# Learning to Identify Reduced Passive Verb Phrases with a Shallow Parser

**Sean Igo**
School of Computing
University of Utah
Salt Lake City, UT 84112
samwibatt@gmail.com

**Ellen Riloff**
School of Computing
University of Utah
Salt Lake City, UT 84112
riloff@cs.utah.edu

## Abstract

Our research is motivated by the observation that NLP systems frequently mislabel passive voice verb phrases as being in the active voice when there is no auxiliary verb (e.g., *"The man arrested had a long record"*). These errors directly impact thematic role recognition and NLP applications that depend on it. We present a learned classifier that can accurately identify *reduced passive voice* constructions in shallow parsing environments.

## Introduction

Natural language processing systems need to distinguish between active and passive voice to correctly identify the thematic roles associated with verb phrases (VPs). In the passive voice, verbs typically occur in the past participle form preceded by a "to be" auxiliary verb.[1] For example, *"YouTube was purchased by Google"* is a passive voice construction which means that YouTube was the purchasee (theme) and Google was the purchaser (agent). In contrast, *"YouTube purchased Google"* is an active voice construction which reverses the thematic roles, meaning that YouTube is the purchaser (agent) and Google is the purchasee (theme). As is often the case, both the agent and theme belong to the same semantic class (in this case, COMPANY) so determining whether the VP is in active or passive voice is critical to correctly assign thematic roles.

Our research is motivated by a recurring problem that we have observed: passive voice VPs are often mislabeled as active voice when the auxiliary verb is missing. For example, consider the following sentences, where the underlined verbs are in the passive voice:

(a) *"A dead bird infected with the West Nile Virus has been discovered in Maine."*
(b) *"The ringleader was John Freire, an attorney arrested in March 1989."*

The verb "infected" in sentence (a) is in the passive voice because its syntactic subject, *"a dead bird"*, is the entity that was infected (its theme). Similarly, the verb "arrested"

in sentence (b) is in the passive voice because its syntactic subject, *"an attorney"*, is the entity that was arrested. We will refer to these types of constructions as *reduced passives*, because they typically occur in reduced relative clauses.

Ideally, NLP systems should be able to accurately identify reduced passives by generating parse trees that capture the corresponding syntactic structures. However, we will show that even full parsers have trouble consistently identifying reduced passives across domains. But more importantly, many NLP applications rely on shallow parsers because of their speed and robustness (Li and Roth 2001), and shallow parsers do not generate the deep syntactic structures that are needed to explicitly capture reduced passive constructions.

We present a novel approach that aims to create a classifier to identify reduced passive voice VPs in shallow parsing environments.

## Motivation and Related Work

Our research was motivated by the observation that our shallow parser, Sundance (Riloff and Phillips 2004), consistently mislabeled reduced passives as being in the active voice. These errors led to mistakes in downstream application systems that rely on thematic role labels. In fact, Sundance is unusual in that it generates active/passive voice labels at all. Most shallow parsers perform syntactic "chunking" (NP, VP, etc.) but do not perform any syntactic analysis beyond that. Sundance uses heuristic rules to distinguish active voice VPs from passive voice VPs based on part-of-speech tags and the presence of auxiliary verbs. But reduced passives look almost identical to active voice VPs based on these properties. We are not aware of research that has focused specifically on identifying reduced passive voice constructions, but the problem has been previously acknowledged and discussed (Merlo and Stevenson 1998). Before embarking on a solution, we wanted to know: (1) how common are reduced passives?, and (2) how well do full parsers identify them?

To answer these questions, we manually identified the reduced passive voice constructions in texts drawn from three corpora: 50 Penn Treebank (WSJ) texts (Bies et al. 1995), 200 MUC4 terrorism articles (MUC-4 Proceedings 1992), and 100 ProMed disease outbreak texts (ProMedmail 2006). Reduced passives occurred in 8.4% of the Treebank sentences, 11.6% of the MUC4 sentences, and 13.3% of the ProMed sentences. On average, they occurred in 1 of

[1]Some other auxiliary verbs can also signal passive voice, such as *"get"* (e.g., *"he got shot"*) and *"become"* (e.g., *"she became stranded"*) (Carter and McCarthy 1999; McEnery and Xiao 2005).

every 9 sentences. We concluded that this phenomenon was common enough to warrant further investigation.

We then evaluated the ability of 3 widely used full parsers to recognize passive voice constructions: Charniak's parser (Charniak 2000), Collins' parser (Collins 1999), and MINI-PAR (Lin 1998). The Charniak and Collins parsers generate Treebank-style parse trees, which do not explicitly assign voice labels. However, the Treebank grammar (Bies et al. 1995) includes phrase structures that capture passive voice constructions, and Treebank parsers should generate these structures when passives occur. So we wrote a program that navigates a Treebank-style parse tree and labels a verb as an ordinary or reduced passive if one of the corresponding phrase structures is found.[2] To measure the effectiveness of our program, we compared its output against the manually identified passive verbs in our gold standard. The program identified the ordinary passives with 98% recall and 98% precision, and the reduced passives with 93% recall and 93% precision. Upon inspection, we attributed the discrepancies to highly unusual phrases and felt satisfied that the program faithfully reflects the passive voice labels that would be assigned by Treebank-style parsers using automated methods.

MINIPAR generates dependency relations, one of which is the head-modifier relationship *vrel*, or "passive verb modifier of nouns", which corresponds to the most common type of reduced passive. We considered all verbs with *vrel* tags to be reduced passives. MINIPAR does not seem to have a relation that directly corresponds to ordinary passive voice constructions, so we did not evaluate MINIPAR on these.

| | WSJ | | | MUC4 | | | PRO | | |
|---|---|---|---|---|---|---|---|---|---|
| | R | P | F | R | P | F | R | P | F |
| *Ordinary Passives* | | | | | | | | | |
| **CH** | .95 | .96 | .95 | .91 | .96 | .93 | .93 | .96 | .94 |
| **CO** | .95 | .95 | .95 | .91 | .94 | .92 | .93 | .94 | .93 |
| *Reduced Passives* | | | | | | | | | |
| **CH** | .90 | .88 | .89 | .77 | .71 | .74 | .77 | .78 | .77 |
| **CO** | .85 | .89 | .87 | .66 | .67 | .66 | .64 | .78 | .70 |
| **MP** | .48 | .57 | .52 | .51 | .72 | .60 | .44 | .68 | .53 |

Table 1: Performance of full parsers

Table 1 shows the performance of the parsers[3] on ordinary and reduced passives. The Charniak (**CH**) and Collins (**CO**) parsers performed well on ordinary passives in all three domains. For reduced passives, these parsers performed well on WSJ texts, but performance dropped considerably on the MUC4 and ProMed corpora. MINIPAR (**MP**) achieved only moderate recall and precision across the board.

The performance of the Charniak and Collins parsers was considerably lower on reduced passives than on ordinary passives. This supports our hypothesis that reduced passives are more difficult to recognize. Second, their per-

formance on reduced passives was substantially lower on MUC4 and ProMed texts than on WSJ texts. Since there was no comparable drop across domains for ordinary passives, reduced passive recognition may be more susceptible to domain portability problems in statistical parsers.[4]

Our goal was to develop a reduced passive voice recognizer that can be used with shallow parsers, which are widely used because they are typically much faster and more robust than full parsers. To illustrate the speed differential, we processed 1,718 sentences from ProMed using our shallow parser, Sundance, as well as the Stanford parser (Klein and Manning 2003), which claims to be a fast full parser. The Stanford parser took 48 minutes, while Sundance took 20 seconds.[5] This speed differential is similar to what we have observed with other full parsers as well. For applications that must process large volumes of text, this speed differential can make the use of full parsers impractical, or at least undesirable. Li and Roth (2001) have also demonstrated that shallow parsers are more robust on lower-quality texts.

## Reduced Passive Classification

The goal of our research is to create a classifier that can identify reduced passive VPs in shallow parsing environments. The features rely on information provided by the Sundance shallow parser, which does part-of-speech tagging and syntactic chunking, as well as syntactic role assignment (subject, direct object, indirect object) and simple clause segmentation. It also assigns general semantic classes (e.g., ANIMAL, DISEASE, etc.) to words via dictionary look-up. In this section, we first describe a test that identifies *Reduced Passive Candidates*. Next, we present features that we hypothesized could be useful for identifying reduced passives. Finally, we describe the machine learning classifiers.

### Reduced Passive Candidates

The first step is to identify verbs that potentially can be reduced passives. We want to rule out verbs that are obviously in active voice or ordinary passive voice constructions. We consider a verb to be a *reduced passive candidate* if it satisfies three criteria: (1) it is in past tense[6] , (2) it is not itself an auxiliary verb ("be", "have", etc.), (3) the VP containing the verb does not contain passive auxiliaries, perfective auxiliaries ("have"), "do", or modals. We define the set of passive auxiliaries as: {"be", "become", "appear", "feel","get", "remain", "seem"}. Only the verbs that satisfy these criteria are used to create training/test instances.

### Features

We created 28 features to identify reduced passives.

---

*Lexical (1):* One feature is the root of the verb.

*Syntactic (7):* 4 binary features represent the presence of 4 syntactic constituents around the verb (subject, direct object, indirect object, and "by" PP); 2 features represent the lexical heads of the subject and the "by" PP; 1 binary feature indicates whether the subject is a nominative pronoun.

*Part of Speech (5):* 3 binary features indicate whether the verb is (a) followed by a verb, auxiliary, or modal POS tag, (b) followed by a preposition, or (c) preceded by a number; 2 features represent the POS tag of the preceding word and the following word.

*Semantic (4):* 4 features represent the semantic class of the head of the verb's subject and the semantic class of the head of a following "by" PP (if one exists). We define 2 features for each case: one with the most general matching semantic class, and one with the most specific matching semantic class, based on Sundance's semantic hierarchy.

*Clausal (6):* 2 binary features indicate the presence of multiple clauses or multiple verb phrases, and 4 binary features represent the context surrounding the current VP: is it followed by an infinitive, a new clause, the end of the sentence, or is it the last VP in the sentence?

### TRANSITIVITY

If a transitive verb occurs in the past tense without a passive auxiliary or a direct object, then it is tempting to conclude that it is a reduced passive. However, transitive verbs often appear without a direct object (e.g., *"He ate after the concert"*).[7] Consequently, simply knowing that a verb is transitive is not sufficient. We hypothesized that it may be useful to empirically determine the *transitiveness* of a verb based on the intuition that some transitive verbs almost always occur with an object, while others occur in mixed settings. If a verb nearly always occurs with an object, then the absence of one is more significant.

We estimate the *transitiveness* of a verb empirically from an unannotated corpus. If a VP does not contain any passive auxiliaries and has both a subject and a direct object, then we assume it is a transitive active voice construction. If a VP is past tense, has a passive auxiliary, and has a subject but no direct object, then we assume it is an ordinary passive. All other instances of the verb are considered to be intransitive. These heuristics are obviously flawed (e.g., all reduced passives are counted as intransitive!). But the heuristics were designed to be a lower bound on a verb's true transitiveness. We then compute **transitiveness** as $P(transitive \mid verb)$ and create the following feature:

*Transitivity (1):* $P(transitive \mid verb)$ is mapped to 6 values based on ranges: $< .20$, .20-.40, .40-.60, .60-.80, $>$ .80, or unknown (verb was not seen in the training corpus).

### THEMATIC ROLE SEMANTICS

Conceptually, recognizing the passive voice is about recognizing that a verb's syntactic subject is its theme, or that a "by" PP contains its agent. Therefore it should be beneficial to know the semantic classes that each verb expects as its theme and agent. To capture this, we derived empirical estimates of the thematic role semantics for verbs.

(a) For each verb, we collected its agents and themes. For transitive active voice constructions, we assumed that the verb's subject is its agent, and its direct object is its theme. For ordinary passives, we assumed that the verb's subject is its theme and (if present) a "by" PP contains its agent.[8]

(b) For each verb, we collected **thematic role semantics** from its agents and themes. We looked up the semantic class of the heads in Sundance's dictionary, and computed $P(semclass \mid agent, verb)$ and $P(semclass \mid theme, verb)$. Using this data, we defined 4 features:

*Plausible Theme (1):* Given a verb and its subject, we assess whether the subject's head is a semantically plausible theme for the verb. Our goal is to determine which semantic classes virtually never occur in a thematic role (e.g., if a verb's subject has a semantic class that has never been seen as a theme, then it is probably not a reduced passive). This feature gets a value of 1 if $P(semclass \mid theme, verb) > .01$, otherwise its value is 0 (which means it was either never seen as a theme, or the few instances were probably noise[9]).

*Plausible Agent ByPP (1):* Given a verb and following "by" PP, we assess whether the head noun of the PP is a semantically plausible agent for the verb using the analogous criteria. If the "by" PP is not a plausible agent, then this verb instance is probably not a reduced passive.

*Plausible Agent Subj (1):* Given a verb and its subject, we assess whether the subject is a semantically plausible agent for the verb, using the same criteria.

*Frequency (1):* This feature takes 4 possible values based on the frequency of the verb's root. We use 4 bins representing a logarithmic scale: 0, 1-10, 11-100, and 100+.

### The Classifier

We created classifiers using decision trees and support vector machines. For the **Dtree** classifiers, we used the Weka J48 decision tree learner (Witten and Frank 2005) with the default settings. For the **SVM** classifiers, we used the SVM$^{light}$ package (Joachims 1999) with the default settings. We built one SVM classifier with a linear kernel, **LSVM**, and one with a degree 3 polynomial kernel, **PSVM**.

## Experimental Results

We conducted two sets of experiments. First, we performed 10-fold cross-validation using 2,069 Wall Street Journal (WSJ) texts from the Penn Treebank. The gold standard answers were the reduced passive voice labels automatically derived from the Penn Treebank parse trees by the program mentioned in the Motivation section. Second, we performed experiments on 3 corpora using manually annotated test sets as the gold standards. To create the test sets, we randomly selected 50 WSJ articles from the Penn Treebank, 200 terrorism articles from the MUC4 corpus, and 100 disease outbreak stories from ProMed. The training set for these experiments consisted of the 2,069 Penn Treebank documents

---

[7]One could argue that there is an implicitly understood direct object, i.e., *"He ate [food] after the concert."*.

[8]With the exception of dates and locations.

[9]For example: parser error, incorrect semantic label, etc.

| Method | XVAL | | | WSJ | | | MUC4 | | | PRO | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rec | Prec | F | Rec | Prec | F | Rec | Prec | F | Rec | Prec | F |
| *Candidacy* | **.86** | .14 | .24 | **.86** | .15 | .25 | **.85** | .17 | .28 | **.86** | .25 | .38 |
| *+MC+NoDO* | .77 | .23 | .36 | .74 | .26 | .38 | .77 | .25 | .38 | .78 | .36 | .50 |
| *+ByPP* | .13 | .68 | .21 | .13 | .69 | .22 | .09 | .75 | .16 | .10 | .90 | .17 |
| *+All* | .12 | .82 | .20 | .12 | .81 | .22 | .09 | **.80** | .16 | .09 | **.93** | .16 |
| *Dtree* | .53 | .82 | .64 | .49 | .81 | .61 | .52 | .78 | .62 | .43 | .81 | .56 |
| *LSVM* | .62 | .82 | .71 | .58 | .79 | .67 | .61 | .76 | .67 | .58 | .81 | **.68** |
| *PSVM* | .65 | **.84** | **.73** | .60 | **.82** | **.69** | .60 | **.80** | **.69** | .54 | .82 | .65 |

Table 2: Recall, precision, and F-measure results for reduced passive voice recognition.

used in the cross-validation experiments.[10] Our WSJ test set did not overlap with the documents in this training set. We report recall and precision for the cross-validation experiments on Treebank documents (XVAL) as well as the separate WSJ, MUC4, and ProMed (PRO) test sets.

We devised 4 baselines to determine how well simple rules can identify reduced passives:

*Candidacy*: Verbs that are past tense, have no modals, no passive, perfect, or "do" auxiliary, and are not themselves auxiliary verbs are labeled as reduced passives. (These are the Reduced Passive Candidate criteria described earlier.)

*+MC+NoDO*: In addition to the *Candidacy* rules, the verb must occur in a multi-clausal sentence and cannot have a direct object.

*+ByPP*: In addition to the *Candidacy* rules, the verb must be followed by a PP with preposition "by".

*+All*: All of the conditions above must apply.

Table 2 shows that the baseline systems can achieve high recall or high precision, but not both. Consequently, these simple heuristics are not a viable solution for shallow parsers.

The last 3 rows of Table 2 show the performance of our classifiers, with the best score for each column in boldface. For the XVAL, WSJ, and MUC4 data sets, the best classifier is the PSVM which achieves 80-84% precision with 60-65% recall. For ProMed, the LSVM performs best, achieving 81% precision with 58% recall. These scores are all substantially above the baseline methods.

We also conducted experiments to determine which features had the most impact. Performance dropped substantially when then lexical, POS, or transitivity features were removed. We also investigated whether any group of features performed well on their own. The POS features performed best but still well below the full feature set.

## Conclusions

Currently, NLP systems that use shallow parsers either consistently mislabel reduced passives, or they rely on heuristics similar to our baselines. As an alternative, we have presented a learned classifier that can accurately recognize reduced passives in shallow parsing environments. This classifier should be beneficial for NLP systems that use shallow parsing but need to identify thematic roles.

---

[10]See http://www.xmission.com/~sgigo/aaai08/rp.html for the document ids used in all of our experiments.

## References

Bies, A.; Ferguson, M.; Katz, K.; and MacIntyre, R. 1995. Bracketing Guidelines for Treebank II Style Penn Treebank Project. Technical report, University of Pennsylvania.

Carter, R., and McCarthy, M. 1999. The English get-passive in spoken discourse: description and implications for an interpersonal grammar. *English Language and Linguistics* 3.1:41–58.

Charniak, E. 2000. A Maximum-Entropy-Inspired-Parser. In *Proceedings of the The First Meeting of the North American Chapter of the Association for Computational Linguistics*.

Collins, M. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. Dissertation, University of Pennsylvania.

Joachims, T. 1999. Making Large-Scale Support Vector Machine Learning Practical. In B. Schölkopf, C. Burges, A. S., ed., *Advances in Kernel Methods: Support Vector Machines*. MIT Press.

Klein, D., and Manning, C. 2003. Fast Exact Inference with a Factored Model for Natural Language Parsing. In *Advances in Neural Information Processing Systems 15*.

Li, X., and Roth, D. 2001. Exploring evidence for shallow parsing. In *Proc. of the Annual Conference on Computational Natural Language Learning (CoNLL)*.

Lin, D. 1998. Dependency-based Evaluation of MINIPAR. In *Workshop on the Evaluation of Parsing Systems, Granada, Spain*.

McEnery, A., and Xiao, R. 2005. Passive constructions in English and Chinese: A corpus-based contrastive study. *The Corpus Linguistics Conference Series* 1(1).

Merlo, P., and Stevenson, S. 1998. What grammars tell us about corpora: the case of reduced relative clauses. In *Proceedings of the Sixth Workshop on Very Large Corpora*.

MUC-4 Proceedings. 1992. *Proceedings of the Fourth Message Understanding Conference (MUC-4)*. Morgan Kaufmann.

ProMed-mail. 2006. *http://www.promedmail.org/*.

Riloff, E., and Phillips, W. 2004. An Introduction to the Sundance and AutoSlog Systems. Technical Report UUCS-04-015, School of Computing, University of Utah.

Witten, I., and Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition*. Morgan Kaufmann.