Incorporating Coercive Constructions into a Verb Lexicon

Claire Bonial*, Susan Windisch Brown*, Jena D. Hwang*, Christopher Parisien**,
Martha Palmer* and Suzanne Stevenson**

*Department of Linguistics, University of Colorado at Boulder
**Department of Computer Science, University of Toronto
{Claire.Bonial, Susan.Brown, hwangd, Martha.Palmer}@colorado.edu
{chris, suzanne}@cs.toronto.edu

Abstract

We take the first steps towards augmenting a lexical resource, VerbNet, with probabilistic information about coercive constructions. We focus on CAUSED-MOTION as an example construction occurring with verbs for which it is a typical usage or for which it must be interpreted as extending the event semantics through coercion, which occurs productively and adds substantially to the relational semantics of a verb. However, through annotation we find that VerbNet fails to accurately capture all usages of the construction. We use unsupervised methods to estimate probabilistic measures from corpus data for predicting usage of the construction across verb classes in the lexicon and evaluate against VerbNet. We discuss how these methods will form the basis for enhancements for VerbNet supporting more accurate analysis of the relational semantics of a verb across productive usages.

1 Introduction

Automatic semantic analysis has been very successful when taking a supervised learning approach on data labeled with sense tags and semantic roles (e.g., see Màrquez et al., 2008). Underlying these recent successes are lexical resources, such as PropBank (Palmer et al., 2005), VerbNet (Kipper et al., 2008), and FrameNet (Baker et al., 1998; Fillmore et al., 2002), which encode the relational semantics of numerous lexical items, especially verbs. However, because authors and speakers use verbs productively in previously unseen ways, semantic analysis systems must not be limited to direct extrapolation from previously seen usages licensed by static lexical resources (cf. Pustejovsky & Jezek, 2008). To achieve more accurate semantic analyses, we must augment such resources with knowledge of the extensibility of verbs.

Central to verb extensibility is the process of semantic and syntactic coercion. Coercion allows a verb to be used in "atypical" contexts that extend its relational semantics, thereby enabling expression of a novel concept, or simply more fluid expression of a complex concept. For example, consider a strictly intransitive action verb such as *blink*. This verb may instead be used in a construction with an object, as in *She blinked the snow off her lashes*, leading to an interpretation of the verb in which the object is causally affected and changes location (the CAUSED-MOTION construction; Goldberg, 1995). This type of constructional coercion is common in language and underlies much extensibility of verb usages. Understanding such coercive processes thus has significant impact on how we should represent knowledge about verbs in a lexical resource.

Importantly, constructional coercion is not an all-or-nothing process – a word must be semantically and syntactically compatible in some respects with a context in order for its use to be extended to that context, but the restrictions on compatibility are not hard-and-fast rules (Langacker, 1987; Kay & Fillmore, 1999; Goldberg, 2006; Goldberg, to appear). Gradience of compatibility plays an important role in coercion, suggesting that a probabilistic approach may be necessary for encoding knowledge of constructional coercion in a verb lexicon (cf. Lapata & Lascarides, 2003).

Our hypothesis here is that, due to this gradient process of productivity, existing verb lexicons do not adequately capture the actual patterns of use of extensible constructions. In this paper, we focus on the CAUSED-MOTION (CM) construction as an initial test case. We first annotate the classes of an extensive verb lexicon, VerbNet, as to whether the CM construction is allowed for all, some, or none of the verbs in the class, noting additionally whether it is a typical or coerced usage. We find that many of the classes that allow the construction for at least some verbs do not include the CM frame in their definition, indicating a significant shortcoming in the relational knowledge encoded in the lexicon. Next, we

develop probabilistic measures for determining to what degree a class is likely to admit the CM construction. We then test our measures over corpus data, manually annotated for use of the CM construction. Finally, we present preliminary work on automatic techniques for calculating the proposed measures in an unsupervised way, to avoid the need for expensive manual annotation. This work forms the preliminary steps toward empirically augmenting VerbNet's predictive capabilities concerning the event semantics of verbs in coercible constructions.

2 Extensible Constructions and VerbNet

Construction grammar has much insight to offer on the topic of productivity and on the resulting statistical patterns and gradience of usages (e.g., Langacker, 1987; Kay & Fillmore, 1999; Goldberg, 2006). A construction is formally defined to be any pairing of linguistic form (e.g., a syntactic frame) and meaning. Words can be used in constructions to the extent that their lexical semantics is compatible with – or can be coerced to be compatible with – the semantic constraints on the construction.

It is this notion of constructional coercion, and degree of coercibility, that accounts for the richness of usages that go beyond those thought of as typical or definitional for a verb: by coercing a verb not normally associated with a particular frame to occur in it, the meaning of the event can take on additional properties not considered a core part of the verb's semantics. For example, in the case of the sentence discussed above, *She blinked the snow off her lashes*, it is not the verb but rather the CM construction itself that licenses the direct object and adds the notion of "motion causally affecting the object" to the event semantics. Amongst other examples of well-known constructional coercions are: (1) The CAUSE-RECEIVE construction has the syntactic form of NP-V-NP-NP. For example, in *Bob painted Sally a picture*, the simple transitive verb *paint* gains the CAUSE TO RECEIVE sense, in which Sally is the recipient and the picture is the transferred item. (2) The WAY construction has the form of NP-V-[POSS *way*]-PP. For example, in *Frank found his way to New York*, the construction allows the verb *find* to gain a motion reading (i.e., "Frank traveled to New York") that would not otherwise be allowed (e.g., **Frank found to New York*).

Recognizing such extensions to the relational semantics of verbs is very important for accurate semantic interpretation in NLP. However, precise specifications for capturing the notion of coercible constructions, such as are needed for a computational resource, have heretofore been lacking.

2.1 VerbNet & Knowledge of Constructions

Computational verb lexicons are key to supporting NLP systems aimed at semantic interpretation. Verbs express the semantics of an event being described as well as the relational information among participants in that event, and project the syntactic structures that encode that information. Verbs are also highly variable, displaying a rich range of semantic and syntactic behavior.

Verb classifications help NLP systems to deal with this complexity by organizing verbs into groups that share core semantic and syntactic properties. For example, VerbNet (derived from Levin's [1993] work, Kipper et al., 2008) is widely used for a number of semantic processing tasks, including semantic role labeling (Swier and Stevenson, 2004), the creation of semantic parse trees (Shi and Mihalcea, 2005), and implicit argument resolution (Gerber and Chai, 2010). The detailed semantic predicates listed with each VerbNet class also have the potential to contribute to text-specific semantic representations and, thereby, to tasks requiring inferencing (Zaenen et al., 2008; Palmer et al., 2009).

VerbNet identifies semantic roles and syntactic patterns characteristic of the verbs in each class makes explicit the connections between the syntactic patterns and the underlying semantic relations that can be inferred for all members of the class. Each syntactic frame in a class has a corresponding semantic representation that details the semantic relations between event participants across the course of the event. For example, one of the characteristic patterns listed for the Pour class is a CAUSED-MOTION pattern, which accounts for sentences like *She poured water from the pitcher into the bowl*. This is represented in VerbNet as follows:

Syntactic representation:
   NP V NP PP PP
   Agent V Theme Source Location

Semantic representation:
   MOTION (DURING(E), THEME)
   NOT (PREP (START(E), THEME, LOCATION))
   PREP (START(E), THEME, SOURCE)
   PREP (END(E), THEME, LOCATION)
   CAUSE (AGENT, E)

This representation details connections between the syntax and semantics using the semantic roles as links, indicating that the Agent is the Subject NP and has CAUSED the Event, and that the Theme is the Object NP and has a new LOCATION at the end of the event. These types of inferences provide the foundation for deep semantic analysis of text.

However, the specifications in VerbNet (as in other predicate lexicons, such as FrameNet, Baker et al., 1998; Fillmore et al., 2002) are seen as definitional – they are restricted to the core usages of the verbs that are valid for all verbs in the class. However, as noted above, people often use verbs productively, in ways that go beyond the boundaries of the verb class structure. It is important to correctly identify these productive usages when they occur, since they may be explicitly adding crucial inferences. If a construction is not recognized in the form of a syntactic frame in VerbNet, such inferences are not possible, greatly reducing VerbNet's utility and coverage. For example, creative uses of a verb, such as *She blinked the snow off her lashes*, would have no corresponding frame in *blink*'s class, the Hiccup class. It contains one intransitive frame:

NP V
Agent V
BODY_PROCESS (E, AGENT)
INVOLUNTARY (E, AGENT)

Sentences that coerce the meaning of *blink* to fit with a CM event would currently be misanalysed. One option might be to augment the Hiccup class with the CM frame from the Pour class, which would ensure that such sentences would be analyzed more accurately. However, given the productive nature of constructional coercion and its widespread applicability, the approach of adding any possible pattern to each class is not appropriate: this would undermine the definitional distinctions between classes and greatly lessen their usefulness.

Complicating the issue is the phenomenon of regular sense extensions (Dang et al., 1998), where what once may have been coercion has become entrenched and is now seen as a different sense of the verb. For example, the verbs in the Push class express the general meaning of exerting force on an object, such as *She pushed on the wall*. Often, the exertion of force moves the object, which can be expressed in a CM construction such as *She pushed the box across the room*. VerbNet accounts for this regular sense extension by including most of the Push verbs in the Carry class as well, which has the CM construction as one of its frames. Deciding when to include a verb in another class based on regular sense extensions, when to add a frame for a construction to a class, or when to reject the frame as a defining part of a class, is made difficult by the graded nature of matches between verbs and a construction. Our goal is to maintain the advantages of the class structure of VerbNet while enhancing it with a graded view of the applicability of a construction for each class. Noting the applicability of a construction will enable the inclusion of its appropriate semantic predicates, and the inferencing over them, which are currently not supported.

3   Our Proposal: Constructional Profiles

We aim to augment VerbNet with knowledge of constructions that are likely to be used extensibly with a range of verbs. Such extensible constructions will be core usages for some classes (such as the CM for the Pour class, as noted above) but will be less characteristic of the fundamental semantics of other verb classes (such as CM for the Hiccup class). We propose to identify such a construction and its varying roles in the different classes by using relevant statistics over usages of verbs in a corpus – what we call a *constructional profile*.

A constructional profile is a probabilistic assessment of the usage of a particular construction by the verbs in a class. We developed the following three measures to capture the relevant behavior, with the goal of providing both type- and token-based views of the behavior of a verb class with respect to a target construction:

P1	$P_{type}(X|C)$: *probability that a verb type in class C is attested in construction X*
	P1 gives a type-based assessment, indicating how widespread the use of the construction is across the verb types in the class. For example, if 8 out of 10 members of a class appear with the construction, we might estimate P1 as 0.8.
P2	$P_{token}(X|C)$: *probability that the instances of a typical verb in class C occur in construction X*
	P2 gives a token-based assessment, indicating, for a typical verb in the class, the relative amount of usage of the construction among all usages of the verb. For example, to estimate this, we might average across all verbs in the class, the percentage of tokens in this construction.
P3:	$P_{token}(X|X\text{-verbs-in-}C)$: *same as P2 but considering only verbs that have been attested in construction X*
	P3 is the same as P2, but looking only at those verbs in the class that have an attested usage of the construction, removing verbs without attested usages.

We hypothesize that these measures will have high values for those classes for which the construction should be definitional; very low values for those classes that are not compatible with the construction; and varying values for those classes that allow coerced usages to a greater or lesser extent.

Although these probabilities are intuitively very simple, estimating them from corpus data poses a significant challenge. Since a construction is a pairing of form with meaning, recognizing the use of a particular

construction is not simply a matter of determining the syntactic pattern of the usage; rather, certain semantic properties and relations must co-occur with the syntactic pattern. Earlier work has shown that a supervised learning method was able to discriminate potential usages of the CM construction given training sentences manually labeled as either CM or not (Hwang et al., 2010). Here, we aim instead to identify usages of the CM construction, but without requiring an expensive manual annotation effort. That is, we seek an unsupervised method for estimating the probabilities in P1–P3 above.

We approach this goal in steps as follows. First, we examine all the classes in VerbNet to see which allow the CM construction (Section 4). This anno-tation reveals shortcomings in VerbNet's representa-tion (classes that allow the CM construction but do not list it) and also provides a gold standard with which to evaluate our method of identifying an exten-sible construction using our constructional profiles. Second, we use the manually annotated CM construction data from Hwang et al. (2010) to estimate probabilities P1–P3 using maximum likelihood formulations (Section 5). An analysis of the predictive power of these constructional profile measures shows a good match with the distinctions made in the human annotation of the classes. Thus, our annotation based constructional profile measures show promise for identifying relevant behaviors of the construction across the classes. Third, we explore automatic methods for estimating the constructional profile measures without the need for manual annotations (Section 6). We use a hierarchical Bayesian model that learns verb classes from corpus data to provide unsupervised estimates of the constructional profiles, which also exhibit the relevant distinctions across the classes.

4 Annotating the VerbNet Resource

We begin with a manual examination of the resource and a thorough annotation of the status of each class with respect to the CM construction. This effort reveals a number of shortcomings in VerbNet, and the need for developing methods that can support the extension of VerbNet to better reflect the coercive uses of constructions across the classes. The annotation described here also forms the basis for the evaluation in the following sections of our new probabilistic measures, by motivating hypotheses about the expected patterns of use of the CM construction across the classes.

4.1 Annotation Guidelines and Results

The first goal of our manual annotation of VerbNet classes was to determine which classes currently represent CM in one of their frames. To this end, we identified which classes contain the following frame:

NP [Agent/Cause]-V-NP [Patient/Theme]-
PP [Source/Destination/Recipient/Location]

These frames correspond to classes such as Slide, with its frame NP-V-NP-PP.Destination: *Carla slid the books to the floor*. We also examined classes with the patterns NP-V-NP-PP.Oblique, NP-V-NP-PP. Theme2, and NP-V-NP-PP.Patient2. In these classes, annotators had to judge whether the final PP was compatible with CM. For example, the Breathe class contains the frame NP-V-NP.Theme-PP.Oblique, *The dragon breathed fire on Mary*, which is compatible with CM; whereas the same basic frame in the Other_cos class is not: NP V NP PP.Oblique, *The summer sun tanned her skin to a golden bronze*.

In addition, we annotated which classes were potentially compatible with CM for either all verbs in the class or only some verbs. The "some" classification has the drawback that it may be applied to classes with very different proportions of compatible verbs; while suitable for our exploratory work here, we plan to make finer distinctions in the future. A secondary determination was whether or not the class was compatible with CM as part of its core semantics, or if it was compatible with CM because it was coercible into the construction. A verb was considered "compatible with CM" and "not coerced" if the verb could be used in the CM construction and its semantics, as reflected in VerbNet's semantic predicates, involved a CAUSE predicate in combination with another predicate such as CONTACT, TRANSFER, (EN)FORCE, EMIT, TAKE_IN (predicates potentially involving movement along some path). For example, although CM is not already included as a frame for the Bend class containing the verb *fold*, the semantics of this class include CAUSE and CONTACT, and the verb can be used in a CM construction: *She folded the note into her journal*. Therefore, this class would have been considered "compatible with CM" but "not coerced". Conversely, a verb was considered "compatible with CM" and "coerced" if the verb could be used in the CM construction, yet its semantics, again as reflected in VerbNet, did not involve CAUSE and MOVEMENT ALONG A PATH (e.g., the verb *wiggle* of the Body_internal_motion class: *She wiggled her foot out of the boot*).

In summary, as presented in the table below, we annotated each class according to whether (1) the CM construction was already represented in VerbNet for this class, (2) the construction was possible for all, some, or

none of the verbs in that class, and (3) the verbs of any class compatible with CM were coerced into the construction or not. The classification for (3) was made regardless of whether "all" verbs or only "some" were compatible with CM. This determination was made uniformly for a class: there were no classes in which only certain CM-compatible verbs were considered "coerced".

VN class example [# of classes like this]	CM in VN	CM is possible	CM is coerced
Banish [50]	Yes	All	No
Nonverbal_Expression [2]	Yes	All	Yes
Cheat [6]	Yes	Some	No
Exhale [18]	No	All	No
Hiccup [30]	No	All	Yes
Fill [46]	No	Some	No
Wish [54]	No	Some	Yes
Matter [64]	No	None	N/A

Notably, we identified 206 classes where at least some of the verbs in that class are compatible with the CM construction; however, VerbNet currently only recognizes the CM construction in 58 classes. There were several classes of interest: First, although it may seem unusual that CM is represented in 6 classes where we found that only "some" verbs were compatible with CM (e.g., Cheat class), these were cases where only more restricted subclasses are compatible with CM, and this syntactic frame is listed for that subclass. This suggests subclasses may provide a more precise characterization of which verbs are compatible with a construction.

Secondly, we identified 18 classes in which all verbs were compatible with CM without coercion; thus, these classes could likely be improved by the addition of the CM syntactic frame. Additionally, we found 30 classes in which all verbs are coercible into the CM construction; however, the actual likelihood of a verb in those classes occurring in a CM construction remains to be investigated in the following sections. Like those classes where it was determined that only "some" verbs are compatible with CM, usefully incorporating the CM construction into classes that require coercion relies on accurately determining the probability that verbs in those classes will actually appear in the CM construction.

For those classes in which "all" verbs are compatible with CM, our intuition was that some aspect of the verb's semantics either inherently includes or allows the verb to be coerced into the CM construction. Conversely, for those classes in which no verbs are compatible with CM, presumably some aspect of the verb's semantics is logically incompatible with CM. Although pinpointing

precisely what aspect of a verb's semantics makes it compatible with CM may not be possible, we can investigate whether or not our intuitions are supported by examining the actual frequencies of CM constructions for given verbs or a given class.

4.2 Hypotheses

Using these annotations, we were able to develop two simple hypotheses.

Hypothesis 1: We expect the constructional profile measures for the CM construction in a given corpus to be highest for those classes in which all verbs were found to be compatible with CM; lower for classes in which only some verbs were found to be compatible; and lowest for classes in which no verbs were found to be compatible.

Hypothesis 2: We expect the constructional profile measures for the CM construction in a given corpus to be highest for verbs that fall into classes where CM is not considered coerced (for either some or all of the verbs in the class); lower for verbs that fall into classes in which the CM construction only works through coercion (for either some or all of the verbs in the class); and lowest for verbs that fall into classes in which no verbs are compatible with CM.

To investigate Hypothesis 1, we grouped the annotated classes according to whether all, some, or no verbs in the class are compatible with CM:

	Class example	# of classes
Allowed by All	Bring, Carry	106
Allowed by Some	Appoint, Lodge	100
Allowed by None	Try, Own	64

To investigate Hypothesis 2, we did a second grouping of the classes according to whether CM is not coerced, CM is coerced, or CM is simply not compatible with the class. This second grouping did not distinguish whether CM was compatible with "all" or "some" of the verbs in a given class.

	Class example	# of classes
Not Coerced	Put, Throw	120
Coerced	Floss, Wink	86
Not Compatible	Differ	64

5 Evaluation using Constructional Profiles

5.1 Annotated data description

Our research uses the data annotated for Hwang et al. (2010), in which 1800 instances in the form NP-V-NP-PP were identified in the Wall Street Journal portion of the Penn Treebank II (Marcus et al., 1994). Each instance

of the data was single annotated with one of the two labels: CM or non-CM. The annotation guidelines were based on the CM analysis of Goldberg (1995).

Our analysis began with the same data but adopted a slightly narrower definition of CM. We diverged from the Hwang et al. (2010) study in the following two ways: (1) sentences where the object NP is an item that is created by the event denoted by the verb were not considered CM (e.g., *Mr. Pilson scribbled a frighteningly large figure on a slip of paper*, where the figure is created through the scribbling event); and (2) sentences in which movement is prevented were not considered CM (e.g., *He kept her at arm's length*). In agreement with Hwang et al., our annotation included both metaphorical senses (e.g., *[It] cast a shadow over world oil markets*) and literal senses (e.g., *The company moved the employees to New York*) of CM. Our annotation using the narrower guidelines resulted in 85.8% agreement with the original annotation.[1] The distribution of labels in our data is 21.8% for CM and 78.2% for NON-CM.

5.2 Annotated data description

Using statistics over the manually annotated data, we calculate maximum likelihood estimates of the three constructional profile measures introduced in Section 3, as follows. First, let the probability that a verb v is used in the CM construction be estimated as:

$$P(CM|v,C) = \frac{\#(CM\ usages\ of\ v \in C)}{\#(CM + non\text{-}CM\ usages\ of\ v \in C)}$$

That is, $P(CM|v,C)$ is estimated as the relative frequency of the CM construction for v out of all annotated usages of v that are labeled as class C. Now let C_{CM} be all verbs v in C with at least one usage annotated as CM; i.e.:

$$C_{CM} = \{v \in C \mid P(CM|v,C) > 0\}$$

Then we calculate estimates of P1–P3 as:

P1: $P_{type}(CM|C) = |C_{CM}|/|C|$
This measure indicates how widespread the use of CM is across the verb types in the class.

P2: $P_{token}(CM|C) = [\sum_{v \in C} P(CM|v,C)]/|C|$
The average over all verbs v in C of $P(CM|v,C)$
This indicates the relative amount of usage of CM among all usages of the verbs in the class.

P3: $P_{token}(CM|v,C) = [\sum_{v \in C_{CM}} P(CM|v,C))]/|C_{CM}|$
The average over all verbs v in C_{CM} of $P(CM|v,C)$
P3 narrows the P2 measure to only those verbs in the

class for which there is an attested usage of CM.

5.3 Analysis of the Constructional Profiles

The tables below provide a summary of the profile measures P1-P3 for the groups of VerbNet classes as defined in section 4.2. For each group listed, we report the averages of P1-P3 over all classes in the group where at least one verb in the class occurred in the data manually annotated for CM usage.

	P1	P2	P3
CM Allowed by All	0.413	0.323	0.437
CM Allowed by Some	0.087	0.078	0.224
CM Not Allowed	0.055	0.055	0.083

As seen here, the constructional profile measures over CM in the data corroborate our Hypothesis 1 (Section 4.2). All three measures on average are highest for the classes that fall into the "all allowed" group, next highest for those in the "some allowed" group, and lowest for the "not allowed" classes.

	P1	P2	P3
CM Non-Coerced	0.354	0.274	0.418
CM Coerced	0.091	0.091	0.185
CM Not Allowed[2]	0.056	0.056	0.083

Furthermore, the second table here confirms our expectations for Hypothesis 2 (Section 4.2). Again, all three measures on average are highest for classes that fall into the "non-coerced" group, next highest for classes in the "coerced" group (in which the construction is achievable only through coercion), and lowest for the "not allowed" group.

Thus, our two hypotheses are borne out, showing that our constructional profile measures, when estimated over manually annotated data, can be useful in capturing important distinctions among classes of verbs with regard to their usage in an extensible construction such as CM.

6 Automatic Creation of Constructional Profiles Using a Bayesian Model

Manually annotating a corpus for usages of a construction can be prohibitively expensive, so we also investigate the use of automatic methods to estimate constructional profile measures. By using a hierarchi-cal Bayesian model (HBM) that acquires latent prob-abilistic verb classes from corpus data, we provide unsupervised

---

[1] We found that 34.0% of the disagreements were directly due to the changes in annotation resulting from our two new criteria.

[2] Note the non-zero values result from actual CM verb usages in the data belonging to classes believed to be not compatible with CM by VerbNet expert annotators.

estimates of the constructional profiles.

6.1 Overview of Model and Data

We use the HBM of Parisien & Stevenson (2011), a model that automatically acquires probabilistic knowledge about verb argument structure and verb classes from large-scale corpora. The model is based on a large body of research in nonparametric Bayesian topic modeling (e.g., Teh et al., 2004), a robust method of discovering syntactic and semantic structure in very large datasets. For each verb encountered in a corpus, the model provides an estimate of the verb's expected overall pattern of usage. By using latent probabilistic verb classes to influence these expected usage patterns, the model can, for example, estimate the probability that a verb like *blink* might occur in a CM construction, even if no such attested usages appear in the corpus.

In this preliminary study, we use the corpus data from Parisien & Stevenson (2011), since the model has been trained and evaluated on this data. As that study was aimed at modeling facts of child language acquisition, it uses child-directed speech from the Thomas corpus (Lieven et al., 2009), part of the CHILDES database (MacWhinney, 2000). In this preliminary study, we use their development dataset containing approx. 170,000 verb usages, covering approx. 1,400 verb types. (We reserve the test set for future experiments.) For each verb usage in the input, a number of features are automatically extracted that indicate the number and type of syntactic arguments occurring with the verb and general semantic properties of the verb. The semantic features are drawn from the set of VerbNet semantic predicates, such as CAUSE, MOTION, and CONTACT. These are automatically extracted from all classes compatible with the verb (with no sense disambiguation).

6.2 Measures for Constructional Profiles

Using the argument structure constructions, verb usage patterns and classes learned by the model, we estimate the three constructional profile measures in Section 3, as follows. First, we note that since the constructions acquired by the model are probabilistic in nature, a particular CM instance may be a partial match to more than one of the model's constructions.

For each verb in the input, we consider the likelihood of use of the CM construction to be the likelihood of a contrived frame intended to capture the important properties of a CM usage. F_{CM} is a usage taking a direct object and a prepositional phrase, and including the semantic features CAUSE and MOTION, with all other semantic features left unspecified. For a given verb v, we

estimate the likelihood of this CM usage, over all constructions in the model, as follows:

$$P(F_{CM}|v) = \sum_k P(F_{CM}|k)P(k|v)$$

Here, $P(F_{CM}|k)$ is the likelihood of the CM usage F_{CM} being an instance of the probabilistic construction k, and $P(k|v)$ is the likelihood that verb v occurs with construction k. These component probabilities are estimated using the probability distributions acquired by the model and averaged over 100 samples from the Markov Chain Monte Carlo simulation, as described in Parisien & Stevenson (2011).

Now, we let C_{CM} be the set of verbs in VerbNet class C where the expected likelihood of a CM usage is non-negligible (akin to the set of verbs with attested usage in Section 5.2):

$$C_{CM} = \{v \epsilon C \mid P(F_{CM}|v) > \lambda \}$$

where λ is a small threshold, here 0.0001. Note that since v is not disambiguated for class in our data, all usages of v contribute to this estimate.

The estimates of P1-P3 are comparable to those in Section 5.2. The difference is that since we are un-able to disambiguate individual usages of the verbs, each usage of v is considered to belong to all possible classes C of which v is a member. P1 is estimated as before; P2 and P3 are averages of $P(F_{CM}|v)$.

6.3 Analysis of the Constructional Profiles

The tables below provide a summary of the profile estimates P1-P3 for the groups of VerbNet classes as given in Section 4.2. For each group listed, we report the averages of P1-P3 over all classes in the group where at least one of the verbs in the class occurred in the training input to the model.

	P1	P2	P3
All allowed	0.569	0.0180	0.0250
Some allowed	0.449	0.0106	0.0192
Not allowed	0.363	0.0044	0.0079

These profile measures align with the hypotheses in Section 4.2 and with the measures based on manually annotated data in Section 5.2. The estimates are high-est for classes where all verbs permit the CM con-struction, second highest for classes where only some permit it, and lowest for classes that do not permit it.

	P1	P2	P3
CM non-coerced	0.546	0.0178	0.0260
CM coerced	0.458	0.0095	0.0167
CM not allowed	0.363	0.0044	0.0079

Again, the overall patterns of the profile measures align with Sections 4.2 and 5.2. The profile estimates are highest for classes annotated to be non-coerced usages of CM, second highest for coerced classes, and lowest for "not allowed".

The measures show the overall differences among classes in the different groups (for both groupings) – i.e., the average behavior among classes in the different groups varies as we predicted. This indicates that the measures are tapping into aspects of construction usage that are relevant to making the desired distinctions in VerbNet, and validates the use of automatic techniques. However, there is a substantial amount of variability in these measures across the classes, so we also consider how well the estimates can predict the appropriate group for individual classes. That is, can we automatically predict whether the CM construction can be used by all, some, or none of the verbs in a given verb class, and can we predict whether such usages are coerced?

We consider the P3 measure as it provides the best separation among the class groupings. The tables below report precision (P), recall (R) and F-measures (F) for each group, where 'all' and 'some' have been collapsed. For exploratory purposes, we pick P3 = 0.006 as the value that optimizes F-measures of this classification. Future work will explore more principled means for setting these thresholds.

	P	R	F
CM allowed	0.880	0.742	0.806
CM not allowed	0.407	0.636	0.497

Only a 2-way distinction can be made reliably for the allowed grouping. The F-score of over 80% for the "allowed" label is very promising. The low precision for the "not allowed" case suggests that the model can't generalize sufficiently due to sparse data.

	P	R	F
CM non-coerced	0.691	0.491	0.574
CM coerced	0.461	0.417	0.438
CM not allowed	0.406	0.709	0.517

We use thresholds of P3 = 0.021 to separate non-coerced from coerced classes, and P3 = 0.007 to separate coerced from not allowed classes. The model estimates show moderate success in distinguishing classes with coerced vs. non-coerced usage of the CM construction. However, our measures simply cannot distinguish non-occurrence due to semantic incompatibility from non-occurrence due to chance, given the expected low frequency of a novel

coerced use of a construction. To separate the allowed cases into whether they are coerced or not requires a more detailed assessment of the semantic compatibility of the class, which means looking at finer-grained features of verb usages that are indicative of the semantic predicates compatible with the particular construction. Moreover, this kind of assessment likely needs to be applied on a verb-specific (and not just class-specific) level, in order to identify those verbs out of a potentially coercible class that are indeed coercible (i.e., identifying the coercible verbs in a class labeled as "some allowed").

7 Conclusion

Our investigation demonstrates that VerbNet does not currently represent the CM construction for all verbs or verb classes that are compatible with this construction, and the existing static representation of verbs is inadequate for analyzing extensions of verb meaning brought about by coercion. The utility of VerbNet would be greatly enhanced by an improved representation of constructions: specifically, the incorporation of probabilities that verbs in a given (sub)class would occur in a particular construction, and whether this constitutes a regular sense extension. This addition to VerbNet would increase the resource's coverage of syntactic frames that are compatible with a given verb, and therefore enable appropriate inferences when coercion occurs. We have made preliminary steps towards developing this probabilistic distribution over both verb instances and classes, based on a large corpus. Unsupervised methods for estimating the probabilities achieve an F-score of over 80% in distinguishing the classes that allow the target construction. However, making distinctions among coerced and non-coerced cases will require us to go beyond these class-based probabilities to finer-grained, corpus-based assessments of a verb's semantic compatibility with a coercible construction.

To move beyond these preliminary findings, we must therefore shift our focus to the behavior of individual verbs. Additionally, to reduce the impact of errors resulting from low-frequency verbs and classes, we plan to expand our research to more data, specifically the OntoNotes TreeBank data (Weischedel et al., 2011). Finally, to achieve our ultimate goal of creating a lexicon that can flexibly account for a variety of constructions, we will examine other constructions as well. While determining the set of coercible constructions in a language is itself a topic of current research, we propose initially to include the widely recognized CAUSE-RECEIVE and WAY constructions in addition to CM.

References

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. *Proceedings of the 17th International Conference on Computational Linguistics (COLING/ACL-98),* pp. 86–90, Montreal.

Dang, HoaTrang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. Investigating regular sense extensions based on intersective Levin classes. *Proceedings of COLING-ACL98*, pp. 293–299.

Fillmore, Charles J., Christopher R. Johnson, and Miriam R.L. Petruck. 2002. Background to FrameNet. *International Journal of Lexicography,* 16(3):235-250.

Gerber, Matthew, and Joyce Y. Chai. 2010. Beyond NomBank: A study of implicit arguments for nominal predicates. *Proceedings of the 48th Annual Meeting of the Association of Computational Linguistics,* pp. 1583–1592, Uppsala, Sweden, July.

Goldberg, A. E. 1995. *Constructions: A construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Goldberg, A. E. 2006. *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

Goldberg, A. E. To appear. Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics.*

Hwang Jena D., Rodney D. Nielsen and Martha Palmer. 2010. Towards a domain-independent semantics: Enhancing semantic representation with construction grammar. *Proceedings of Extracting and Using Constructions in Computational Linguistic Workshop,* held with NAACL HLT 2010, Los Angeles, June.

Kay, P., and C. J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The What's X Doing Y? construction. *Language,* 75:1–33.

Kipper, Karin, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation Journal,* 42:21–40.

Langacker, R. W. 1987. *Foundations of cognitive grammar: Theoretical prerequisites.* Stanford, CA: Stanford University Press.

Lapata, M., and A. Lascarides. 2003. Detecting novel compounds: The role of distributional evidence. *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*(EACL03), pp.235–242. Budapest, Hungary.

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

Lieven, E., D. Salomo, and M. Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20(3):481–507.

MacWhinney, B. 2000.*The CHILDES Project: Tools for analyzing talk* (3rd ed., Vol. 2: The Database). Erlbaum.

Màrquez, L., X. Carreras, K. Litkowski, and S. Stevenson. 2008. Semantic role labeling: An introduction to the special issue. *Computational Linguistics,* 34(2): 145–159.

Martha Palmer, Jena D. Hwang, Susan Windisch Brown, Karin Kipper Schuler and Arrick Lanfranchi. 2009. Leveraging lexical resources for the detection of event relations. *Proceedings of the AAAI 2009 Spring Symposium on Learning by Reading*, Stanford, CA, March.

Palmer, Martha, Daniel Gildea, and Paul Kingsbury.2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics,* 31(1):71–106.

Parisien, Christopher, and Suzanne Stevenson. 2011. To appear in *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society,* Boston, MA, July.

Pustejovsky, J., and E. Jezek. 2008. Semantic coercion in language: Beyond distributional analysis. *Italian Journal of Linguistics/RivistaItaliana di Linguistica* 20(1): 181–214.

Shi, Lei, and Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. *Proceedings of the 6th International Conference on Intelligent Text Processing and Computational Linguistics,* Mexico City, Mexico.

Swier, R., and S. Stevenson. 2004. Unsupervised semantic role labeling. *Proceedings of the 2004 Conf. on Empirical Methods in Natural Language Processing,* pp. 95–102, Barcelona, Spain.

Teh, Y. W., M. I. Jordan, M. J.Beal, and D. M.Blei.2006. Hierarchical Dirichlet processes. *Jrnl of the American Statistical Asscn*, 101(476): 1566–1581.

Weischedel, R., E. Hovy, M. Marcus, M. Palmer, .R. Belvin, S. Pradan, L. Ramshaw and N. Xue. 2011.OntoNotes: A Large Training Corpus for Enhanced Processing. In Part 1: Data Acquisition and Linguistic Resources of *The Handbook of Natural Language Processing and Machine Translation: Global Automatic Language Exploitation,* Eds.: Joseph Olive, Caitlin Christianson, John McCary. Springer Verlag, pp. 54-63.

Zaenen, A., C. Condoravdi, and D. G. Bobrow. 2008. The encoding of lexical implications in VerbNet. *Proceedings of LREC 2008,* Morocco, May.