linguistics, than the argument structure of verbs. As with our work on NOMLEX, we are hoping that NomBank will substantially contribute to improving the NLP community's ability to understand and process noun argument structure.

Acknowledgments

Nombank is supported under Grant N66001-001-1-8917 from the Space and Naval Warfare Systems Center San Diego. This paper does not necessarily reflect the position or the policy of the U.S. Government.

We would also like to acknowledge the people at the University of Pennsylvania who helped make NomBank possible, including, Martha Palmer, Scott Cotton, Paul Kingsbury and Olga Babko-Malaya. In particular, the use of PropBank's annotation tool and frame files proved invaluable to our effort.

References

T. Fontenelle. 1997. *Turning a bilingual dictionary into a lexical-semantic database*. Lexicographica Series Maior 79. Max Niemeyer Verlag, Tübingen.

D. Gildea and D. Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28:245–288.

M. Gross. 1981. 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.

M. Gross. 1982. 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*.

N. Habash and B. Dorr. 2003. CatVar: A Database of Categorial Variations for English. In *Proceedings of the MT Summit*, pages 471–474, New Orleans.

P. Kingsbury and M. Palmer. 2002. From treebank to propbank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Spain.

P. Kingsbury, M. Palmer, and Mitch Marcus. 2002. Adding semantic annotation to the penn treebank. In *Proceedings of the Human Language Technology Conference*, San Diego, California.

B. Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press, Chicago.

C. Macleod, R. Grishman, and A. Meyers. 1998a. COMLEX Syntax. *Computers and the Humanities*, 31(6):459–481.

C. Macleod, R. Grishman, A. Meyers, L. Barrett, and R. Reeves. 1998b. Nomlex: A lexicon of nominalizations. In *Proceedings of Euralex98*.

I. A. Mel'ĉuk. 1988. *Dependency Syntax: Theory and Practice*. State University Press of New York, Albany.

I. A. Mel'ĉuk. 1996. 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.

A. Meyers, C. Macleod, R. Yangarber, R. Grishman, Leslie Barrett, and Ruth Reeves. 1998. Using NOMLEX to Produce Nominalization Patterns for Information Extraction. In *Coling-ACL98 workshop Proceedings: the Computational Treatment of Nominals*.

A. Meyers, R. Grishman, M. Kosaka, and S. Zhao. 2001a. Covering Treebanks with GLARF. In *ACL/EACL Workshop on Sharing Tools and Resources for Research and Education*.

A. Meyers, M. Kosaka, S. Sekine, R. Grishman, and S. Zhao. 2001b. Parsing and GLARFing. In *Proceedings of RANLP-2001*, Tzigov Chark, Bulgaria.

A. Meyers, R. Reeves, Catherine Macleod, Rachel Szekeley, Veronkia Zielinska, Brian Young, and R. Grishman. 2004. The Cross-Breeding of Dictionaries. In *Proceedings of LREC-2004*, Lisbon, Portugal. To appear.

MUC-6. 1995. *Proceedings of the Sixth Message Understanding Conference*. Morgan Kaufman. (MUC-6).

D. M. Perlmutter and P. M. Postal. 1984. 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.

M. Poesio and R. Vieira. 1998. A Corpus-based Investigation of Definite Description Use. *Computational Linguistics*, 24(2):183–216.

C. G. Rosen. 1984. 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.

University of Pennsylvania. 2002. Annotation guidelines for PropBank. `http://www.cis.upenn.edu/~ace/propbank-guidelines-feb02.pdf`.