

Those Other NomBank Dictionaries – Manual for Dictionaries that Come with NomBank

Adam Meyers

1 Introduction

The NomBank frame files and the morphological dictionary, which are essential to using NomBank, are described in the document *Annotation Guidelines for NomBank – Noun Argument Structure for PropBank 2007* (Referred to below as the NOMBANK Guidelines). The focus of this document is on other dictionaries which were a side effect of creating NomBank.

These other dictionaries were created to help with development of NomBank. We decided to release these along with NomBank so that NomBank users could use these dictionaries as well. These dictionaries include:

1. NOMLEX-PLUS, an extension of NOMLEX [3], provides a dictionary based resource for identifying noun argument structure.
2. ADJADV is a dictionary for mapping adjective modifiers of nouns to corresponding instances of adverbial modifiers of verbs. For example, it would help identify a correspondence between *slow* in *the slow evaporation of the moisture* and *slowly* in *The moisture evaporated slowly*.
3. NOMADV is a resource for mapping nouns to adverbs. It is useful for identifying nouns that take adverbial arguments, e.g., *particular* in *In particular, I like it when you sing jazzy songs*; and it is also useful for identifying nouns that are modifiers of their arguments, for example *The improbability of life on the moon* implies an ARG-MOD modification of the proposition licensed by *life*, i.e., *Something probably does not live on the moon*.
4. COMNOM is an automatically enriched version of COMLEX Syntax [2]. The focus was to enrich the noun complement structure from the NOMLEX entries, adding PP complements for the first time, as well as increasing coverage of sentential complements. Automatic routines were used to derive these additional complements.

```

(X :FEATURE1 "value"
  :FEATURE2 ("1" "2" "3")
  :FEATURE3 (Y)
  :FEATURE4 ((Q :FEATUREa T)
              (R :FEATUREb "M")
              (S)))

```

Figure 1: A Lexical Entry Schema of the COMLEX style

5. Nombank-Dict is a lisp-like version of the frames files that have been automatically enhanced to include generalized descriptors (GENDESCR). These entries are equivalent to their xml counterparts, except that a heuristic was applied to minimize the different descriptor labels being used. For example, ACTOR becomes AGENT, as does most ARG0 labels that end in "ER" (EXPERIENCER is an exception).

2 Format

The dictionaries are in a lisp-like notation like COMLEX Syntax [2] and NOMLEX [3]. Each entry is a typed feature structure consisting of a type and a set of feature value pairs. Each value can be either an atom (usually T), a string (indicated with quotation marks), a list of strings, a typed feature structure or a list of typed feature structures. A type can have zero or more feature valued pairs. For example, Figure 1 is a possible lexical entry in a hypothetical dictionary in this format. X is the type of feature structure which can occur with FEATURES 1, 2, 3 and 4. FEATURE1 takes a string as a value; FEATURE2 takes a list of strings as a value, FEATURE3 takes a single typed feature structure Y as a value (and Y occurs with no features); FEATURE4 takes a list of typed feature structures as its value – Q occurs with FEATUREa, R with FEATUREb and S with no features. :FEATUREa takes T as its value and FEATUREb takes a string as its value.

3 Versions Sensitive to Training Corpora

There are three versions of NOMLEX-PLUS and COMNOM corresponding to how much corpus training influenced their creation:

- The unsuffixed name (NOMLEX-PLUS and COMNOM) plus their version numbers, e.g., 1.0, is used for the versions trained on the entire Wall Street Journal PropBank/NomBank (Penn Treebank II) corpus.

- The -tr version was trained on just sections 02–21 of the corpus. These versions are designed for those who wish to divide the WSJ corpus in the customary manner. These versions contain information only from the standard training corpus.
- The -clean versions do not use corpus training at all. This is for those who would rather not use the corpus-based enhancements that NYU implemented.

4 Distribution

With the exception of COMNOM, all these dictionaries will be distributed with NomBank. I will have to figure out another way to distribute COMNOM, probably via the LDC, who holds the COMLEX license.

5 NOMLEX-PLUS

NOMLEX [3] is a dictionary describing 1025 nouns that are morphologically related to verbs. NOMLEX-PLUS is an extension of this dictionary to over 8000 entries using a variety of semi-automatic processes described in part in [4]. There are now nearly 4000 nominalizations of verbs (NOM) and an additional 1600 entries with similar dictionary structures (NOMLIKE, NOMING, ABLE-NOM). The remaining entries fall into various other classes which are also described in the NOMBANK manual. Choice of category (NOM, NOMLIKE, etc.) was manual, but the filling in of the details of the entries was based on a combination of COMLEX entries for corresponding verbs plus some heuristic rules, e.g., the NOMLEX-PLUS entry for the noun *abuse* was based on the COMLEX entry for the verb *abuse*. In a similar manner, the non-nominalization entries were based partially on defaults and partially on lexical specific information. In addition, there was sporadic editing by hand, hence, these entries are semi-automatically created. Note that it took several people years of focused work to create NOMLEX's 1000 entries, whereas the additional 7000 entries of NOMLEX-PLUS were created by one person here and there over the years. The original NOMLEX entries were manually created and fine-tuned, whereas the original NOMLEX-PLUS entries were created by heuristics.

5.1 Nominalizations of Verbs and Similar Entries

Nominalizations (NOM) and the similar entries (NOMLIKE, NOMING and ABLE-NOM) of NOMLEX-PLUS are mostly the same as the NOM entries for the regular-

ized version of the original NOMLEX described in [5] (<http://nlp.cs.nyu.edu/nomlex/nomlex-alpha-reg-manual.ps>). There are, however, the following additions:

1. :IND-OBJ-ATTRIBUTE – An additional feature, similar to :SUBJ-ATTRIBUTE and :OBJ-ATTRIBUTE, but lists semantic features of indirect objects.
2. :SEMI-AUTOMATIC – :SEMI-AUTOMATIC T indicates that an entry is not from the original NOMLEX and is the product of procedures which at least partially rely on heuristics.¹
3. :SVERB – for a small subset of items, possible support verbs are indicated with this feature.
4. The VALUES for :SUBJ-ATTRIBUTE, :OBJ-ATTRIBUTE and :IND-OBJ-ATTRIBUTE have been expanded to a larger set, listed in the table 1. These features represent a form of semantic selection restrictions. Some of these correspond to Named Entity classes and Complex Syntax semantic classes (values of :FEATURES).
5. The VMOTION feature of COMLEX verbs is now also marked on nominalizations. This indicates that a noun (or verb) can occur with a locative adverbs indicating motion in a direction, e.g., in *The advance upward*. For nominalizations, adjectives with locative meanings can also be used, e.g., *The diagonal advance*. These adjectives are marked with the feature LOC&DIR-ADV in the ADJADV dictionary discussed below.
6. There are the following additional types of argument nominalizations as indicated by values of :NOM-TYPE –
 - (a) COMP – Indicates that the nominalization incorporates the meaning of a sentential of predicate complement of the verb. For example, *nickname* acts like the predicative NP complement of the verb, *John's nickname* can corresponds to the *JJ* in *They nicknamed John JJ*.
 - (b) P-OBJ – The nominalization corresponds to a prepositional object of the verb (that is arguably not an indirect object). For example, *culmination* is a (P-OBJ :PVAL (“with” “in”)) type of nominalization. For example, suppose *The culmination of the process* is coreferential with

¹Although SEMI-AUTOMATIC indicates entries that were not created by hand, all entries may have been automatically updates. Specifically, using the annotated corpus, we added additional prepositions to the NOMLEX entries as well as selection restrictions (values of :SUBJ-ATTRIBUTE, :OBJ-ATTRIBUTE and :IND-OBJ-ATTRIBUTE).

the phrase *the invention of the spoon*. Then *The culmination of the process* can be paraphrased as *The process culminated with the invention of the spoon*.

- (c) INSTRUMENT – The nominalization corresponds to an instrument of the corresponding verb, e.g., *The cure for the disease* refers to the thing used by someone to cure the disease.
- (d) OBJECT-PART, SUBJECT-PART, and P-OBJ-PART are subject, object and prepositional object nominalizations that incorporate a particle. For example, *runaway* is a (SUBJECT-PART :ADVAL (“away”)) nominalization of *run* corresponding to the subject of *run away*.

As discussed, there are several nominalization-like classes in addition to the NOM class, which like the original NOMLEX, includes only nominalizations that are morphologically related to verbs and that have basically the same meaning as the verbal form. The other classes can be described as follows:

1. NOMING – These are ING forms of verbs (gerunds or present participles) that occur in a noun environment in the PENN TREEBANK corpus. There is some gray area as to what a noun environment is and there is also some gray area as to whether these -ing forms have or have not been lexicalized, i.e., function as true nouns regardless of their context. As noted in NOMBANK Guidelines, we mostly rely on Penn’s part of speech annotation to determine that these are nouns (except for obvious errors).
2. ABLE-NOM – These are morphologically related to a verb and share argument structure, but have slightly different meanings which can be paraphrased using the noun *ability*. Furthermore, these nouns usually include one of the following suffixes: *-ability*, *-ibility*, *-able*, *-ism* or *-ness*. For example, *accountability* and *affordability* mean something like *the ability to account* and *ability to afford*.
3. NOMLIKE – These entries are for nouns with rich argument structure that based on a verb which they are not morphologically related to, e.g., the NOMLIKE entry for *aberration* is based on the verb *err*. In most cases, there is no morphologically related verb. However, in other cases, the intended meaning of the noun is closer to a verb that it is not morphologically related to, e.g., the NOMLIKE entry for *whitewash* is related to the verb *gloss* as in *gloss over* even though there is another NOM entry related to the verb *whitewash*. Finally, there are some NOMLIKE entries which attempt link the noun to a verb that is found in PropBank because the morphologically related verb is not found in PropBank, e.g., there is a NOMLIKE entry

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 <i>with</i> 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	<i>about</i> phrases, <i>that</i> 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

Table 1: Selectional Features – Values of :SUBJ-/:OBJ-/:IND-OBJ-ATTRIBUTE

for *yearning* based on the verb *desire* as *yearn* was not found in PropBank. This last feature is an artifact of the task used to develop NOMLEX-PLUS, e.g., NOMLEX entries were used to develop initial frame files which were then edited to fit actual usage, so development of NOMBANK went more smoothly if we had a PropBank model for such nouns.

5.2 NOMADJ and NOMADJLIKE

The entries for nominalization of Adjectives (NOMADJ) and entries based on adjectival meanings (NOMADJLIKE) are similar to the verbal nominalizations. The initial entries were based on heuristics, defaults, the COMLEX entry for the corresponding adjectives and an additional binary interpretation of whether or not the subject of the corresponding adjective tended to be human (an agent or experiencer) and/or a beneficiary.²³

The subcategorizations (values of :ADJ-SUBC) from Table 2 are used, along with the same sort of mapping information as NOMLEX. For most of the instances in the table, the corresponding COMLEX Syntax adjective class is derived by removing the NOM- prefix. Detailed definitions of the :SUBC features for adjectives are found in the COMLEX Syntax manual [6] and related COMLEX documentation. Each item in the table is listed along with the NOMBANK subcategorization phrase that is assumed to go with it when applicable (see the NOMBANK manual for details). In the case of NOM-ADJ-PP/ADJ-PP, a :PVAL is given indicating that a PP complement with the a preposition from the list indicated is allowed – a one to one mapping is assumed between the PP complement of the noun and the PP complement of the corresponding adjective. In the case of the intransitives, we divide up the cases where the phrase corresponding to the subject of the adjective is constrained to be humanlike (NOM-INTRANS) or is not so-constrained (NOM-OBJ-INTRANS)⁴. For example, the noun *absence* is marked NOM-OBJ-INTRANS because the corresponding adjective *absent* places no restriction on the corresponding adjective. In contrast, *wisdom* takes a NOM-INTRANS complement because only humans can be *wise*. The NOM-INTRANS-RECIP class marks an alternation between a plural subject and a singular subject with a PP complement (typically with the preposition *with*). The class is taken from COMLEX Syntax verbs and (for a few cases) has been extended to COMNOM adjectives (section 8) and the NOMADJ/NOMADJLIKE entries derived therefrom. For ex-

²This determination is sometimes stipulated and sometimes deduced heuristically from the set of adjective complements.

³The initial entries are occasionally edited by hand, hence the SEMI-AUTOMATIC classification.

⁴Note that intransitive is not marked as a SUBC for adjectives in COMLEX since it is assumed to be a possibility for all adjectives (other than those marked with the feature ATTRIBUTIVE)

NOMADJ/NOMADJLIKE values of :ADJ-SUBC	
NOMADJ CLASS	NOTE/SUBC/PVAL
NOM-EXTRAP-ADJ-FOR-TO-INF-NP-OMIT	(FOR-TO-INF)
NOM-EXTRAP-ADJ-FOR-TO-INF-RS	(FOR-TO-INF)
NOM-EXTRAP-ADJ-FOR-TO-INF	(FOR-TO-INF)
NOM-EXTRAP-ADJ-S-SUBJUNCT	(SENT :S-SUBJUNCT T)
NOM-EXTRAP-ADJ-WH-S	(P-WH)
NOM-EXTRAP-ADJ-S	(SENT :THAT-S T)
NOM-EXTRAP-ADJ-THAT-S	(SENT :THAT-S T)
NOM-ADJ-TO-INF	(TO-INF :SC T)
NOM-FOR-TO-ADJ	(FOR-TO-INF)
NOM-S-ADJ	(SENT :THAT-S T)
NOM-THAT-S-ADJ	(SENT :THAT-S T)
NOM-S-SUBJUNCT-ADJ	(SENT :S-SUBJUNCT T)
NOM-S-WH-ADJ	(P-WH :WH-S T)
NOM-ADJ-PP	:PVAL
NOM-INTRANS	HUMAN subject (not adjective class)
NOM-OBJ-INTRANS	any subject (not adjective class)
NOM-INTRANS-RECIP	COMNOM class*
NOM-COMPARATIVE	SPECIAL

Table 2: Adjective ↔ Noun Subcategorization Mapping Features

ample, the NOMADJ *compatibility* is marked NOM-INTRANS-RECIP, as is the adjective *compatible* in COMNOM, e.g., *Mary and John are compatible*, *Mary is compatible with John*, *John and Mary's compatibility*, *Mary's compatibility with John*. There are exactly 3 nouns that are classified as NOM-COMPARATIVE: *better*, *elder* and *lesser*. These are subject nominalizations of comparative adjectives spelled and pronounced the same way.

The only possible value of :FEATURES associated with NOMADJ and NOMADJLIKE entries is GRADABLE. As with the corresponding COMPLEX Syntax adjective entries, this indicates whether the property associated with the adjective can occur in degrees. A GRADABLE ADJNOMs and ADJNOMLIKES like *ability*, *abundance* and *adventurism* include gradable properties, e.g., it is possible to have a little or a lot of *ability* or *adventurism*. There can be great *abundance* or just a little. When GRADABLE NOMADJ/NOMADJLIKES are also ATTRIBUTE nouns (see the NOMBANK manual), the VALUE role sometimes corresponds to

this quantification of the property.

Most NOMADJ and NOMADJLIKE entries are given the NOM-TYPE ADJ-NOM (similar to the VERB-NOM type used for verbal nominalizations). However, SUBJECT nominalizations like *artist* and *expert* are also possible.⁵

5.3 Entries for NOMBANK's 16 Classes

NOMBANK introduces 16 classes of nouns that need not be associated with another part of speech: *RELATIONAL*, *JOB*, *HALLMARK*, *PARTITIVE*, *SHARE*, *GROUP*, *ENVIRONMENT*, *ABILITY*, *WORK-OF-ART*, *VERSION*, *TYPE*, *ATTRIBUTE*, *FIELD*, *ISSUE*, *CRISS-CROSS*, *EVENT*. Entries for these classes are in the form of figure 3, where for a given type (where type is a member of the 16 class names), only a subset of the following features are used: {*:SUBJECT*, *:OBJECT*, *:IND-OBJ*, *:OBLIQUE1*, *:OBLIQUE2*, *:NOM-SUBC*, *:SUBJ-ATTRIBUTE*, *:OBJ-ATTRIBUTE*, *:IND-OBJ-ATTRIBUTE*}. The remaining features are either obligatory {*:ORTH*, *:NOM-TYPE*, *:SEMI-AUTOMATIC*} or possible {*:PLURAL*, *:FEATURES*} for any entry belonging to one of the 16 classes.⁶ In table 3, the values of the features are given as either strings, lists of strings, the atom *NONE*, the atom T or a list of typed feature structures. Most of these features and their values have already been defined in the NOMBANK manual and this document. However, there are some minor differences that pertain to the usage of these features with respect to these 16 classes, as discussed below.⁷

Entries schematized as figure 3 are essentially a merger of one of the typed feature structures found as a value of *:VERB-SUBC* or *:ADJ-SUBC* (giving *:SUBJECT*, *:OBJECT*, etc. features) and the typed feature structures describing the main entries (giving *:ORTH*, *:PLURAL*, *:NOM-TYPE*, etc. features). For non-nominalizations, *:SUBJECT*, *:OBJECT*, *:IND-OBJ* and the corresponding *SUBJ-ATTRIBUTE*, *:OBJ-ATTRIBUTE* *:IND-OBJ-ATTRIBUTE* features have a somewhat different interpretation than with nominalizations. Clearly, these entries should not be interpreted as instructions for mapping nominalizations to verbs. Rather, these are indications of how particular arguments of the nouns appear. The *:SUBJECT*, *:OBJECT* and *:IND-OBJ* are respectively translations of the NOMBANK ARG0, ARG1 and ARG2 roles. It is convenient to use these terms for a number of reasons, not the least being that the same predictions should be made regardless of

⁵There are also a lot of NOMADJLIKE subject nominalizations based on the adjective *expert*, e.g., *buff*, *iczar*, *genius*.

⁶If NOMLEX-PLUS was hand-corrected as part of a future project, *:SEMI-AUTOMATIC* would become an optional feature.

⁷This section will focus on understanding the lexical entries. For a better understanding of the semantic differences between these classes, please refer to the NOMBANK manual.

(TYPE	:ORTH	“string”
	:PLURAL	“string” or list of strings or *NONE*
	:NOM-TYPE	list of TFS
	:FEATURES	list of TFS
	:SUBJECT	list of TFS
	:OBJECT	list of TFS
	:IND-OBJ	list of TFS
	:OBLIQUE1	list of TFS
	:OBLIQUE2	list of TFS
	:NOM-SUBC	list of TFS
	:SUBJ-ATTRIBUTE	list of TFS
	:OBJ-ATTRIBUTE	list of TFS
	:IND-OBJ-ATTRIBUTE	list of TFS
	:SEMI-AUTOMATIC	T)

Table 3: A Lexical Entry Schema for Non-nominalizations in NomBank

whether one looks at the nonnominalization entry or the nominalization entry for a word that has two entries, e.g., *director*.

Different sets of :NOM-TYPE values are used for nonnominalizations, as indicated in table 4. Many of these nouns are given a generic NOM-REL type. However, CRISSCROSS and EVENT nouns which have their own idiosyncratic types; HALLMARK and RELATIONAL nouns which are given the SUBJECT type (as they are similar to subject nominalizations); and ATTRIBUTE nouns are typically marked (IND-OBJ :REF). ATTRIBUTE nouns in NOMBANK are typically assigned the ARG2 (value) argument when this argument is not present in the noun phrase. For example, the attribute noun *weight* in *John’s weight* is treated as the ARG2 of weight because the whole phrase is typically associated with this argument by coreference or predication, e.g., *John’s weight was 180 pounds*. In addition, nonnominalizations sometimes seem to incorporate the ARG1 (represented here as OBJECT). For example, *nationalist* takes the argument structure of an ATTRIBUTE noun, but incorporates the OBJECT role, the one assigned the nationalism attribute. Similarly, *brainchild* acts like the argument nominalization of a WORK-OF-ART noun, filling the OBJECT role (the thing created). Finally, there is a productive subclass of JOB nouns that incorporate object role, in this case, the title being sought, e.g., *chairmanship*, *dictatorship*, *directorship*, *mayoralty*, *presidency*.

Entity Type	Possible :NOM-TYPE values
ABILITY	(NOM-REL)
ATTRIBUTE	(IND-OBJ :REF T) (OBJECT)
CRISSCROSS	(CRISSCROSS)
ENVIRONMENT	(NOM-REL)
EVENT	(EVENT)
FIELD	(NOM-REL)
GROUP	(NOM-REL)
HALLMARK	(SUBJECT)
ISSUE	(NOM-REL)
JOB	(NOM-REL) (OBJECT)
PARTITIVE	(NOM-REL)
RELATIONAL	(SUBJECT)
SHARE	(NOM-REL)
TYPE	(NOM-REL)
VERSION	(NOM-REL)
WORK-OF-ART	(NOM-REL) (OBJECT)

Table 4: :NOM-TYPE Values for Nonnominalizations

There are a number of additional values of :FEATURES that are possible for nonnominalizations. The most useful feature is probably TRANSPARENT. As discussed in the NOMBANK manual, a noun marked with the TRANSPARENT feature can take the semantic properties of its :OBJECT argument. For example, *a variety of sandwiches* and *a bottle of beer* can be interpreted as *sandwiches* and *beer* rather than *varieties* or *bottles*. Thus one can eat *a variety of sandwiches* and drink *a bottle of beer*. This feature is captured by marking both *variety* and *bottle* with the feature (TRANSPARENT). As discussed in the NOMBANK manual, nouns with this feature can also participate in SUPPORT structures, e.g., *He made a variety of mistakes*. By default, a subset of partitives (those with PARTITIVE-QUANT rolesets) and SHARE nouns are assumed to be transparent. In addition since GROUP nouns are assumed to be somewhat redundant with PARTITIVES, a noun that is marked both PARTITIVE-PART and GROUP in NOMBANK will end up as a transparent GROUP noun in NOMLEX-PLUS.

Additional values of :FEATURE are used to differentiate subtypes of PARTITIVE nouns, e.g., LOCATIVE-UNIT and TEMPORAL-UNIT respectively identify nouns like *mile*, *acre*, *inch* and *month*, *second*, *week*. The features MERONYM and PIECE identify PARTITIVE/MERONYM and PARTITIVE/PIECE subclasses.⁸

⁸The roleset names for these contain the strings “MERONYM” and “PARTITIVE-PART”.

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

Table 5: COMLEX Syntax ADVERB FEATURES ↔ NOMBANK ARGMs

The subclasses of relational nouns can be identified as follows: (i) DEFREL nouns do not have :IND-OBJ features, but ACTREL nouns do; and (ii) The feature CONSULTANT identifies the CONSULTANT subtype of ACTREL nouns.

Finally, the features OBJ-IS-NOM-SUBC and IND-OBJ-IS-NOM-SUBC are used to indicate that the nom-subc feature fills the same role as the object (obj-is-nom-subc) or indirect object (ind-obj-is-nom-subc). ABILITY, ENVIRONMENT and WORK-OF-ART nouns are typically assigned the OBJ-IS-NOM-SUBC feature and ATTRIBUTE nouns is typically assigned the IND-OBJ-IS-NOM-SUBC feature.

6 ADJADV

ADJADV is a dictionary with the purpose of mapping adjectives to adverbs using COMLEX Syntax adverbial classes. The main purpose is to provide a basis to semantically classify adjective modifiers of nouns using adverbial classes, e.g., to give *the slow evaporation of the liquid* a similar interpretation as *The liquid evaporated slowly*. The features of COMLEX that are used include: DEGREE-ADV EVAL-ADV GRADABLE LOC&DIR-ADV MANNER-ADV META-ADV TEMPORAL-ADV TIMETAG. Please look at the COMLEX Syntax manual [6] for detailed definitions. Table 5, repeated from the NOMBANK manual shows approximately how these classes can be interpreted for purposes of NOMBANK annotation.

Most ADJADV dictionary entries include the following features :FEATURES :ORTH :SEMI-AUTOMATIC and :COMMENT features. In addition to features for identifying features involving their modification properties, these adjectives: (i) inherit their TIMETAG and GRADABLE features from their COMLEX syntax adjective entries (see the COMLEX Syntax manual for details); and (ii) are marked :SEMI-AUTOMATIC just like NOMBANK entries.

We will now discuss subtypes of ADJADV dictionary entries

More ADJADV Classes			
Class	Probable Tag/ Role Descriptor	Description	Example
FOCUSADJ	ARGM-ADV	Focus adjectives	<i>only</i>
LOCADJ	ARGM-LOC	Locative Adjective	<i>rural</i>
NATIONALITY	ARGM-LOC	Nationality Adjective	<i>Uruguayan</i>
ORGADJ	ARG0, TOPIC, RECIPIENT, BENEFICIARY	organization adjective	<i>democratic</i>
STYLEADJ	ARGM-MNR	style-of adjectives	<i>orwellian</i>
TOPICADJ	ARGM-MNR, TOPIC	ARGM-MNR, TOPIC	<i>financial</i>

Table 6: Mapping ADJADV classes to NomBank ARGM

6.1 ADJADV and ADJADVLIKE

These are by far the most numerous entries in ADJADV. ADJADV entries map between adjectives and adverbs that are both morphologically and semantically related. ADJADVLIKE entries map between adjectives and semantically related adverbs that are unrelated morphologically. The feature :ADV identifies the related adverb. Usually, this adverb is found in COMLEX Syntax with the same (or similar) features.

6.2 FOCUSADJ, LOCADJ, NATIONALITY, ORGADJ, STYLEADJ and TOPICADJ

These are adjectives which are not mapped to any particular adverb, but still seem to serve some adverbial function as indicated in the table 6. These can correspond to certain ARGMs in NomBank or sometimes the numbered arguments.

Note that some of it is unclear whether some of the words with these entries are nouns or adjectives. As a result, they are sometimes marked :PLURAL *NONE* indicating that if they are nouns, they only occur in the singular. Also, if they are given :NOMTYPE values, it is typically the same as their type (STYLEADJ, etc.)

6.3 ADJVERB and ADJVERBLIKE

We have also classified a small number (19 as of this writing) of adjectives that are related to verbs and included them in the ADJADV dictionary. For the purposes of this dictionary, the crucial property is that these nouns can fill the ARGM-PNC role. Examples include: the adjective *acquisitive* which is morphologically related

(NOMADV	:ORTH	“addition”
	:ADV	“additionally”
	:FEATURES	((META-ADV :CONJ T))
	:STRING	“in addition”
	:FEATURES	((GRADABLE) (MANNER-ADV))
	:SEMI-AUTOMATIC	T)

Figure 2: NOMADV entry for *addition*

to the verb *acquire* and therefore has an ADJVERB entry; and *mercenary* which has an ADJVERBLIKE entry based on the verb *hire*. An obvious area for future work would be to investigate adjectives related to verbs on a larger scale both in terms of identifying more instances and investigating their role as predicates in a manner similar to NOMLEX.

7 NOMADV

There are two types of NOMADV entries: NOMADV entries relate nouns and adverbs that are both morphologically and semantically related, whereas NOMADV-LIKE relates nouns and adverbs that related only semantically. The only features these entries permit are: :ORTH (the orthography of the lemma), :FEATURES, :ADV, :SEMI-AUTOMATIC and :STRING. The entries include the same values of :FEATURES as the ADJADV entries and their values can be interpreted similarly in some cases. NOMADV, like ADJADV entries, indicate the corresponding adverb with :ADV feature. :SEMI-AUTOMATIC also has the same interpretation. The feature :STRING is used to indicate idiomatic or semi-idiomatic expressions for which this entry may apply.

As discussed in the NOMBANK manual, these entries are similar to NOMBANK entries in that sometimes correspond to adverbial rolesets for the nouns. For example, the (META-ADV :CONJ T) feature suggests that the noun can probably function as part of a discourse connective, e.g., the entry for *addition* is provided in figure 2. The (META-ADV :CONJ T) feature in combination with the :STRING value “*in addition*” should indicate its discourse adverbial usage.

These nouns also can appear as modifiers of their own arguments like NOMBANK CRISSCROSS nouns. The figure 3 examples are from the NOMBANK manual.

1. *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
2. *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
3. *the bitterness of the battle*
REL = bitterness, ARG2-REF = bitterness, ARG1 = of the battle
REL = battle, ARGM-MNR = bitterness

Figure 3: Crisscrossing ARGM Attribute Nouns

8 COMNOM

COMNOM is an expansion of COMLEX Syntax based on NOMLEX-PLUS with a few additional features thrown in. This section will focus only on these additional features and will not try to recap the entire COMLEX Syntax manual. There are two main changes that were made based on NOMBANK entries. Prepositions were collected from NOMLEX-PLUS subcategorization (:VERB-SUBC, :OBJECT, :SUBJECT, etc.) found in all types of NOMLEX-PLUS entries. It was assumed that nouns took a prepositional phrase (PP) complement with these prepositions as the value of :PVAL. This is significant because COMLEX Syntax did not include PP complements. In addition, for each value of :VERB-SUBC, various sentential complements were deduced and added to the existing COMLEX entry. As noted above (section 3), there are different versions of NOMLEX-PLUS based on which sections of the Penn Treebank II Wall Street Journal corpus were included in its creation. Consequently, there are also three versions of COMNOM: COMNOM, COMNOM-tr and COMNOM-clean.

8.1 More Additions

Occasionally, our automatic update procedure would try to augment a noun entry that was not there. We simply added a new noun entry along with a comment line :COMMENT “automatically generated - not marked noun in COMLEX”. There are 689 such nouns.

8.2 Minor Additions to other Parts of Speech

We added alternation features SUBJ-OBJ-ALT to verbs whose subjects and objects alternate as per the dictionary based on the index of [1] (available by ftp ???)

We added the INTRANS-RECIP :SUBC class to selected adjective entries, e.g., *compatible*.

We made various corrections to the dictionary when they came up.

8.3 Distribution of COMNOM

The Linguistic Data Consortium (LDC) licenses COMLEX Syntax. Since COMNOM includes all of COMLEX, COMNOM must also be distributed through the LDC.

9 Nombank-Dict

Nombank-Dict is a lisp-like version of the frames files that have been automatically enhanced to include generalized descriptors (GENDESCR).⁹ These entries are equivalent to their xml counterparts, except that a heuristic was applied to minimize the different descriptor labels being used. For example, ACTOR becomes AGENT, as does most ARG0 labels that end in “ER” (EXPERIENCER is an exception). These heuristics involve the interaction of: suffixes in :DESCR, a table of common near-synonymous roles, the verbnet role provided in corresponding PROPBANK roleset, a procedure for choosing a single word from a multiworded name, the numbered role of the argument, among other factors.

The sample Nombank-Dict entry in figure 4 corresponds to the frame file content provided as figure 5. Each part of the xml has a lisp-style counterpart. For the most part, the keywords in the Nombank-Dict entry correspond to features in the frame file, e.g., the keyword :DESCR under the type ROLE corresponds to the feature “descr=” in the xml tag of type *role*. In this simple case the :GENDESCR feature is the same as the :DESCR feature. In general, this identity does not hold. Table 7 provides some sample correspondences. The :GENDESCR values are intended to provide a level of generalization for nombank arguments, e.g., it is easier to find statistical information about AGENTS than about ABANDONERS. There are about a third as many different :GENDESCR labels than :DESCR labels (613 versus 1880). Furthermore, the most frequent 91 :GENDESCR labels cover 90% of all label instances in the dictionary. In contrast, the 411 most frequent :DESCR

⁹There is another minor difference between these dictionary entries and the frame files. The possible function tags are listed (:F) inside the descriptions of the roles. OPT stands for optional and OTHER means some function tag that is not part of the normal adjunct list.

:DESCR	:GENDESCR	Sample Words
SECONDARY-THEME	THEME2	%, 1/10th, amount
ABANDONER	AGENT	abandonment
ADVISOR	AGENT	counseling
CLIMBER ACTION ALLOWED	ACTION	authorization, password
BENEFACTOR	BENEFICIARY	aid
BECAUSE OF WHAT	CAUSE	reward
AUDIENCE	RECIPIENT	broadcast, airing
BUYER	RECIPIENT,	asking, charge
THING SHARED	SHARED	common, sharing
SHARED FEELINGS OR IDEAS	SHARED	solidarity

Table 7: Converting :DESCR to :GENDESCR

labels cover 90% of the label instance in the dictionary. Thus a role labeling system that attempts to use role descriptors to predict argument numbers would may be better off using the :GENDESCR labels than the :DESCR labels, in spite of any possible error introduced by the heuristics.

References

- [1] B. Levin. *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press, Chicago, 1993.
- [2] C. Macleod, R. Grishman, and A. Meyers. COMLEX Syntax. *Computers and the Humanities*, 31:459–481, 1998.
- [3] C. Macleod, R. Grishman, A. Meyers, L. Barrett, and R. Reeves. Nomlex: A lexicon of nominalizations. In *Proceedings of Euralex98*, 1998.
- [4] 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.
- [5] R. Reeves, C. Macleod, and A. Meyers. *Manual of NOMLEX: The Regularized Version*. Proteus Project, New York University, 1999. <http://nlp.cs.nyu.edu/nomlex/nomlex-alpha-reg-manual.ps>.
- [6] S. R. Wolff, C. Macleod, and A. Meyers. *Complex Word Classes Manual*. Proteus Project, New York University, 1998.

```

(PBNOUN  \\  

:ORTH  ''ability'' \\  

:ROLE-SETS \\  

  ((ROLE-SET1  \\  

    :ID  ''ability.01''  \\  

    :NAME ''ABILITY/nomadj-able''  \\  

    :ROLES \\  

      ((ROLE :DESCR ''AGENT''  \\  

        :N  ''0''  \\  

        :F (OPT OTHER) \\  

        :GENDESCR ''AGENT'') \\  

      (ROLE :DESCR ''ACTION''  \\  

        :N  ''1'' :F (PRD) \\  

        :GENDESCR ''ACTION'')) \\  

    :EXAMPLES \\  

      ((EXAMPLE  \\  

        :NAME ''AUTOGEN1''  \\  

        :TEXT (''the government 's ability to pay its bills'') \\  

        :ARGS \\  

          ((ARG :N  ''0''  \\  

            :STRINGS (''the government 's'')) \\  

          (REL :STRINGS (''ability'')) \\  

          (ARG :N  ''1''  \\  

            :F  ''PRD''  \\  

            :STRINGS (''to pay its bills'')))) \\  

      (EXAMPLE :NAME ''AUTOGEN2''  \\  

        :TEXT (''the ability to read block printing'')  \\  

        :ARGS \\  

          ((REL :STRINGS (''ability'')) \\  

          (ARG :N  ''1''  \\  

            :F  ''PRD''  \\  

            :STRINGS (''to read block printing'')))))) \\  


```

Figure 4: Sample Nombank-Dict entry

```

<frameset>
<predicate lemma="ability">
<roleset id="ability.01" name="ability/nomadj-able">
<roles>

<role descr="agent" n="0"></role>

<role descr="action" n="1"></role>

</roles>

<example name="autogen1">
<text>
the government 's ability to pay its bills
</text>
<arg n="0">the government 's</arg>
<rel>ability</rel>
<arg f="PRD" n="1">to pay its bills</arg>
</example>

<example name="autogen2">
<text>
the ability to read block printing
</text>
<rel>ability</rel>
<arg f="PRD" n="1">to read block printing</arg>
</example>

</roleset>
</predicate>
</frameset>

```

Figure 5: XML frame file for *ability*