No sentence is too confusing to ignore

Paul Cook
Department of Computer Science
University of Toronto
Toronto, Canada
`pcook@cs.toronto.edu`

Suzanne Stevenson
Department of Computer Science
University of Toronto
Toronto, Canada
`suzanne@cs.toronto.edu`

Abstract

We consider sentences of the form *No X is too Y to Z*, in which X is a noun phrase, Y is an adjective phrase, and Z is a verb phrase. Such constructions are ambiguous, with two possible (and opposite!) interpretations, roughly meaning either that "Every X Zs", or that "No X Zs". The interpretations have been noted to depend on semantic and pragmatic factors. We show here that automatic disambiguation of this pragmatically complex construction can be largely achieved by using features of the lexical semantic properties of the verb (i.e., *Z*) participating in the construction. We discuss our experimental findings in the context of construction grammar, which suggests a possible account of this phenomenon.

1   No noun is too adjective to verb

Consider the following two sentences:

(1)   No interest is too narrow to deserve its own newsletter.

(2)   No item is too minor to escape his attention.

Each of these sentences has the form of *No X is too Y to Z*, where X, Y, and Z are a noun phrase, adjective phrase, and verb phrase, respectively. Sentence (1) is generally taken to mean that *every* interest deserves its own newsletter, regardless of how narrow it is. On the other hand, (2) is typically interpreted as meaning that *no* item escapes his attention, regardless of how minor it is. That is, sentences with the identical form of *No X is too Y to Z* either can mean that *"every* X Zs", or can mean the opposite—that *"no* X Zs"![1]

---

[1]Note that in examples (1) and (2), the nouns *interest* and *item* are the subjects of the verbs *deserve* and *escape*, respec-

This "verbal illusion" (Wason and Reich, 1979), so-called because there are two opposite interpretations for the very same structure, is of interest to us for two reasons. First, the contradictory nature of the possible meanings has been explained in terms of pragmatic factors concerning the relevant presuppositions of the sentences. According to Wason and Reich (1979) (as explained in more detail below), sentences such as (2) are actually nonsensical, but people coerce them into a sensible reading by reversing the interpretation. One of our goals in this work is to explore whether computational linguistic techniques—specifically automatic corpus analysis drawing on lexical resources—can help to elucidate the factors influencing interpretation of such sentences across a collection of actual usages.

The second reason for our interest in this construction is that it illustrates a complex ambiguity that can cause difficulty for natural language processing applications that seek to semantically interpret text. Faced with the above two sentences, a parsing system (in the absence of specific knowledge of this construction) will presumably find the exact same structure for each, giving no basis on which to determine the correct meaning from the parse. (Unsurprisingly, when we run the C&C Parser (Curran et al., 2007) on (1) and (2) it assigns the same structure to each sentence.) Our second goal in this work is thus to explore whether increased linguistic understanding of this phenomenon could be used to disambiguate such examples automatically. Specifically, we use this construction as an example of the kind of difficulties faced in semantic interpretation when meaning may be determined by pragmatic or other extra-syntactic factors, in order to explore whether

---

tively. In this construction the noun can also be the object of the verb, as in the title of this paper which claims no sentence can/should be ignored.

lexical semantic features can be used as cues to resolving pragmatic ambiguity when a complex semantico-pragmatic model is not feasible.

In the remainder of this paper, we present the first computational study of the *No X is too Y to Z* phenomenon, which attempts to automatically determine the meaning of instances of this semantically and pragmatically complex construction. In Section 2 we present previous analyses of this construction, and our hypothesis. In Section 3, we describe the creation of a dataset of instances that verifies that both interpretations ("every" and "no") indeed occur in corpora. We then analyze the human annotations in this dataset in more detail in Section 4. In Section 5, we present the feature model we use to describe the instances, which taps into the lexical semantics and polarity of the constituents. In Section 6, we describe machine learning experiments and classification results that support our hypothesis that the interpretation of this construction largely depends on the semantics of its component verb. In Section 7 we suggest that our results support an analysis of this phenomenon within construction grammar, and point to some future directions in our research in Section 8.

2   Background and our proposal

The *No X is too Y to Z* construction was investigated by Wason and Reich (1979), and discussed more recently by Pullum (2004) and Liberman (2009a,b). Here we highlight some of the most important properties of this complex phenomenon. Our presentation owes much to the lucid discussion and clarification of this topic, and of the work of Wason and Reich specifically, by Liberman.

Wason and Reich argue that the compositional interpretation of sentences of the form of (1) and (2) is "every X Zs". Intuitively, this can be understood by considering a sentence identical to sentence (1), but without a negative subject: *This interest is too narrow to deserve its own newsletter*, which means that "this interest is so narrow that it does not deserve a newsletter". This example indicates that the meaning of *too narrow to deserve its own newsletter* is "so narrow that it does not deserve a newsletter". When this negative "too" assertion is compositionally combined with the *No interest* subject of sentence (1), it results in a meaning with two negatives: "No interest is so narrow that it does not deserve a newsletter", or simply, "Every interest deserves a newslet-

ter". Wason and Reich note that in sentences such as (1), the compositional "every" interpretation is consistent with common beliefs about the world, and thus refer to such sentences as "pragmatic".

By contrast, the compositional interpretation of sentences such as (2) does not correspond to our common sense beliefs. Consider an analogous (non-negative subject) sentence to sentence (2)—i.e., *This item is too minor to escape his attention*. It is nonsensical that "This item is so minor that it does not escape his attention", since being more "minor" entails more likelihood of escaping attention, not less. The compositional interpretation of (2) is similarly nonsensical—i.e., that "No item is so minor that it does not escape his attention"; Such sentences are thus termed "non-pragmatic" by Wason and Reich, who argue that the complexity of the non-pragmatic sentences—arising in part due to the number of negations they contain—causes the listener or reader to misconstrue them. According to their reasoning, listeners choose an interpretation that is consistent with their beliefs about the world—namely that "no X Zs", in this case that "No item escapes his attention"—instead of the compositional interpretation ("Every item escapes his attention").

While Wason and Reich focus on the compositional semantics and pragmatics of these sentences, they also note that the non-pragmatic examples typically use a verb that itself has some aspect of negation, such as *ignore*, *miss*, and *overlook*. This property is also pointed out by Pullum (2004), who notes that *avoid* in his example of the construction means "manage to **not** do" something. Building on this observation, we hypothesize that lexical properties of the component constituents of this construction, particularly the verb, can be important cues to its semantico-pragmatic interpretation. Specifically, we hypothesize that the pragmatic ("every" interpretation) and non-pragmatic ("no" interpretation) sentences will tend to involve verbs with different semantics. Given that verbs of different semantic classes have different selectional preferences, we also expect to see the "every" and "no" sentences associated with semantically different nouns and adjectives.

3   Dataset

3.1   Extraction

To create a dataset of usages of the construction *no NP is too AP to VP*—referred to as the tar-

get construction—we use two corpora: the British National Corpus (Burnard, 2000), an approximately one hundred million word corpus of late-twentieth century British English, and The New York Times Annotated Corpus (Sandhaus, 2008), approximately one billion words of non-newswire text from the New York Times from the years 1987–2006. We extract all sentences in these corpora containing the sequence of strings *no*, *is too*, and *to* separated by one or more words. We then manually filter all sentences that do not have *no NP* as the subject of *is too*, or that do not have *to VP* as an argument of *is too*. After removing duplicates, this results in 170 sentences. We randomly select 20 of these sentences for development data, leaving 150 sentences for testing.

Although we find only 170 examples of the target construction in 1.1 billion words of text, note that our extraction process is quite strict and misses some relevant usages. For example, we do not extract sentences of the form *Nothing is too Y to Z* in which the subject NP does not contain the word *no*. Nor do we extract usages of the related construction *No X is too Y for Z*, where Z is an NP related to a verb, as in *No interest is too narrow for attention*. (We would only extract the latter if there were an infinitive verb embedded in or following the NP.) In the present study we limit our consideration to sentences of the form discussed by Wason and Reich (1979), but intend to consider related constructions such as these—which appear to exhibit the same ambiguity as the target construction—in the future.

We next manually identify the noun, adjective, and verb that participate in the target construction in each sentence. Although this could be done automatically using a parser (e.g., Collins, 2003) or chunker (e.g., Abney, 1991), here we want to ensure error-free identification. We also note a number of sentences containing co-ordination, such as in the following example.

(3) These days, no topic is too recent or specialized to disqualify it from museum apotheosis.

This sentence contains two instances of the target construction: one corresponding to the noun-adjective-verb triple *topic*, *recent*, *disqualify*, and the other to the triple *topic*, *specialized*, *disqualify*. In general, we consider each unique noun-adjective-verb triple participating in the target construction as a separate instance.

3.2 Annotation

We used Amazon Mechanical Turk (AMT, `https://www.mturk.com/`) to obtain judgements as to the correct interpretation of each instance of the target construction in both the development and testing datasets. For each instance, we generated two paraphrases, one corresponding to each of the interpretations discussed in Section 1. We then presented the given instance of the target construction along with its two paraphrases to annotators through AMT, as shown in Table 1. In generating the paraphrases, one of the authors selected the most appropriate paraphrase, in their judgement, where *can* in the paraphrases in Table 1 was selected from *can*, *should*, *will*, and $\emptyset$. Note that the paraphrases do not contain the adjective from the target construction. In the case of multiple instances of the target construction with differing adjectives but the same noun and verb, we only solicited judgements for one instance, and used these judgements for the other instances. In our dataset we observe that all instances obtained from the same sentence which differ only with respect to their noun or verb have the same interpretation. We therefore believe that instances with the same noun and verb but a different adjective are unlikely to differ in their interpretation.

Instructions:

- Read the sentence below.
- Based on your interpretation of that sentence, select the answer that most closely matches your interpretation.
- Select "I don't know" if neither answer is close to your interpretation, or if you are really unsure.

That success was accomplished in large part to tight control on costs , and no cost is too small to be scrutinized .

- Every cost can be scrutinized.
- No cost can be scrutinized.
- I don't know.

Enter any feedback you have about this HIT. We greatly appreciate you taking the time to do so.

Table 1: A sample of the Amazon Mechanical Turk annotation task.

We also allowed the judges to optionally enter any feedback about the annotation task which in some cases—discussed in the following section—was useful in determining whether the judges found a particular instance difficult to annotate.[2]

For each instance of the target construction we obtained three judgements from unique workers on AMT. For approximately 80% of the items, the judgements were unanimous. In the remaining cases we solicited four additional judgements, and used the majority judgement. We paid $0.05 per judgement; the average time spent on each annotation was approximately twenty seconds, resulting in an average hourly wage of about $10.

The development data was also annotated by three native English speaking experts (computational linguists with extensive linguistic background, two of whom are also authors of this paper). The inter-annotator agreement among these judges is very high, with pairwise observed agreements of 1.00, 0.90, and 0.90, and corresponding unweighted Kappa scores of 1.00, 0.79, and 0.79. The majority judgements of these annotators are the same as those obtained from AMT on the development data, giving us confidence in the reliability of the AMT judgements. These findings are consistent with those of Snow et al. (2008) in showing that AMT judgements can be as reliable as those of expert judges.

Finally, we remove a small number of items from the testing dataset which were difficult to paraphrase due to ellipsis of the verb participating in the target construction, or an extra negation in the verb phrase. We further remove one sentence because we believe the paraphrases we provided are in fact misleading. The number of sentences and of instances (i.e., noun-verb-adjective triples) of the target construction in the development and testing datasets is given in Table 2. 160 of the 199 testing instances (80%) have the "every" interpretation, with the remainder having the "no" interpretation.

4 Analysis of annotation

We now more closely examine the annotations obtained from AMT to better determine the extent to

Dataset	# sentences	# instances
Development	20	33
Test	140	199

Table 2: The number of sentences containing the target construction, and the number of resulting instances.

which they are reliable. We also consider specific instances of the target construction that are judged inconsistently to establish some of the causes of disagreement.

One of the three experts who annotated the development items (discussed in Section 3.2) also annotated twenty items selected at random from the testing data. In this case two instances are judged differently than the majority judgement obtained from AMT. These instances are given below with the noun, adjective and verb in the target construction underlined.

(4) When it comes to the clash of candidates on national television, no detail, it seems, is too minor for negotiation, no <u>risk</u> too <u>small</u> to <u>eliminate</u>.

(5) Lectures by big-name Wall Street felons will show why no <u>swindler</u> is too <u>big</u> to <u>beat</u> the rap by peaching on small-timers.

For sentence (4), the AMT judgements were unanimously for the "no" interpretation whereas the expert annotator chose the "every" interpretation. We are uncertain as to the reason for this disagreement, but are convinced that the "every" interpretation is the intended one.

In the case of sentence (5), the AMT judgements were split four–three for the "every" and "no" interpretations, respectively, while the expert annotator chose the "no" interpretation. For this sentence the provided paraphrases were *Every swindler can beat the rap* and *No swindler can beat the rap*. If attention in the sentence is restricted to the target construction—i.e., *no swindler is too big to beat the rap by peaching on small-timers*—either of the "no" and "every" interpretations is possible. That is, this clause alone can mean that "no swindler is 'big' enough to be able to beat the rap" (the "no" interpretation), or that "no swindler is 'big' enough that they

are above peaching on small-timers" (or in other words, "every swindler is able to beat the rap by peaching on small-timers", the "every" interpretation). However, the intention of the sentence as the "no" interpretation is clear from the referral in the main clause to *big-name Wall Street felons*, which implies that "big" swindlers have *not* beaten the rap. Since the AMT annotators may not be devoting a large amount of attention to the task, they may focus only on the target construction and not the preliminary disambiguating material. In this event, they may be choosing between the "every" and "no" interpretations based on how cynical they are of the ability (or lack thereof) of the American legal system to punish Wall Street criminals.

We also examine a small number of examples in the testing set which do not receive a clear majority judgement from AMT. For this analysis we consider items for which the difference in the number of judgements for each of the "every" and the "no" interpretations is one or less This gives four instances of the target construction, one of which we have already discussed above, example (5); the others are presented below, again with the noun, adjective, and verb participating in the target construction underlined:

(6) Where are our priorities when we so carefully weigh costs and medical efficacy in deciding to offer a medical lifeline to the elderly, yet no <u>amount</u> of money is too <u>great</u> to <u>spend</u> on the debatable paths we've taken in our war against terror?

(7) No <u>neighborhood</u> is too <u>remote</u> to <u>diminish</u> Mr. Levine's determination to discover and announce some previously unheralded treat.

(8) No <u>one</u> is too <u>remote</u> anymore to be <u>concerned</u> about style, Ms. Hansen suggested.

In example (6) the author is using the target construction to express somebody else's viewpoint that "any amount should be spent on the war against terror". Therefore the literal reading of the target construction appears to be the "every" interpretation. However, this construction is being used rhetorically (as part of the overall sentence) to express the author's belief that "too much money is being spent on the war against terror", which is close in meaning to the "no" interpretation. It appears that the annotators are split between these two readings. For sentence (7) the

atypicality of *neighbourhood* as the subject of *diminish* may make this instance particularly difficult for the judges. Sentence (8) appears to us to be a clear example of the "every" interpretation. The paraphrases for this usage are "Everyone should be concerned about style" and "No one should be concerned about style". In this case it is possible that the judges are biased by their beliefs about whether one should be concerned about style, and that this is giving rise to the lack of agreement. These examples illustrate that some of these usages are clearly complex for people to annotate. Such complex examples may require more context to be annotated with confidence.

5   Model

To test our hypothesis that the interaction of the semantics of the noun, adjective, and verb in the target construction contributes to its pragmatic interpretation, we represent each instance in our dataset as a vector of features that capture aspects of the semantics of its component words.

WordNet   To tap into general lexical semantic properties of the words in the construction, we use features that draw on the semantic classes of words in WordNet (Fellbaum, 1998). These binary features each represent a synset in WordNet, and are turned on or off for the component words (the noun, adjective, and verb) in each instance of the target construction. A synset feature is on for a word if the synset occurs on the path from all senses of the word to the root, and off otherwise. We use WordNet version 3.0 accessed using NLTK version 2.0 (Bird et al., 2009).

Polarity   Because of the observation that the verb in the target construction, in particular, has some property of negativity in the "no" interpretation, we also use features representing the semantic polarity of the noun, adjective, and verb in each instance. The features are tertiary, representing positive, neutral, or negative polarity. We obtain polarity information from the subjectivity lexicon provided by Wilson et al. (2005), and consider words to be neutral if they have both positive and negative polarity, or are not in the lexicon.

6   Experimental results

6.1   Experimental setup

To evaluate our model we conduct a 5-fold cross-validation experiment using the items in the test-

ing dataset. When partitioning the items in the testing dataset into the five parts necessary for the cross-validation experiment, we ensure that all the instances of the target construction from a single sentence are in the same part. This ensures that no instance used for training is from the same sentence as an instance used for testing. We further ensure that the proportion of items in each class is roughly the same in each split.

For each of the five runs, we linearly scale the training data to be in the range $[-1, 1]$, and apply the same transformation to the testing data. We train a support vector machine (LIBSVM version 2.9, Chang and Lin, 2001) with a radial basis function kernel on the training portion in each run, setting the cost and gamma parameters using cross-validation on just the training portion, and then test the classifier on the testing portion for that run using the same parameter settings. We micro-average the accuracy obtained on each of the five runs. Finally, we repeat each 5-fold cross-validation experiment five times, with five random splits, and report the average accuracy over these trials.

6.2 Results

Results for experiments using various subsets of the features are presented in Table 3. We restrict the component word—the noun, adjective, or verb—for which we extract features to those listed in column "Word", and extract only the features given in column "Features" (WordNet, polarity, or all). The majority baseline is 80%, corresponding to always selecting the "every" interpretation. Accuracies shown in boldface are significantly better than the majority class baseline using a paired t-test. (In all cases where the difference is significant, we obtain $p \ll 0.01$.)

We first consider the results using features extracted only for the noun, adjective, or verb individually, using all features. The best accuracy in this group of experiments, 87%, is achieved using the verb features, and is significantly higher than the majority baseline. On the other hand, the classifiers trained on the noun and adjective features individually perform no better than the baseline. These results support our hypothesis that lexical semantic properties of the component verb in the *No X is too Y to Z* construction do indeed play an important role in determining its interpretation. Although we proposed that selectional constraints

from the verb would also lead to differing semantics of the nouns and adjectives in the two interpretations, our WordNet features are likely too simplistic to capture this effect, if it does hold. Before ruling out the semantic contribution of these words to the interpretation, we need to explore whether a more sophisticated model of selectional preferences, as in Ciaramita and Johnson (2000) or Clark and Weir (2002), yields more informative features for the noun and adjective.

Experimental setup		% accuracy
Word	Features	
Noun	All	80
Adjective	All	80
Verb	All	**87**
All	WordNet	**88**
All	Polarity	80
All	All	**88**
Majority baseline		80

Table 3: % accuracy on testing data for each experimental condition and the majority baseline. Accuracies in boldface are statistically significantly different from the baseline.

We now consider the results using the WordNet and polarity features individually, but extracted for all three component words. The WordNet features perform as well as the best results using all features for all three words, which gives further support to our hypothesis that the semantics of the components of the target construction are related to its interpretation. The polarity features perform poorly. This is perhaps unsurprising as polarity is a poor approximation to the property of "negativity" that we are attempting to capture. Moreover, many of the nouns, adjectives, and verbs in our dataset either have neutral polarity or are not in the polarity lexicon, and therefore the polarity features are not very discriminative. In future work, we plan to examine the WordNet classes of the verbs that occur in the "no" interpretation to try to more precisely characterize the property of negativity that these verbs tend to have.

6.3 Error analysis

To better understand the errors our classifier is making, we examine the specific instances which are classified incorrectly. Here we focus on the experiment using all features for all three component words. There are 23 instances which are

consistently mis-classified in all runs of the experiment. According to the AMT judgements, each of these instances corresponds to the "no" interpretation. These errors reflect the bias of the classifier towards the more frequent class, the "every" interpretation.

We further note that two of the instances discussed in Section 4—examples (4) and (6)—are among those instances consistently classified incorrectly. The majority judgement from AMT for both of these instances is the "no" interpretation, while in our assessment they are in fact the "every" interpretation. We are therefore not surprised to see these items "mis-classified" as "every".

Example (8) was incorrectly classified in one trial. In this case we agree with the gold-standard label obtained from AMT in judging this instance as the "every" interpretation; nevertheless, this does appear to be a difficult instance given the low agreement observed for the AMT judgements.

It is interesting that no items with an "every" interpretation are consistently misclassified. In the context of our overall results showing the impact of the verb features on performance, we conclude that the "no" interpretation arises due to particular lexical semantic properties of certain verbs. We suspect then that the consistent errors on the 21 truly misclassified expressions (23 minus the 2 instances discussed above that we believe to be annotated incorrectly) are due to sparse data. That is, if it is indeed the verb that plays a major role in leading to a "no" interpretation, there may simply be insufficient numbers of such verbs for training a supervised model in a dataset with only 39 examples of those usages.

7 Discussion

We have presented the first computational study of the semantically and pragmatically complex construction *No X is too Y to Z*. We have developed a computational model that automatically disambiguates the construction with an accuracy of 88%, reducing the error-rate over the majority-baseline by 40%. The model uses features that tap into the lexical semantics of the component words participating in the construction, particularly the verb. These results demonstrate that lexical properties can be successful in resolving an ambiguity previously thought to depend on complex pragmatic inference over presuppositions (as in Wason and Reich (1979)).

These results can be usefully situated within the context of linguistic and psycholinguistic work on semantic interpretation processing. Beginning around 20 years ago, work in modeling of human semantic preferences has focused on the extent to which properties of lexical items influence the interpretation of various linguistic ambiguities (e.g., Trueswell and Tanenhaus, 1994). While semantic context and plausibility are also proposed to play a role in human interpretation of ambiguous sentences (e.g., Crain and Steedman, 1985; Altmann and Steedman, 1988), it has been pointed out that it would be difficult to "operationalize" the complex interactions of presuppositional factors with real-world knowledge in a precise algorithm for disambiguation (Jurafsky, 1996). Although not intended as proposing a cognitive model, the work here can be seen as connected to these lines of research, in investigating the extent to which lexical factors can be used as proxies to more "hidden" features that underlie the appropriate interpretation of a pragmatically complex construction.

Moreover, as in the approach of Jurafsky (1996), the phenomenon we investigate here may be best considered within a constructional analysis (e.g., Langacker, 1987), in which both the syntactic construction and the particular lexical items contribute to the determination of the meaning of a usage. We suggest that a clause of the form *No X is too Y to Z* might be the (identical) surface expression of *two* underlying constructions—one with the "every" interpretation and one with the "no" interpretation—which place differing constraints on the semantics of the verb. (E.g., in the "no" interpretation, the verb typically has some "negative" semantic property, as noted in Section 2.) Looked at from the other perspective, the lexical semantic properties of the verb might determine which *No X is too Y to Z* construction (and associated interpretation) it is compatible with. Our results support this view, by showing that semantic classes of verbs have predictive value in selecting the correct interpretation.

Note that such a constructional analysis of this phenomenon assumes that both interpretations of these sentences are linguistically valid, given the appropriate lexical instantiation. This stands in contrast to the analysis of Wason and Reich (1979), which presumes that people are applying some higher-level reasoning to "correct" an ill-formed statement in the case of the "no" in-

terpretation. While such extra-grammatical inference may play a role in support of language understanding when people are faced with noisy data, it seems unlikely to us that a construction that is used quite readily and with a predictable interpretation is nonsensical according to rules of grammar. Our results point to an alternative linguistic analysis, one whose further development may also help to improve automatic disambiguation of instances of *No X is too Y to Z*. In the next section, we discuss directions for future work that could elaborate on these preliminary findings.

8 Future Work

One limitation of this study is that the dataset used is rather small, consisting of just 199 instances of the target construction. As discussed in Section 3.1, the extraction process we use to obtain our experimental items has low recall; in particular it misses variants of the target construction such as *Nothing is too Y to Z* and *No X is too Y for Z*. In the future we intend to expand our dataset by extracting such usages. Furthermore, the data used in the present study is primarily taken from news text. While we do not adopt the view of some that usages of the target construction having the "no" interpretation are errors, it could be the case that such usages are more frequent in less formal text. In the future we also intend to extract usages of the target construction from datasets of less formal text, such as blogs (e.g., Burton et al., 2009).

Constructions other than *No X is too Y to Z* exhibit a similar ambiguity. For example, the construction *X didn't wait to Y* is ambiguous between "X did Y right away" and "X didn't do Y at all" (Karttunen, 2007). In the future we would like to extend our study to consider more such constructions which are ambiguous due to the interpretation of negation.

In Section 4 we note that for some instances the complexity of the sentences containing the target construction may make it difficult for the annotators to judge the meaning of the target. In the future we intend to present simplified versions of these sentences—which retain the noun, adjective, and verb from the target construction in the original sentence—to the judges to avoid this issue. Such an approach will also help us to focus more clearly on observable lexical semantic effects.

We are particularly interested in further exploring the hypothesis that it is the semantics of the component verb that gives rise to the meaning of the target construction. Recall Pullum's (2004) observation that the verb in the "no" interpretation involves explicitly *not* acting. Using this intuition, we have informally observed that it is largely possible to (manually) predict the interpretation of the target construction knowing only the component verb. We are interested in establishing the extent to which this observation holds, and precisely which aspects of a verb's meaning give rise to the interpretation of the target construction.

Our current model of the semantics of the target construction does not capture Wason and Reich's (1979) observation that the compositional meaning of instances having the "no" interpretation is non-pragmatic. While we do not adopt their view that these usages are somehow "errors", we do think that their observation can indicate other possible lexical semantic properties that may help to identify the correct interpretation. Taking the classic example from Wason and Reich, *no head injury is too trivial to ignore*, one clue to the "no" interpretation is that generally a head injury is not something that is ignored. On the other hand, considering Wason and Reich's example *no missile is too small to ban*, it is widely believed that missiles should be banned. We would like to add features that capture this knowledge to our model.

In preliminary experiments we have used co-occurrence information as an approximation to this knowledge. (For example, we would expect that *head injury* would tend to co-occur less with *ignore* than with antonymous verbs such as *treat* or *address*.) Although our early results using co-occurrence features do not indicate that they are an improvement over the other features considered (WordNet and polarity), it may also be the case that our present formulation of these co-occurrence features does not effectively capture the intended knowledge. In the future we plan to further consider such features, especially those that model the selectional preferences of the verb participating in the target construction.

These several strands of future work—increasing the size of the dataset, improving the quality of annotation, and exploring additional features in our computational model—will enable us to extend our linguistic analysis of this interesting phenomenon, as well as to improve performance on automatic disambiguation of this complex construction.

References

Steven Abney. 1991. Parsing by chunks. In Robert Berwick, Steven Abney, and Carol Tenny, editors, *Principle-Based Parsing: Computation and Psycholinguistics*, pages 257–278. Kluwer Academic Publishers.

Gerry T. M. Altmann and Mark Steedman. 1988. Interaction with context during human sentence processing. *Cognition*, 30(3):191–238.

Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*. O'Reilly Media Inc.

Lou Burnard. 2000. *The British National Corpus Users Reference Guide*. Oxford University Computing Services.

Kevin Burton, Akshay Java, and Ian Soboroff. 2009. The ICWSM 2009 Spinn3r Dataset. In *Proceedings of the third Annual Conference on Weblogs and Social Media (ICWSM 2009)*. San Jose, CA.

Chih-Chung Chang and Chih-Jen Lin. 2001. *LIBSVM: a library for support vector machines*. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Massimiliano Ciaramita and Mark Johnson. 2000. Explaining away ambiguity: Learning verb selectional preference with Bayesian networks. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, pages 187–193. Saarbrücken, Germany.

Stephen Clark and David Weir. 2002. Class-based probability estimation using a semantic hierarchy. *Computational Linguistics*, 28(2):187–206.

Michael Collins. 2003. Head-driven statistical models for natural language parsing. *Computational Linguistics*, 29(4):589–637.

Stephen Crain and Mark Steedman. 1985. On not being led up the garden path: The use of context by the psychological syntax processor. In David R. Dowty, Lauri Karttunen, and Arnold M. Zwicky, editors, *Natural language parsing: Psychological, computational, and theoretical perspectives*, pages 320–358. Cambridge University Press, Cambridge.

James Curran, Stephen Clark, and Johan Bos. 2007. Linguistically motivated large-scale NLP with C&C and Boxer. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 33–36. Prague, Czech Republic.

Christiane Fellbaum, editor. 1998. *Wordnet: An Electronic Lexical Database*. Bradford Books.

Daniel Jurafsky. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20(2):137–194.

Lauri Karttunen. 2007. Word play. *Computational Linguistics*, 33(4):443–467.

Ronald W. Langacker. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*, volume 1. Stanford University Press, Stanford.

Mark Liberman. 2009a. No detail too small. Retrieved 9 February 2010 from `http://languagelog.ldc.upenn.edu/nll/`.

Mark Liberman. 2009b. No wug is too dax to be zonged. Retrieved 9 February 2010 from `http://languagelog.ldc.upenn.edu/nll/`.

Geoffrey K. Pullum. 2004. Too complex to avoid judgment? Retrieved 7 April 2010 from `http://itre.cis.upenn.edu/~myl/languagelog/`.

Evan Sandhaus. 2008. The New York Times Annotated Corpus. Linguistic Data Consortium, Philadelphia, PA.

Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Y. Ng. 2008. Cheap and fast — But is it good? Evaluating non-expert annotations for natural language tasks. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, pages 254–263. Honolulu, HI.

John Trueswell and Michael J. Tanenhaus. 1994. Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In Charles Clifton, Lyn Frazier, and Keith Rayner, editors, *Perspectives on Sentence Processing*, pages 155–179. Lawrence Erlbaum, Hillsdale, NJ.

Peter Wason and Shuli Reich. 1979. A verbal illusion. *The Quarterly Journal of Experimental Psychology*, 31(4):591–597.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, pages 347–354. Vancouver, Canada.