GLML: A Generative Lexicon Markup Language

James Pustejovsky, Anna Rumshisky,
Jessica L. Moszkowicz, Olga Batiukova

June 22, 2008

Abstract

This document describes GLML, a mark-up language based on Generative Lexicon Theory. While most annotation systems capture surface relationships, GLML captures derivational relationships. We provide a brief overview of GL before moving on to our proposed methodology for annotating with GLML. Annotation is task-driven to ease the load for the annotator since GL annotation could be very complex. There are four main tasks:

1. Compositional mechanisms of argument selection
2. Qualia in argument selection
3. Qualia in modification constructions
4. Type selection in modification of dot objects

We explain what each task includes and provide screenshots of the annotation interface. We also include the XML format for GLML including the BNFs for the GLML tags and complete examples of annotated sentences.

Contents

1 Introduction

In this document, we discuss how compositional operations in language can be annotated.[1] Most annotation schemes over propositional and predicative content focus on the identification of the predicate type, the argument extent, and the semantic role (or label) assigned to that argument by the predicate (see Palmer et al., 2005; Ruppenhofer et al., 2006). The focus here will be on annotating the compositional derivation rather than the surface types of the entities involved in function application. So, for example, the distinction in the types appearing in the subject position in (1) below is captured by entity typing, but not by any sense tagging from FrameNet or PropBank.

(1) a. Mary called yesterday.

b. The Boston office called yesterday.

While this has been treated as *type coercion* or *metonymy* in the literature (cf. Hobbs, Pustejovsky, others), the relevant point here is that PropBank and FrameNet should treat the sentences on par with one another. Yet, how is this possible, if the entity typing given to the subject in (1a) is HUMAN and that given for (1b) is ORGANIZATION or BUILDING?[2]

Similarly, in (2) below, sense annotation for the verb *enjoy* should arguably assign similar values to both (2a) and (2b).

(2) a. Mary enjoyed drinking her beer .

b. Mary enjoyed her beer.

[1]The idea for annotating a corpus according to GL arose during a discussion at GL 2007, between one of the authors (J. Pustejovsky) and Nicoletta Calzolari and Pierrette Bouillon.

[2]The SemEval Metonymy task, Markert and Nissim (2007) describe the best consensus in the community on how to handle metonymic relations over a large data set. This task has two basic types with metonymic variants:

i. **Categories for Locations**: literal, place-for-people, place-for-event, place-for-product;

ii. **Categories for Organizations**: literal, organization-for-members, organization-for-event, organization-for-product, organization-for-facility .

Most of these relation types can be subsumed under the current specification. Interestingly, the relations introduced in this document are not covered by the scope of that task definition.

The consequence of this, however, is that the mapping to a syntactic realization for a given sense is made more complex, and is in fact, perplexing for a clustering or learning algorithm operating over subcategorization types for the verb.

Pustejovsky (2007) and Asher and Pustejovsky (2006) distinguish the following modes of composition in natural language:

(3) a. PURE SELECTION (Type Matching): the type a function requires is directly satisfied by the argument;
b. ACCOMMODATION: the type a function requires is inherited by the argument;
c. TYPE COERCION: the type a function requires is imposed on the argument type. This is accomplished by either:
i. *Exploitation*: taking a part of the argument's type to satisfy the function;
ii. *Introduction*: wrapping the argument with the type required by the function.

Each of these will be identified as a unique relation between the predicate and a given argument. In this annotation effort, we restrict the possible relations between the predicate and a given argument to *selection* and *coercion*. A more fine-grained typology of relations may be applied at a later point. Furthermore, qualia structure values are identified in both argument selection and modification contexts.

The rest of this document proceeds as follows: Section 2 provides some background information on Generative Lexicon Theory. In Section 3, we describe our general methodology and architecture for GL annotation. Section 4 gives an overview of each of the annotation tasks. A more complete description that also includes the resulting GLML markup for each task is provided in Section 5. The actual specification of GLML can be found in Section 6. Finally, in Section 7, we mention some possible extensions to GLML.

2 Background of GL

Generative Lexicon introduces a knowledge representation framework which offers a rich and expressive vocabulary for lexical information. The motivations for this are twofold. Overall, GL is concerned with explaining the creative use of language; we consider the lexicon to be the key repository holding much of the information underlying this phenomenon. More specifically, however, it is the notion of a constantly evolving lexicon that GL attempts to emulate; this is in contrast to views of static lexicon design, where the set of contexts licensing the use of words is determined in advance, and there are no formal mechanisms offered for expanding this set.

One of the most difficult problems facing theoretical and computational semantics is defining the representational interface between linguistic and non-linguistic knowledge. GL was initially developed as a theoretical framework for encoding selectional knowledge in natural language. This in turn required making some changes in the formal rules of representation and composition. Perhaps the most controversial aspect of GL has been the manner in which lexically encoded knowledge is exploited to construct interpretations for linguistic utterances. Following standard assumptions in GL, the computational resources available to a lexical item consist of the following four levels:

(4) a. LEXICAL TYPING STRUCTURE: giving an explicit type for a word positioned within a type system for the language;
b. ARGUMENT STRUCTURE: specifying the number and nature of the arguments to a predicate;
c. EVENT STRUCTURE: defining the event type of the expression and any subeventual structure it may have;
d. QUALIA STRUCTURE: a structural differentiation of the predicative force for a lexical item.

The qualia structure, inspired by Moravcsik (1975)'s interpretation of the *aitia* of Aristotle, is defined as the modes of explanation associated with a word or phrase in the language. These are defined as follows in Pustejovsky (1991):

(5) a. FORMAL: the basic category of which distinguishes the meaning of a word within a larger domain;
b. CONSTITUTIVE: the relation between an object and its constituent parts;
c. TELIC: the purpose or function of the object, if there is one;
d. AGENTIVE: the factors involved in the object's origins or "coming into being".

Pustejovsky (2001) separates the domain of individuals into three distinct type levels:

(6) a. NATURAL TYPES: Natural kind concepts consisting of reference only to Formal and Constitutive qualia roles;

b. ARTIFACTUAL TYPES: Concepts making reference to purpose or function.
c. COMPLEX TYPES: Concepts making reference to an inherent relation between types.

In this annotation effort, we will type the arguments using a modified shallow ontology derived from the Brandeis Shallow Ontology (Pustejovsky et al., 2004).

3 General Methodology and Architecture

In this section, we describe the set of tasks for annotating compositional mechanisms within the GL framework. The current GL markup will include the following tasks:

(7) a. Compositional Mechanisms of Argument Selection: Verb-based Annotation
b. Qualia in Argument Selection: Noun-based Annotation
c. Qualia in Modification Constructions
d. Type Selection in Modification of Dot Objects

These will be discussed in Section 4.

3.1 System Architecture

Figure 1 below shows the general architecture of a GLML annotation. The first two phases are designed to prepare the data for annotation. The prepared data is then presented to the annotator in the annotation environment. The result of the annotation is stored in a database. At that point, an XML generator is used to output the annotation. In addition, a second database may be used to store the GL logical form version of the annotation, although the logical form is still under development.

The first step in the process is to construct the data set for annotation. This involves the following:

(8) a. Select targets for each task
b. Identify the senses for each target
c. Associate type templates for each sense

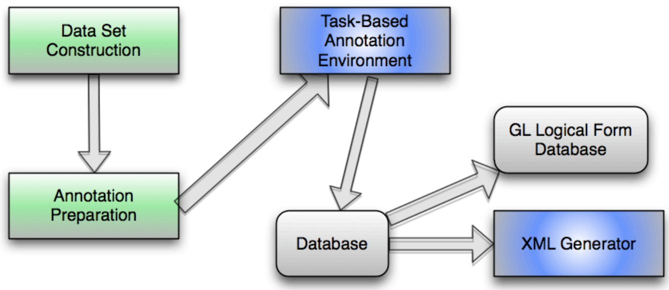


Figure 1: GLML Architecture

These steps will differ slightly for each of the major GLML annotation tasks. For example, Task 1 focuses on predicates. To prepare for this task, we must choose which predicates we want to annotate. We then have to identify the senses for each of these predicates and associate type templates with each sense. Since the goal of Task 1 is to distinguish between selection and coercion, we may want to choose predicates that sometimes select and other times coerce.

The Data Set Construction step tells us what will be annotated. The next step performs some basic preprocessing that is needed for the annotation environment including:

(9) a. Highlight the targets for each task
    b. Perform argument and adjunct parsing
    c. Identify argument and adjunct heads

Once the data has been prepared, task-based annotation can be performed. Task-based annotation is the best choice for GL annotation due to the complexity of the task. The annotation environment is designed so that the annotator can focus on one facet of the annotation at a time, rather than trying to do everything at once. The environment used for each task is described later in this document.

Finally, we envision having multiple ways to represent the completed annotation. In this document, we focus on the XML format of GLML, but the annotation can also be stored in a database or converted to GL logical form.

3.2 The Type System for Annotation

The type system we have chosen for annotation is purposefully shallow, but we also aimed to include types that would ease the complexity of the annotation task. The type system is not structured in a hierarchy, but rather it is presented as a set of types. For example, we include both HUMAN and ANIMATE in the type system along with PHYSICAL OBJECT. While HUMAN is a subtype of both ANIMATE and PHYSICAL OBJECT, the annotator does not need to be concerned with this. This allows the annotator to simply choose the HUMAN type when necessary rather than having to deal with type inheritance.

In addition to the top level of types used in the annotation, we also include a second level of more specific types for some of the top types. For the first round of annotation, we include more specific types only for ABSTRACT ENTITY, but it is conceivable that other top types could also have a second level. These more specific types are present for the first round of annotation so that we can refine what types are most appropriate to include in the annotation.

The following table shows the system of types we use for annotation, including the set of types for the more specific ABSTRACT ENTITY types:

Top Types	**Abstract Entity Subtypes**
abstract entity	attitude
human	emotion
animate	property
organization	obligation
physical object	rule
artifact	
event	
proposition	
information	
sensation	
location	
time period	

Table 1: Type System for Annotation

The Brandeis Shallow Ontology (BSO) is a shallow hierarchy of types selected for their prevalence in manually identified selection context patterns. As used by the CPA technique (Pustejovsky et al., 2004, see also

Rumshisky et al., 2006), there are just 65 types, in terms of which patterns for the first one hundred verbs have been analyzed. New types are added occasionally, but only when all possibilities of using existing types prove inadequate. Once the set of manually extracted patterns is sufficient, the type system will be re-populated and become pattern-driven.

4 Task Description

4.1 Compositional Mechanisms of Argument Selection: Verb-based Annotation

This annotation task involves choosing which selectional mechanism is used by the predicate over a particular argument. Unlike PropBank and FrameNet annotation, where the argument is identified with an argument identifier and semantic role label associated with the verb, here we are interested in the compositional history of the way a function acts over its argument. The possible relations between the predicate and a given argument will, for now, be restricted to *selection* and *coercion*. In *selection*, the argument NP satisfies the typing requirements of the predicate, as in (10).

(10) a.The spokesman denied the statement (PROPOSITION).
b. The child threw the ball (PHYSICAL OBJECT).
c. The audience didn't believe the rumor (PROPOSITION).

Coercion encompasses all cases when any of the type-shifting operations mentioned above (e.g. exploitation or introduction) must be performed on the complement NP in order to satisfy selectional requirements of the predicate, as in (11). Note that coercion operations may apply to any argument position in a sentence, including the subject, as seen in (11) and (11). Coercion can also be seen as an object of a proposition as in (11).

(11) a. The president denied the attack (EVENT → PROPOSITION).
b. The White House (LOCATION → HUMAN) denied this statement.
c. This book (PHYS • INFO → HUMAN) explains the theory of relativity.
d. The Boston office called with an update (EVENT → INFO).

An initial set of verbs has been selected for annotation, with a set number of sentences chosen randomly for each verb. For each sentence, the compositional relationship of the verb with every argument and adjunct will be annotated. The target types for each argument are provided in a *type template* shown in Figure 2 that is associated with the sense of the verb in the given sentence. The first subtask, therefore, is to disambiguate the verb senses.

Figure 2: Type Template Schematic

During the annotation, we will also ask the annotator to specify a source type for the adjunct in the case of coercion. This will allow us to collect data on what types can be coerced and into what target types. As such, the annotator will be responsible for marking an example such as (11) as a coercion, but he or she should also annotate that *attack* is usually of type `event`. The details of the steps involved are provided in Section 5.1.

4.2 Qualia in Argument Selection: Noun-based Annotation

This annotation task involves choosing which quale associated with the noun is acted on by the verb. The identification of the semantic nature of the link between an argument and the verb selecting it is needed because the mere indication of the grammatical relation holding between both is not enough to determine the interpretation. Evidence of this can be seen in other domains as well. For example, semantic labeling is necessary to differentiate different types of intransitive constructions, namely, unaccusative and unergative. Consider the contrast between *John fell* and *John ran*. These are distinguished by the the fact that in the former sentence, John is assigned the semantic role of Patient or Theme while in the latter, John is the agent. This difference also accounts for their different syntactic behavior in certain languages.

An initial set of nouns has been selected for annotation, with a set number of sentences chosen randomly for each noun. For each sentence, the argument position occupied by the target noun is annotated with respect to the quale acted on by the predicate. As in the previous task, the first step will be to disambiguate the noun senses.

The following qualia relations will be identified: *Formal*, *Constitutive*, *Telic*, *Agentive*, and *Inverse Constitutive*. The first four relations correspond to the qualia as defined in classic GL (see (5) in section 2). Examples are given below in (12) - (15).

The qualia of the noun being activated by the verb is shown in parentheses.

(12) a. Mary drank a <u>cup</u> of water. (TELIC)

b. John bought a <u>cup</u> of coffee. (FORMAL)

c. The child enjoyed her <u>cup</u> of juice. (TELIC)

(13) a. John accidentally broke his <u>pen</u>. (TELIC)

b. I need to fill my <u>pen</u>. (CONST)

c. This <u>pen</u> leaks. (CONST)

d. John brought his <u>pen</u> to the exam. (FORMAL)

(14) a. Please set the <u>table</u> for dinner. (TELIC)

b. John banged the <u>table</u> with his shoe. (FORMAL)

c. Mary left the <u>table</u> abruptly. (TELIC)

(15) a. Can you shine the <u>lamp</u> over here? (TELIC)

b. Mary hung the <u>lamp</u> in the kitchen. (FORMAL)

c. John assembled the <u>lamp</u>. (AGENTIVE)

The *Inverse Constitutive* quale is used when the noun is part of another entity of the same type. This type is used in the adjective-noun and noun-noun compounds described in the following subsection, but it does not seem to appear in verb-argument contexts.

4.3 Qualia Selection in Modification Constructions

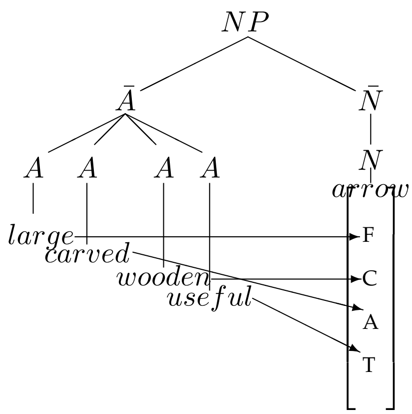
In this task, the annotation identifies the manner in which the modifying expression semantically relates to the target element, typically a noun. We follow Bouillon (1997) in distinguishing different modification relations of an adjective over a head as cases of distinct qualia selection.

For GLML annotation, the relevant semantic relations are defined in terms of the qualia structure. We examine two kinds of constructions in this task: adjectival modification of nouns and nominal compounds.

4.3.1 Adjectival Modification of Nouns

This task involves annotating how particular noun qualia values are bound by the adjectives. Following Pustejovsky (2000), we assume that the properties grammatically realized as adjectives "bind into the qualia structure of nouns, to select a narrow facet of the noun's meaning." The different types of binding of the adjectival modification can be better understood if we examine the modification structure of *large carved wooden useful arrow* in (16):

(16)



Large refers to the arrow as a physical object, its FORMAL type, so that adjective is associated with that quale. Similarly, *carved* is associated with the creation of the arrow (AGENTIVE), *wooden* is associated with a material part of the arrow (CONSTITUTIVE), and *useful* is associated with how the arrow is used (TELIC).

Many adjectives appear specialized with respect to the qualia they bind and, in these cases, they identify the concrete value of the relevant quale. See, for example, (17) - (20).

(17) CONST
a. wooden house
b. wrinkled face
c. mountainous region

(18) FORMAL
heavy, red, large, sweet, raw, rough, hard, simple, responsible, happy, short, narrow, poor, bitter, new

(19) TELIC
useful, effective, good (knife, table, teacher)

(20) AGENTIVE
a. carved figure
b. hand-made shoes
c. synthetic material
d. natural light

Each adjective-noun pairing will be presented in a full sentential context as shown in Section 5.

4.3.2 Nominal Compounds

This task explores the semantic relationship between elements in nominal compounds. The general relations presented in Levi (1978) are a useful guide for beginning a classification of compound types, but the relations between compound elements quickly proves to be too coarse-grained. Warren's comprehensive work (Warren, 1978) is a valuable resource for differentiating relation types between compound elements.

The class distinction in compound types in language can be broken down into three forms (cf. Spencer, 1991):

(21) a. ENDOCENTRIC COMPOUNDS: One element in the construction functions as the head.
b. EXOCENTRIC COMPOUNDS (BAHUVRIHI): loudmouth
c. DVANDVA COMPOUNDS: a simple conjunction of two elements, without a dependency holding between them;

Following Bisetto and Scalise (2005), however, it is possible to distinguish three slightly differently constructed classes of compounds, each exhibiting endocentric and exocentric behavior:

(22) a. SUBORDINATING: the head acts functionally over N_1, incorporating it as an argument.
b. ATTRIBUTIVE: a general modification relation.
c. COORDINATE: the dvandva construction mentioned above.

We will focus on the two classes of *subordinating* and *attributive* compounds. Within each of these, we will distinguish between *synthetic* and *non-synthetic* compounds. The latter are deverbal nouns, and when acting functionally (subordinating), take the sister noun as an argument, as in (23).

(23) a. bus driver
b. window cleaner

The non-synthetic counterparts of (23) are shown below in (24), where the head is not deverbal in any obvious way.

(24) a. pastry chef
b. bread knife

While Bisetto and Scalise's distinction is a useful one, it does little to explain how non-relational sortal nouns such as *chef* and *knife* act functionally over the accompanying noun in the compound, as above.

This construction and the semantic phenomenon generally has been examined within GL by Johnston and Busa (1999). We will assume much of that analysis in our definition of the task described here. Our basic assumption regarding the nature of the semantic link between both parts of compounds is that it is generally similar to the one present in adjectival modification. The only difference is that in nominal compounds, for instance, the qualia of a head noun are activated or exploited by a different kind of modifier, a noun.

Following Johnston and Busa (1999), consider the following [N_1 N_2] constructions in English and the corresponding [N_2 P N_1] constructions in Italian.

(25) a. coltello da pane
"bread knife"
b. bicchiere da vino
"wine glass"
c. foro di pallottola
"bullet hole"
d. succo di limone
"lemon juice"
e. porta a vetri
"glass door"

In compounds (25a,b), the relation between N_1 and N_2 can be identified as the Telic role for the heads, *knife* and *glass*, while in (25c,d), the relation can be identified with the Agentive of the respective heads, *hole* and *juice*. In (25e), on the other hand, *glass* is the Constitutive of the head *door*. Interestingly, Johnston and Busa (1999) illustrate how in Italian, the choice between *da* and *di* in compounds is not in free variation, but rather conditioned by the semantic relation between the noun. Specifically, they argue that it is the particular quale binding the two nouns that determines the choice. They correlate the use of *da* with the Telic quale while *di* can be associated with either Agentive or Constitutive.[3]

Using the strategy of qualia selection outlined above by Johnston and Busa (1999), we can identify a broad range of semantic relations in noun compound constructions as qualia-based. As illustration, consider the compounds in (26)-(27).

(26) $[N_1\ N_2]$: N_1 is the TELIC of N_2:
  a. fishing rod
  b. magnifying glass
  c. swimming pool
  d. shopping bag
  e. drinking water

(27) $[N_1\ N_2]$: N_1 is the CONST of N_2:
  a. paper napkins
  b. metal cup
  c. gold filling

Synthetic subordinating compounds may also be characterized as qualia relations, even though they are acting functionally. For example, the examples in (28) are both subordinating and AGENTIVE-selecting compounds:

(28) $[N_1\ N_2]$: N_1 is the AGENTIVE of N_2:
  a. food infection
  b. heat shock

Interestingly, there are corresponding non-synthetic compounds, which also act functionally and are AGENTIVE-selecting:

(29) $[N_1\ N_2]$: N_1 is the AGENTIVE of N_2:
  a. university fatigue
  b. automobile accident
  c. sun light

[3]Rosario and Hearst (2001) catalogue a range of semantic relations in noun compounds using domain-specific lexical knowledge.

4.4 Type Selection in Modification of Dot Objects

This task involves annotating how particular types within dot objects are exploited in adjectival and nominal modification constructions. *Dot objects* or *complex types* (Pustejovsky, 1995) are defined as the product of a type constructor • ("dot"), which creates dot objects from any two types a and b , creating $a \bullet b$. Complex types are unique because they are made up of seemingly conflicting types such as FOOD and EVENT.

Given a complex type $c = a \bullet b$, there are three possible options:

1. the modifier applies to both a and b
2. the modifier applies to a only
3. the modifier applies to b only

Option 1 would be illustrated by examples such as *good book [+info, +physobj]* and *long test [+info, +event]*. Options 2 and 3 can be illustrated by:

(30) a. lunch (EVENT • FOOD):
delicious lunch (FOOD) vs. long lunch (EVENT)

b. book (INFO • PHYSOBJ):
boring book (INFO) vs. heavy book (PHYSOBJ)

c. rumor (ACTIVITY • PROPOSITION):
false rumor (PROPOSITION) vs. persistent rumor (ACTIVITY)

d. lecture (EVENT • INFO):
morning lecture (EVENT) vs. interesting lecture (INFO)

e. lamb (ANIMAL • FOOD):
roast lamb (FOOD) vs. newborn lamb (ANIMAL)

f. construction (PROCESS • RESULT):
wooden construction (RESULT) vs. road construction (PROCESS)

g. concert (EVENT • INFO):
open-air concert (EVENT) vs. orchestral concert (INFO)

A listing of these dot objects has been provided, first in Pustejovsky (2005) and expanded in Rumshisky et al. (2007), cf. Table 2.

Dot type	Example
ACTION • PROPOSITION	promise, allegation, lie, charge
STATE • PROPOSITION	belief
ATTRIBUTE • VALUE	temperature, weight, height, tension, strength
EVENT • INFO	lecture, play, seminar, exam, quiz, test
EVENT • HUMAN	appointment
EVENT • (INFO • SOUND)	concert, sonata, symphony, song
EVENT • PHYSOBJ	lunch, breakfast, dinner, tea
INFO • PHYSOBJ	article, book, CD, DVD, dictionary, diary, email, essay, letter, novel, paper
ORGANIZATION • (INFO • PHYSOBJ)	newspaper, magazine, journal
ORGANIZATION • LOCATION • HUMANGROUP	university, city
EVENT • LOCATION • HUMANGROUP	class
APERTURE • PHYSOBJ	door, window
PROCESS • RESULT	construction, imitation, portrayal, reference, decoration, display, documentation, drawing, enclosure, entry, instruction, design, invention, simulation, illustration, agreement, approval, recognition, damage, compensation, contribution, discount, donation, acquisition, deduction, endowment, gift, classification, purchase
PRODUCER • PRODUCT	Honda, IBM, BMW
TREE • FRUIT	apple, orange, coffee
TREE • WOOD	oak, elm, pine
Pseudo-dot type	**Example**
ANIMAL • FOOD	anchovy, catfish, chicken, eel, herring, lamb, octopus, rabbit, squid, trout
ANIMAL • ARTIFACT • FIR	mink
CONTAINER • CONTENTS	bottle, bucket, carton, crate, cup, flask, keg, pot, shovel, spoon

Table 2: Some examples of dot objects of different complex types, as well as "pseudo-dots" that exhibit dot-like behavior due to coercion.

5 Task Specification and Annotation Guidelines

In this section, we describe the annotation process by detailing the steps involved in each task and the way they are presented to the annotators. For each task, we describe the annotation interface and give the resulting GLML markup for the annotated examples. The annotation will create two link types: CompLink and QLink. CompLink is a composition link representing the relation between the predicate (viewed as function) and its argument. The QLink expresses a relation between a predicate and its argument that involves a specific qualia value. These will be formally defined in Section 6 below.

For each task, a set of target words is selected. For each target word, a sense inventory is compiled at the Data Set Construction stage (cf. Section 3.1). A set of sentences for each target is then selected randomly from a corpus. Each sentence is automatically parsed. [4] The sentences are then organized according to the grammatical relations involving the target word. The annotators are presented with a set of sentences corresponding to a given grammatical relation for each target word. Except for the task involving dot objects (cf. Section 5.5), the first stage involves sense disambiguation of the target word.

5.1 Compositional Mechanisms of Argument Selection: Verb-based Annotation

This annotation task is divided into two subtasks, presented successively to the annotator:

1. Word sense disambiguation of the target predicate;
2. Identification of the compositional relationship between target predicate and its arguments.

We describe the subtasks below.

5.1.1 Subtask 1: Predicate Sense Disambiguation

This annotation task is set up by selecting a set of coercive verbs. A sense inventory is compiled for each verb as described in Rumshisky and Batiukova

[4]Currently we have experimented with RASP (Briscoe and Carroll, 2002) and off the shelf head-based dependency parsers.

(2008). The senses are differentiated by the types associated with the arguments, i.e., a type template is associated with each sense. For example, one of the senses of the verb *deny* is glossed as "State or maintain that something is untrue". The following type template is associated with that sense:

(31) HUMAN deny PROPOSITION

The type template is built in the way similar to the context patterns as defined in Corpus Pattern Analysis (CPA) (Hanks and Pustejovsky, 2005; Pustejovsky et al., 2004).

In the first subtask, the annotator is presented with a set of sentences containing the target verb and the chosen grammatical relation. Both the verb and the headword of the dependent noun phrase are highlighted. The annotator is asked to select the most fitting sense of the target verb, or to throw out the example (pick the "N/A" option) if no sense can be chosen either due to insufficient context, because the appropriate sense does not appear in the inventory, or simply no disambiguation can be made in good faith. The interface is shown in Figure 3.

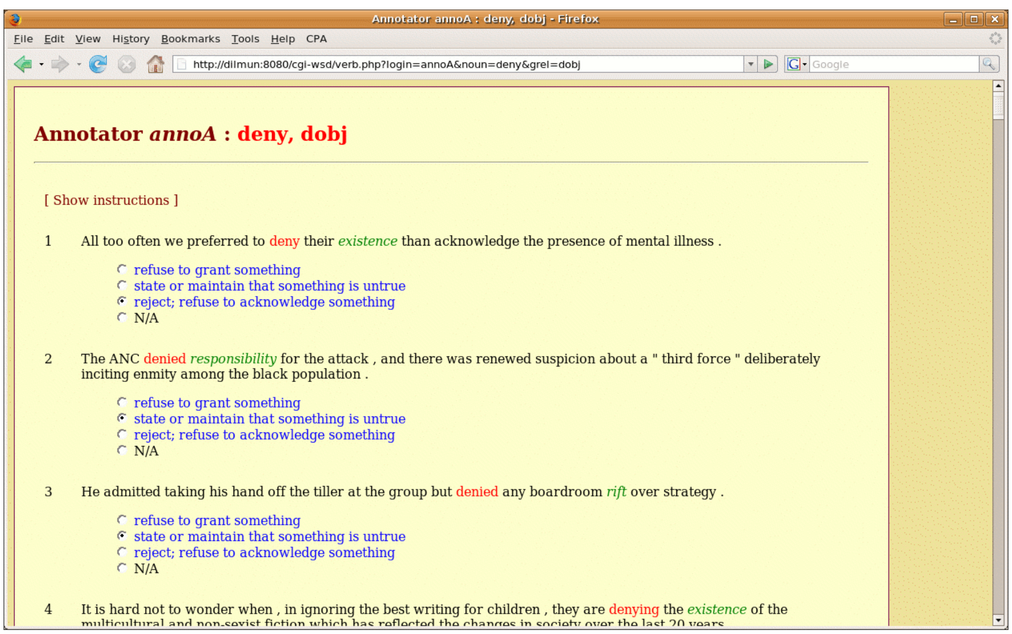


Figure 3: Predicate Sense Disambiguation for *deny*

After this step is completed by the annotator, the appropriate sense is saved into the database, along with the associated type template. The annotator then chooses which grammatical relation to annotate, picking from a list of relations to be annotated next and proceeds to the next step.

5.1.2 Subtask 2: Identifying Compositional Relationship

In this subtask, the annotator is presented with a list of sentences in which the target verb is used in the same sense. The annotator is then asked to determine whether the argument in the specified grammatical relation to the target belongs to the type associated with that sense in the corresponding template. The illustration of this can be seen in Figure 4.

Figure 4: Identifying Compositional Relationship for *deny* (Step 1)

If the argument belongs to the appropriate type, the "yes" box is clicked. This generates a corresponding CompLink with relType="SELECTION". If "no" is selected, a type selection menu pops up below the first question, and the annotator is asked to pick a type from a list of shallow types which is usually associated with the argument. For example, as illustrated in Figure 5, *existence* does not usually denote a PROPOSITION, and the annotator is asked to identify the type it is usually associated with.

The word *existence* happens to be associated with the type ABSTRACT ENTITY which has a number of subtypes. A subtype selection menu pops

Figure 5: Identifying Compositional Relationship for *deny* (Step 2)

up below, and the annotator is asked to choose a subtype, if any (see Figure 6). In case of *existence*, no subtype is specified, and the CompLink with compType="COERCION" is created with the source type ABSTRACT ENTITY and the target type PROPOSITION, signifying that a coercion from ABSTRACT ENTITY to PROPOSITION has taken place in the annotated sentence. The way the other examples for *deny* are annotated is shown in Figure 7.

Figure 8 shows an example of annotation done for the verb *fill* and its indirect object, the NP governed by the preposition *with*. For the examples shown, no type coercion occurs. However, if the source type has several subtypes, the annotator is still asked to specify which subtype is appropriate. In the first round of annotation, the annotators will also be allowed to specify a type not present in the shallow ontology. If the appropriate subtype is not given, the subtype selection "other" may be chosen and specified.

The resulting GLML markup for all examples shown for *fill* and *deny* can be seen in Section 5.1.3 below.

Figure 6: Identifying Compositional Relationship for *deny* (Step 3)

5.1.3 Resulting GLML markup

Examples given in Figure 7 for the direct object of the verb *deny* are saved into the database. The GLML markup generated from the database looks as follows:

1. *All too often we preferred to deny their existence than acknowledge the presence of mental illness.*

```
All too often we preferred to
<SELECTOR sid="s1">deny</SELECTOR>
their
<NOUN nid="n1">existence</NOUN>
than acknowledge the presence of mental illness .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="dobj"
compType="COERCION" sourceType="ABSTRACT_ENTITY" targetType="PROPOSITION"/>
```

2. *Sir Nicholas Lyell, Attorney General, denies a cover-up.*

```
Sir Nicholas Lyell , Attorney General ,
<SELECTOR sid="s1">denies</SELECTOR>
a <NOUN nid="n1">cover-up</NOUN> .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="dobj"
compType="COERCION" sourceType="EVENT" targetType="PROPOSITION"/>
```

Figure 7: Identifying Compositional Relationship for *deny* (Step 4)

3. *Klaus denied accusations by opposition parties that he was copying Poland's "shock therapy".*

```
Klaus
<SELECTOR sid="s1">denied</SELECTOR>
<NOUN nid="n1">accusations</NOUN>
by opposition parties that he was copying Poland 's " shock therapy " .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="dobj"
compType="SELECTION" sourceType="PROPOSITION" targetType="PROPOSITION"/>
```

The GLML markup generated from the database for the examples for the verb *fill* given in Figure 8 looks as follows:

1. *Her mother's voice would fill with outrage.*

```
Her mother 's voice would
<SELECTOR sid="s1">fill</SELECTOR>
with
<NOUN nid="n1">outrage</NOUN> .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="dobj"
compType="SELECTION" sourceType="ABSTRACT_ENTITY"
targetType="ABSTRACT_ENTITY" sourceSubtype="EMOTION"/>
```

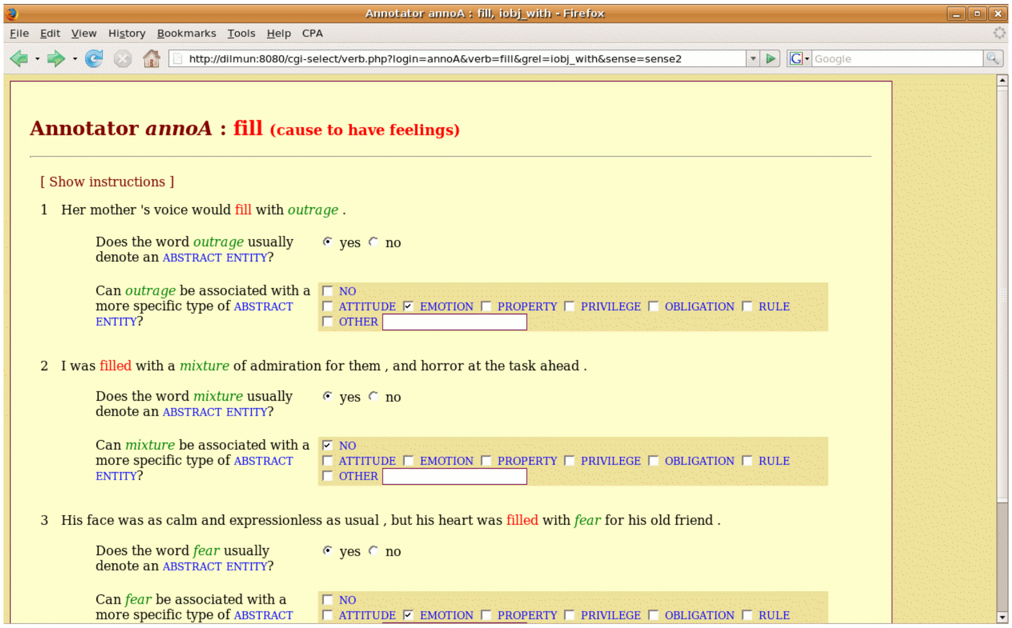


Figure 8: Identifying Compositional Relationship for *fill*

2. *I was filled with a mixture of admiration for them, and horror at the task ahead.*

```
I was
<SELECTOR sid="s1">filled</SELECTOR>
with a <NOUN nid="n1">mixture</NOUN>
of admiration for them , and horror at the task ahead .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="dobj"
compType="SELECTION" sourceType="ABSTRACT_ENTITY"
targetType="ABSTRACT_ENTITY"/>
```

3. *His voice was as calm and as expressionless as usual, but his heart was filled with fear for his old friend.*

```
His voice was as calm and as expressionless as usual, but
his heart was
<SELECTOR sid="s1">filled</SELECTOR>
with
<NOUN nid="n1">fear</NOUN> for his old friend .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="dobj"
compType="SELECTION" sourceType="ABSTRACT_ENTITY"
targetType="ABSTRACT_ENTITY" sourceSubtype="EMOTION"/>
```

5.2 Qualia in Argument Selection: Noun-based Annotation

This annotation task is divided into two subtasks, presented successively to the annotator:

1. Word-sense disambiguation for the target noun;
2. Identification of the qualia relationship between the noun and the governing predicate.

We describe these subtasks below.

5.2.1 Subtask 1: Sense Disambiguation for Nouns

This annotation task is set up by selecting a set of nouns for annotation. A sense inventory is compiled for each noun in a manner similar to the one described above in Section 5.1.

In the first subtask, the annotator is presented with a set of sentences containing the target noun. The annotator is asked to select the most fitting sense of the target noun, or to throw out the example if no choice can be made. The interface is shown in Figure 9. In this example, the noun *cup* is disambiguated between the PHYSICAL OBJECT sense and the EVENT sense.

After this step is completed by the annotator, the appropriate sense is saved into the database for each sentence. The annotator then proceeds to the next step.

5.2.2 Subtask 2: Identifying the Qualia Involved

In this subtask, the annotator is presented with a list of sentences in which the target noun occurs in the same sense. A verb in the chosen grammatical relation to the target noun is highlighted, as well as the target noun itself. The annotator is asked to determine which quale (or qualia) associated with the noun is acted on by the verb. The illustration of this step can be seen in Figures 10 and 11. Once the appropriate quale is chosen, the QLink with the corresponding relType is created. If more than one qualia role is selected, multiple links are created. The annotator is also given an option of declining to annotate the sentence ("unclear from context").

The questions presented to the annotator in order to determine the appropriate qualia roles will change according to the type associated with target noun. Thus, for the PHYSICAL OBJECT-denoting nouns, the question corresponding to the Agentive role involves "making or destroying"

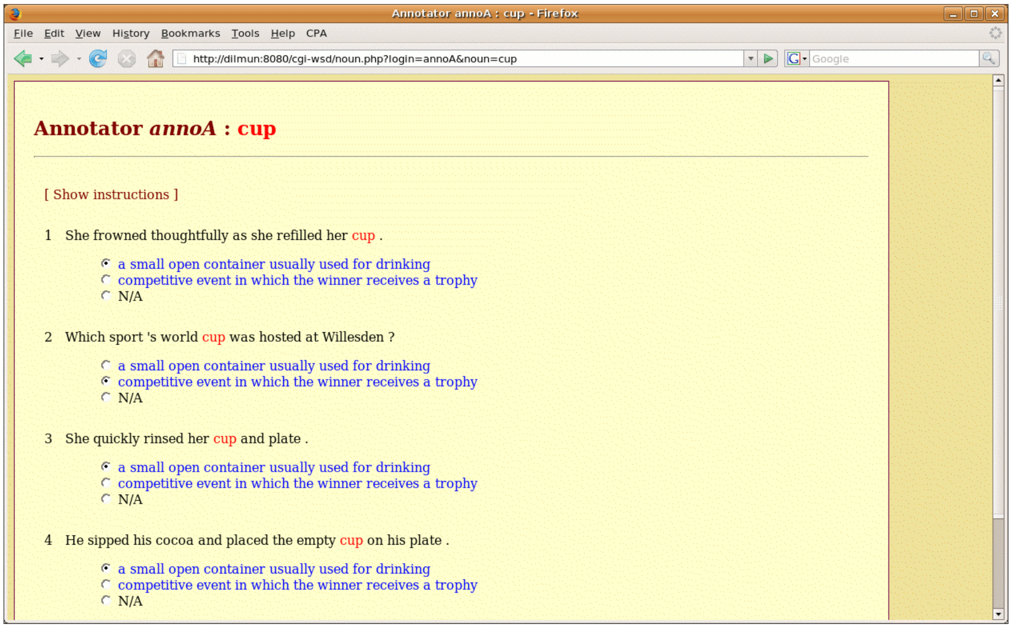


Figure 9: Subtask 1: Sense Disambiguation for Nouns

the object, while for the EVENT-denoting nouns, the same question involves "beginning or ending" the event. Also, the questions aiming to elicit the response about Constitutive role are absent from the question set for EVENT-denoting nouns.

The resulting GLML markup for both senses of *cup* can be seen in Section 5.2.3 below.

5.2.3 Resulting GLML markup

Examples given in Figure 10 for the noun *cup* as a PHYSICAL OBJECT, generate the GLML markup that looks as follows:

1. *She frowned thoughtfully as she refilled her cup.*

```
She frowned thoughtfully as she
<SELECTOR sid="s1">refilled</SELECTOR>
her
<NOUN nid="n1">cup</NOUN> .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="TELIC"/>
<QLink qid="qid2" sID="s1" relatedToNoun="n1" qType="FORMAL"/>
```

2. *I remember sitting alone in the cafeteria, slowly drinking my cup of coffee.*

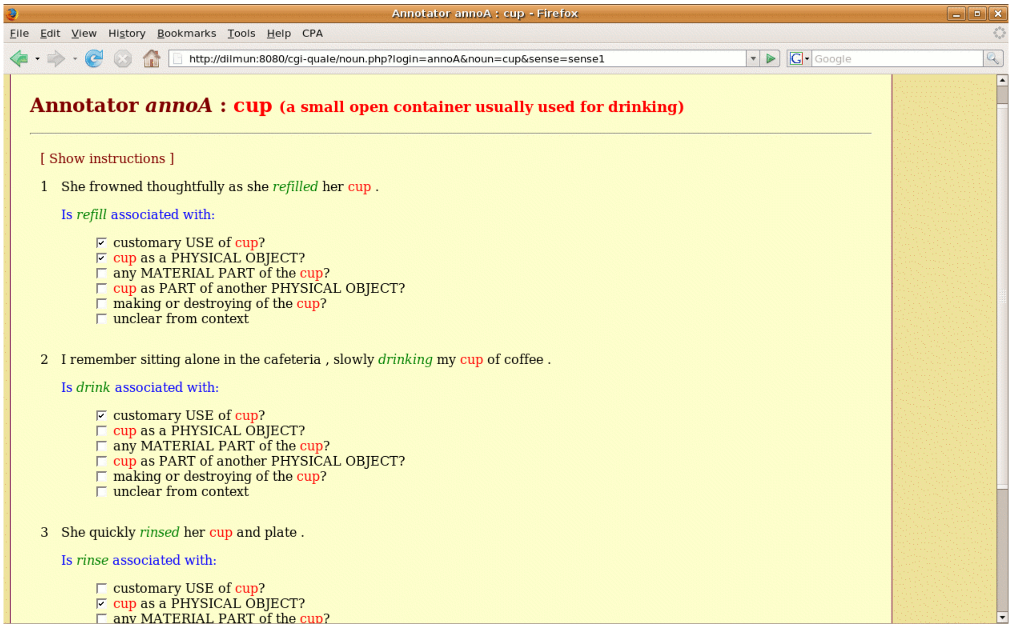


Figure 10: Subtask 1: Qualia Relations for *cup* as a PHYSICAL OBJECT

```
I remember sitting alone in the cafeteria , slowly
<SELECTOR sid="s1">drinking</SELECTOR>
my
<NOUN nid="n1">cup</NOUN>
of coffee .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="TELIC"/>
```

3. *She quickly rinsed her cup and plate.*

```
She quickly
<SELECTOR sid="s1">rinsed</SELECTOR>
her
<NOUN nid="n1">cup</NOUN>
and plate .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="FORMAL"/>
```

When examples given in Figure 11 for the noun *cup* as an EVENT are saved to the database, the GLML markup generated from the database looks as follows:

1. *We will get the horses used to the arena before the cup begins in earnest.*

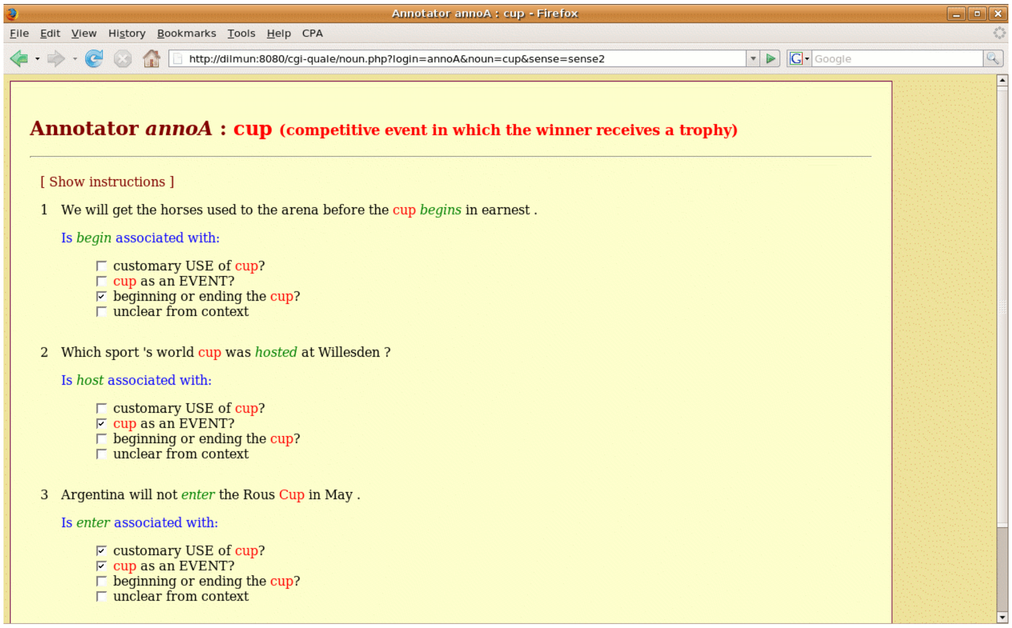


Figure 11: Subtask 1: Qualia Relations for *cup* as an EVENT

```
We will get the horses used to the arena before the
<NOUN nid="n1">cup</NOUN>
<SELECTOR sid="s1">begins</SELECTOR>
 in earnest .
<QLink sID="s1" relatedToNoun="n1" qType="AGENTIVE"/>
```

2. *Which sport's world cup was hosted at Willesden?*

```
Which sport 's world
<NOUN nid="n1">cup</NOUN>
was
<SELECTOR sid="s1">hosted</SELECTOR>
at Willesden ?
<QLink sID="s1" relatedToNoun="n1" qType="FORMAL"/>
```

3. *Argentina will not enter the Rous Cup in May.*

```
Argentina will not
<SELECTOR sid="s1">enter</SELECTOR>
the Rous
<NOUN nid="n1">Cup</NOUN>
in May .
<QLink sID="s1" relatedToNoun="n1" qType="TELIC"/>
<QLink sID="s1" relatedToNoun="n1" qType="FORMAL"/>
```

5.3 Qualia Selection in Modification Constructions: Adjectival Modification of Nouns

Once again, the first step in the annotation of adjectival modification of nouns is sense disambiguation. This step will proceed just as it did for the noun-based annotation for qualia in argument selection. The second subtask will involve answering questions that will supply the information needed for the creation of QLinks.

5.3.1 Subtask 1: Sense Disambiguation for Nouns

This task is identical to the one described in Section 5.2.1.

5.3.2 Subtask 2: Identifying the Qualia Involved

For this subtask, the annotator will see a list of sentences that all include that target noun in the same sense highlighted. This sense is provided at the top of the screen. The modifying adjective is also highlighted in each sentence. Each sentence is followed by a list of questions aimed at helping the annotator identify which quale is being acted on by the adjective. See figures 12 and 13 for a demonstration of this interface.

QLinks will be created based on which qualia the annotator chooses. If the annotator feels that more than one of the qualia is acted on, then as many QLinks as are needed can be created. The annotator also has the option of declining to annotate a particular sentence if he or she thinks choosing one or more of the qualia relations is not possible (”unclear from context”).

The reader is encouraged to read through Section 5.2 since that task is very close to this one. In the next section, we provide the resulting GLML markup for the examples included in Figures 12 and 13.

5.3.3 Resulting GLML markup

Examples given in Figure 12 for the noun *table*, generate the GLML markup that looks as follows:

1. *A polished table was in one corner, with dining chairs around it.*

```
A <SELECTOR sid="s1">polished</SELECTOR>
<NOUN nid="n1">table</NOUN>
was in one corner, with dining chairs around it .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="FORMAL"/>
```

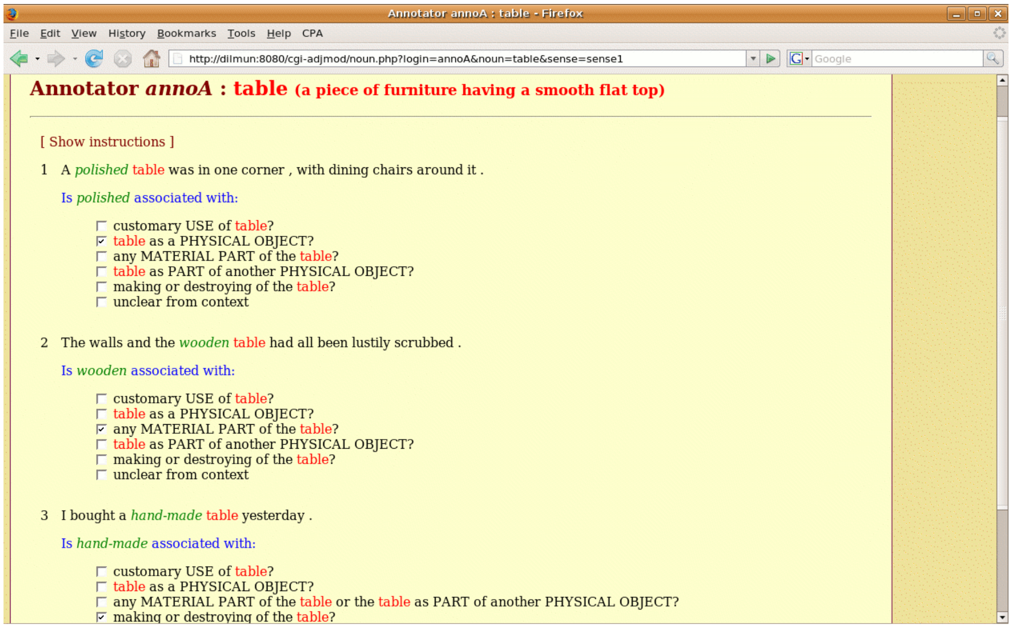


Figure 12: Qualia in Adjectival Modification for *table*

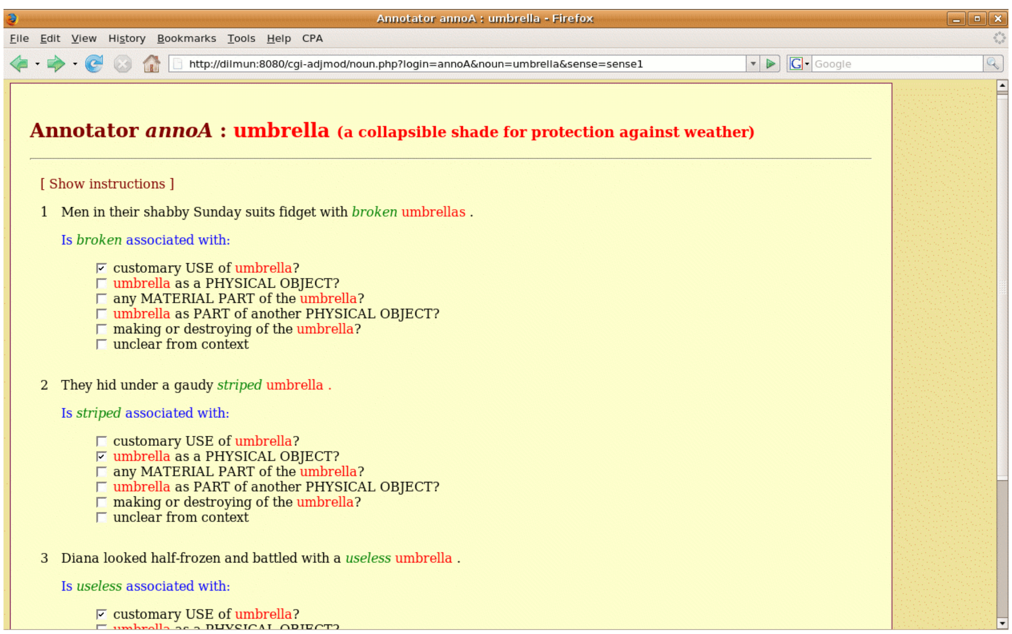


Figure 13: Qualia in Adjectival Modification for *umbrella*

2. *The walls and the wooden table had all been lustily scrubbed.*

```
The walls and the
<SELECTOR sid="s1">wooden</SELECTOR>
<NOUN nid="n1">table</NOUN>
had all been lustily scrubbed .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="CONST"/>
```

3. *I bought a hand-made table yesterday.*

```
I bought a
<SELECTOR sid="s1">hand-made</SELECTOR>
<NOUN nid="n1">table</NOUN>
yesterday .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="AGENTIVE"/>
```

Examples given in Figure 13 for the noun *umbrella*, generate the GLML markup that looks as follows:

1. *Men in their shabby Sunday suits fidget with broken umbrellas.*

```
Men in their shabby Sunday suits fidget with
<SELECTOR sid="s1">broken</SELECTOR>
<NOUN nid="n1">umbrellas</NOUN> .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="TELIC"/>
```

2. *They hid under a gaudy striped umbrella.*

```
They hid under a gaudy
<SELECTOR sid="s1">striped</SELECTOR>
<NOUN nid="n1">umbrella</NOUN> .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="FORMAL"/>
```

3. *Diana looked half-frozen and battled with a useless umbrella.*

```
Diana looked half-frozena and battled with a
<SELECTOR sid="s1">useless</SELECTOR>
<NOUN nid="n1">umbrella</NOUN> .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="TELIC"/>
```

5.4 Qualia Selection in Modification Constructions: Noun-headed Compounds

5.4.1 Subtask 1: Sense Disambiguation for Nouns

This task is identical to the one described in Section 5.2.1.

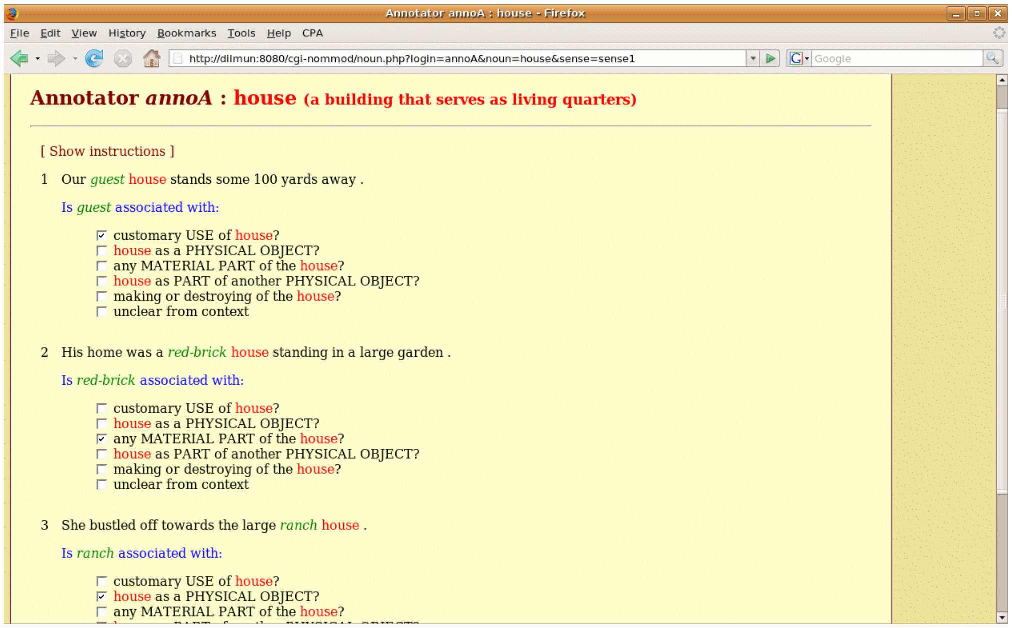


Figure 14: Qualia in Noun-headed Compounds for *house*

5.4.2 Subtask 2: Identifying the Qualia Involved

This subtask is similar to the one defined in section 5.3.2. Figure 14 shows the annotation interface for this task. The next section illustrates the GLML created for these examples.

5.4.3 Resulting GLML markup

Examples given in Figure 14 for the noun *house*, generate the GLML markup that looks as follows:

1. *Our guest house stands some 100 yards away.*

```
Our
<SELECTOR sid="s1">guest</SELECTOR>
<NOUN nid="n1">house</NOUN>
stands some 100 yards away .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="TELIC"/>
```

2. *His home was a red-brick house standing in a large garden.*

```
His home was a
<SELECTOR sid="s1">red-brick</SELECTOR>
<NOUN nid="n1">house</NOUN>
standing in a large garden .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="CONST"/>
```

3. *She bustled off towards the large ranch house.*

```
She bustled off towards the large
<SELECTOR sid="s1">ranch</SELECTOR>
<NOUN nid="n1">house</NOUN> .
<QLink qid="qid1" sID="s1" relatedToNoun="n1" qType="FORMAL"/>
```

5.5 Type Selection in Modification of Dot Objects

This task asks the annotator to choose which part of a dot object (complex type) is exploited in modification contexts. As always, the first step is disambiguation of the target, which, in this case is nouns denoting dot objects. Subtask 2 for this task is to choose the exploited type from a list of component types for the complex noun. A CompLink is then created with the exploited type specified in the the targetType field.

5.5.1 Subtask 1: Sense Disambiguation for Nouns

This task is identical to the one described in Section 5.2.1. Note that the sense inventory for dots will include only homonyms. So the nouns such as *bank* will need to be disambiguated between the "river bank" sense and the "financial institution" sense (contrastive senses), but not between the financial organization itself and the building where it is located (complementary senses).

5.5.2 Subtask 2: Identifying the Type Selected

Each dot object will have a list of types associated with it during the Data Set Construction stage. The annotator will be presented with a list of sentences that all include the same target dot object highlighted. The modifier is also highlighted. The list of dot type components follows each sentence. The annotator should choose the type that is being exploited by the modifier. In some cases, more than one or all of the types may be used as in the case of *long test* (INFO • EVENT). In cases such as this, the annotator should check some or all of the boxes, depending on the example. Figures 15 and 16 show a sample annotation for this task. The resulting GLML markup is provided in the next section.

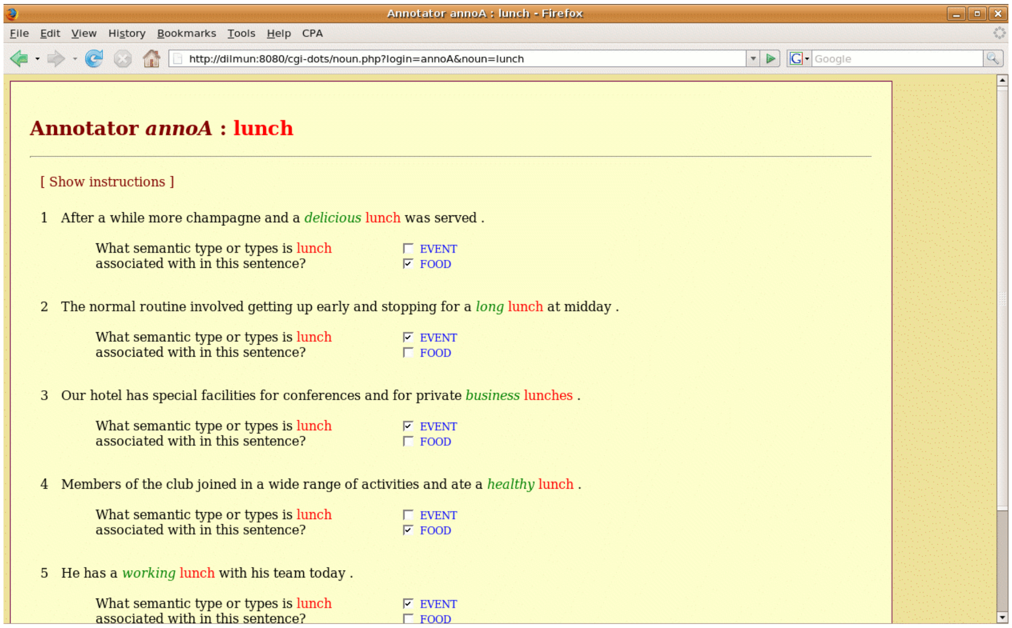


Figure 15: Identifying Selected Type for *lunch*

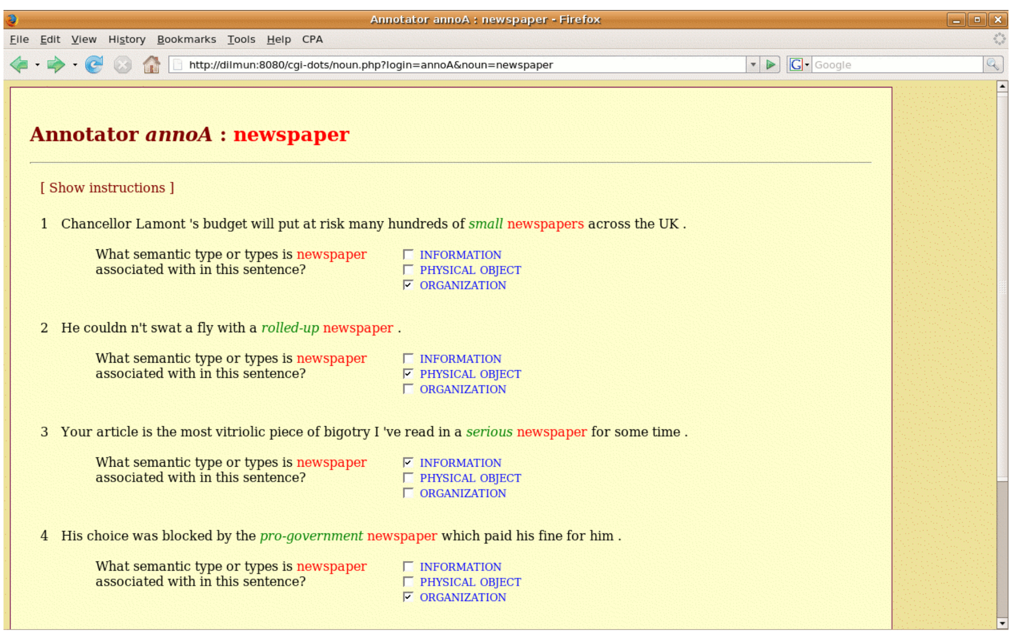


Figure 16: Identifying Selected Type for *newspaper*

5.5.3 Resulting GLML markup

Examples given in Figure 15 for the noun *lunch* generate the GLML markup that looks as follows:

1. *After a while more champagne and a delicious lunch was served.*

```
After a while more champagne and a
<SELECTOR sid="s1">delicious</SELECTOR>
<NOUN nid="n1">lunch</NOUN>
was served .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="mod"
compType="SELECTION" sourceType="[PHYS_OBJ,EVENT]"
targetType="PHYS_OBJ" />
```

2. *The normal routine involved getting up early and stopping for a long lunch at midday.*

```
The normal routine involved getting up early and stopping for a
<SELECTOR sid="s1">long</SELECTOR>
<NOUN nid="n1">lunch</NOUN>
at midday .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="mod"
compType="SELECTION" sourceType="[PHYS_OBJ,EVENT]"
targetType="EVENT" />
```

3. *Our hotel has special facilities for conferences and for private business lunches.*

```
Our hotel has special facilities for conferences and for private
<SELECTOR sid="s1">business</SELECTOR>
<NOUN nid="n1">lunches</NOUN> .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="mod"
compType="SELECTION" sourceType="[PHYS_OBJ,EVENT]"
targetType="EVENT" />
```

4. *Members of the club joined in a wide range of activities and ate a healthy lunch.*

```
Members of the club joined in a wide range of activities and ate a
<SELECTOR sid="s1">healthy</SELECTOR>
<NOUN nid="n1">lunch</NOUN> .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1"  gramRel="mod"
compType="SELECTION" sourceType="[PHYS_OBJ,EVENT]"
targetType="PHYS_OBJ" />
```

5. *He has a working lunch with his team today.*

```
He has a
<SELECTOR sid="s1">working</SELECTOR>
<NOUN nid="n1">lunch</NOUN>
with his team today .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1"  gramRel="mod"
compType="SELECTION" sourceType="[PHYS_OBJ,EVENT]"
targetType="EVENT" />
```

Examples given in Figure 16 for the noun *newspaper*, generate the GLML markup that looks as follows:

1. *Chancellor Lamont's budget will put at risk many hundreds of small newspapers across the UK.*

```
Chancellor Lamont 's budget will put at risk many hundreds of
<SELECTOR sid="s1">small</SELECTOR>
<NOUN nid="n1">newspapers</NOUN>
across the UK .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="mod"
compType="SELECTION" sourceType="[ORGANIZATION,[PHYS_OBJ,INFO]]"
 targetType="ORGANIZATION"/>
```

2. *He couldn't swat a fly with a rolled-up newspaper.*

```
He could n't swat a fly with a
<SELECTOR sid="s1">rolled-up</SELECTOR>
<NOUN nid="n1">newspaper</NOUN> .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="mod"
compType="SELECTION" sourceType="[ORGANIZATION,[PHYS_OBJ,INFO]]"
targetType="PHYSICAL_OBJECT"/>
```

3. *Your article is the most vitriolic piece of bigotry I 've read in a serious newspaper for some time.*

```
Your article is the most vitriolic piece of bigotry I 've read in a
<SELECTOR sid="s1">serious</SELECTOR>
<NOUN nid="n1">newspaper</NOUN>
for some time .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="mod"
compType="SELECTION" sourceType="[ORGANIZATION,[PHYS_OBJ,INFO]]"
targetType="INFORMATION"/>
```

4. *His choice was blocked by the pro-government newspaper which paid his fine for him.*

```
His choice was blocked by the
<SELECTOR sid="s1">pro-government</SELECTOR>
<NOUN nid="n1">newspaper</NOUN>
which paid his fine for him .
<CompLink cid="cid1" sID="s1" relatedToNoun="n1" gramRel="mod"
compType="SELECTION" sourceType="[ORGANIZATION,[PHYS_OBJ,INFO]]"
targetType="ORGANIZATION"/>
```

6 Specification of GLML

There are (currently) two basic link types: CompLink and QLink. CompLink is a composition link. It represents the relation between the predicate (viewed as function) and its argument. It is specified as follows.

```
attributes ::= [cid]
       selectorID
       relatedToNoun
       compType
       type
       targetType
       sourceType
       [sourceSubtype]
       [origin][comment][syntax]
cid ::= ID
{cid ::= ComplinkID
 ComplinkID ::= c<integer>}
sID ::= IDREF
{sID ::= selectorID
 selectorID ::= s<integer>}}
relatedToNoun ::= IDREF
{relatedToNoun ::= NounID}
compType ::= 'SELECTION' | 'COERCION'
type ::= 'HUMAN' | 'ANIMATE' | 'ORGANIZATION' | 'PHYSICAL_OBJECT' | 'ARTIFACT' |
               'EVENT' | 'ABSTRACT_ENTITY' | 'PROPOSITION' | 'INFORMATION' |
               'SENSATION' | 'LOCATION' | 'TIME_PERIOD'
sourceType ::= type | [type, type]
targetType ::= type | [type, type]
sourceSubtype ::= 'ATTITUDE' | 'EMOTION' | 'PROPERTY' | 'PRIVILEGE' |
               'OBLIGATION' | 'RULE' | OtherSubtype
{OtherSubtype ::= CDATA}
gramRel ::= 'dobj' | 'subj' | 'iobj' | 'mod'
origin ::= CDATA
comment ::= CDATA
syntax ::= CDATA
```

QLink is a "Qualia link" and expresses a relation between a predicate and its argument that involves a specific qualia value. QLink is specified as follows:

```
attributes ::= [qid]
       selectorID
       relatedToNoun
       gramRel
       qType
       [origin][comment][syntax]
qid ::= ID
{qid ::= QLinkID
 QLinkID ::= q<integer>}
sID ::= IDREF
{sID ::= selectorID
 selectorID ::= s<integer>}
```

```
relatedToNoun ::= IDREF
{relatedToNoun ::= NounID}
qType ::= 'AGENTIVE' | 'TELIC' | 'CONST' | 'FORMAL' | 'INVERSE_CONST'
gramRel ::= 'dobj' | 'subj' | 'iobj' | 'mod'
origin ::= CDATA
comment ::= CDATA
syntax ::= CDATA
```

7 Extensions to GLML

This document describes how the basic mechanisms of GL can be identified with an annotated corpus in a systematic fashion. There are still many areas of GL theory that are not addressed in this document, however, and many extensions to GLML are proposed. Among these additions are the following linguistic constructions:

1. **Modal Attribution in Adjectival Modification** (cf. Bouillon, 1997):
   a noisy crowd, existential.
   a noisy room, modal.

2. **Adverbial Modification**:

   (a) agentive adverbials: associated with AGENTIVE;

   (b) purpose and rationale clauses: associated with TELIC;
   John bought a pizza to eat.
   John bought a fancy car to impress his friends.

3. **Agentive Nominal Classification** (cf. Busa, 1996, Pustejovsky, 1995):

   (a) Individual level nominals (ILN) associated with the TELIC role:
   1. expressing ability or capacity: *violinist*, *swimmer*
   2. expressing a habit: *smoker*, *drinker*

   (b) Stage level nominals (SLN) expressing a temporal property, associated with the AGENTIVE role:
   passenger, *batter*, *pedestrian*.

As the tasks become more refined, the extensions will also become clearer, Furthermore, as other languages are examined for annotation, new tasks will emerge reflecting perhaps language-specific constructions.

References

N. Asher and J. Pustejovsky. 2006. A type composition logic for generative lexicon. *Journal of Cognitive Science*, 6:1–38.

A. Bisetto and S. Scalise. 2005. The classification of compounds. *Lingue e Linguaggio*, 2:319–332.

P. Bouillon. 1997. *Polymorphie et semantique lexical: le case des adjectifs*. PhD dissertation, Paris VII, Paris.

T. Briscoe and J. Carroll. 2002. Robust accurate statistical annotation of general text. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002), Las Palmas, Canary Islands, May 2002*, pages 1499–1504.

F. Busa. 1996. *Compositionality and the Semantics of Nominals*. PhD dissertation, Brandeis University, Waltham, MA.

P. Hanks and J. Pustejovsky. 2005. A pattern dictionary for natural language processing. *Revue Française de Linguistique Appliquée*.

M. Johnston and F. Busa. 1999. The compositional interpretation of compounds. In E. Viegas, editor, *Breadth and Depth of Semantics Lexicons*, pages 167–167. Dordrecht: Kluwer Academic.

J. N. Levi. 1978. *The Syntax and Semantics of Complex Nominals*. Academic Press, New York.

K. Markert and M. Nissim. 2007. Metonymy resolution at semeval i: Guidelines for participants. In *Proceedings of the ACL 2007 Conference*.

J. M. Moravcsik. 1975. Aitia as generative factor in aristotle's philosophy. *Dialogue*, 14:622–636.

M. Palmer, D. Gildea, and P. Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

J. Pustejovsky, P. Hanks, and A. Rumshisky. 2004. Automated Induction of Sense in Context. In *COLING 2004, Geneva, Switzerland*, pages 924–931.

J. Pustejovsky. 1991. The generative lexicon. *Computational Linguistics*, 17(4).

J. Pustejovsky. 1995. *Generative Lexicon*. Cambridge (Mass.): MIT Press.

J. Pustejovsky. 2000. Events and the semantics of opposition. In C. Tenny and J. Pustejovsky, editors, *Events as Grammatical Objects*, pages 445–482. Center for the Study of Language and Information (CSLI), Stanford, CA.

J. Pustejovsky. 2001. Type construction and the logic of concepts. In *The Syntax of Word Meaning*. Cambridge University Press, Cambridge.

J. Pustejovsky. 2005. A survey of dot objects. Technical report, Brandeis University.

J. Pustejovsky. 2007. Type Theory and Lexical Decomposition. In P. Bouillon and C. Lee, editors, *Trends in Generative Lexicon Theory*. Kluwer Publishers (in press).

B. Rosario and M. Hearst. 2001. Classifying the semantic relations in noun compounds via a domain-specific lexical hierarchy. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*.

A. Rumshisky and O. Batiukova. 2008. Polysemy in verbs: systematic relations between senses and their effect on annotation. In *COLING Workshop on Human Judgement in Computational Linguistics (HJCL-2008)*, Manchester, England. submitted.

A. Rumshisky, P. Hanks, C. Havasi, and J. Pustejovsky. 2006. Constructing a corpus-based ontology using model bias. In *The 19th International FLAIRS Conference, FLAIRS 2006*, Melbourne Beach, Florida, USA.

A. Rumshisky, V. A. Grinberg, and J. Pustejovsky. 2007. Detecting Selectional Behavior of Complex Types in Text. In P. Bouillon, L. Danlos, and K. Kanzaki, editors, *Fourth International Workshop on Generative Approaches to the Lexicon*, Paris, France.

J. Ruppenhofer, M. Ellsworth, M. Petruck, C. Johnson, and J. Scheffczyk. 2006. *FrameNet II: Extended Theory and Practice*.

A. Spencer. 1991. *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Blackwell Textbooks in Linguistics, Oxford, UK and Cambridge, USA.

B. Warren. 1978. *Semantic Patterns of Noun-Noun Compounds*. Acta Universitatis Gothoburgensis, Göteborg.