Automatic Interpretation of the English Possessive

Stephen Tratz †
Army Research Laboratory
Adelphi Laboratory Center
2800 Powder Mill Road
Adelphi, MD 20783
stephen.c.tratz.civ@mail.mil

Eduard Hovy †
Carnegie Mellon University
Language Technologies Institute
5000 Forbes Avenue
Pittsburgh, PA 15213
hovy@cmu.edu

Abstract

The English *'s* possessive construction occurs frequently in text and can encode several different semantic relations; however, it has received limited attention from the computational linguistics community. This paper describes the creation of a semantic relation inventory covering the use of *'s*, an inter-annotator agreement study to calculate how well humans can agree on the relations, a large collection of possessives annotated according to the relations, and an accurate automatic annotation system for labeling new examples. Our 21,938 example dataset is by far the largest annotated possessives dataset we are aware of, and both our automatic classification system, which achieves 87.4% accuracy in our classification experiment, and our annotation data are publicly available.

1 Introduction

The English *'s* possessive construction occurs frequently in text—approximately 1.8 times for every 100 hundred words in the Penn Treebank[1] (Marcus et al., 1993)—and can encode a number of different semantic relations including ownership (*John's car*), part-of-whole (*John's arm*), extent (*6 hours' drive*), and location (*America's rivers*). Accurate automatic possessive interpretation could aid many natural language processing (NLP) applications, especially those that build semantic representations for text understanding, text generation, question answering, or information extraction. These interpretations could be valuable for machine translation to or from languages that allow different semantic relations to be encoded by the possessive/genitive.

This paper presents an inventory of 17 semantic relations expressed by the English *'s*-construction, a large dataset annotated according to the this inventory, and an accurate automatic classification system. The final inter-annotator agreement study achieved a strong level of agreement, 0.78 Fleiss' Kappa (Fleiss, 1971) and the dataset is easily the largest manually annotated dataset of possessive constructions created to date. We show that our automatic classication system is highly accurate, achieving 87.4% accuracy on a held-out test set.

2 Background

Although the linguistics field has devoted significant effort to the English possessive (§6.1), the computational linguistics community has given it limited attention. By far the most notable exception to this is the line of work by Moldovan and Badulescu (Moldovan and Badulescu, 2005; Badulescu and Moldovan, 2009), who define a taxonomy of relations, annotate data, calculate inter-annotator agreement, and perform automatic classification experiments. Badulescu and Moldovan (2009) investigate both *'s*-constructions and *of* constructions in the same context using a list of 36 semantic relations (including OTHER). They take their examples from a collection of 20,000 randomly selected sentences from Los Angeles Times news articles used in TREC-9. For the 960 extracted *'s*-possessive examples, only 20 of their semantic relations are observed, including OTHER, with 8 of the observed relations occurring fewer than 10 times. They report a 0.82 Kappa agreement (Siegel and Castellan, 1988) for the two computational semantics graduates who annotate the data, stating that this strong result "can be explained by the instructions the annotators received

†The authors were affiliated with the USC Information Sciences Institute at the time this work was performed.

[1] Possessive pronouns such as *his* and *their* are treated as *'s* constructions in this work.

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 372–381,
Sofia, Bulgaria, August 4-9 2013.

prior to annotation and by their expertise in Lexical Semantics."

Moldovan and Badulescu experiment with several different classification techniques. They find that their *semantic scattering* technique significantly outperforms their comparison systems with its F-measure score of 78.75. Their SVM system performs the worst with only 23.25% accuracy—suprisingly low, especially considering that 220 of the 960 *'s* examples have the same label.

Unfortunately, Badulescu and Moldovan (2009) have not publicly released their data[2]. Also, it is sometimes difficult to understand the meaning of the semantic relations, partly because most relations are only described by a single example and, to a lesser extent, because the bulk of the given examples are *of*-constructions. For example, why *President of Bolivia* warrants a SOURCE/FROM relation but *University of Texas* is assigned to LOCATION/SPACE is unclear. Their relations and provided examples are presented below in Table 1.

Relation	Examples
POSSESSION	Mary's book
KINSHIP	Mary's brother
PROPERTY	John's coldness
AGENT	investigation of the crew
TEMPORAL	last year's exhibition
DEPICTION-DEPICTED	a picture of my niece
PART-WHOLE	the girl's mouth
CAUSE	death of cancer
MAKE/PRODUCE	maker of computer
LOCATION/SPACE	Univerity of Texas
SOURCE/FROM	President of Bolivia
TOPIC	museum of art
ACCOMPANIMENT	solution of the problem
EXPERIENCER	victim of lung disease
RECIPIENT	Josephine's reward
ASSOCIATED WITH	contractors of shipyard
MEASURE	hundred (sp?) of dollars
THEME	acquisition of the holding
RESULT	result of the review
OTHER	state of emergency

Table 1: The 20 (out of an original 36) semantic relations observed by Badulescu and Moldovan (2009) along with their examples.

3 Dataset Creation

We created the dataset used in this work from three different sources, each representing a distinct genre—newswire, non-fiction, and fiction. Of the 21,938 total examples, 15,330 come from sections 2–21 of the Penn Treebank (Marcus et al., 1993). Another 5,266 examples are from *The History of the Decline and Fall of the Roman Empire* (Gibbon, 1776), a non-fiction work, and 1,342 are from *The Jungle Book* (Kipling, 1894), a collection of fictional short stories. For the Penn Treebank, we extracted the examples using the provided gold standard parse trees, whereas, for the latter cases, we used the output of an open source parser (Tratz and Hovy, 2011).

4 Semantic Relation Inventory

The initial semantic relation inventory for possessives was created by first examining some of the relevant literature on possessives, including work by Badulescu and Moldovan (2009), Barker (1995), Quirk et al. (1985), Rosenbach (2002), and Taylor (1996), and then manually annotating the large dataset of examples. Similar examples were grouped together to form initial categories, and groups that were considered more difficult were later reexamined in greater detail. Once all the examples were assigned to initial categories, the process of refining the definitions and annotations began.

In total, 17 relations were created, not including OTHER. They are shown in Table 3 along with approximate (best guess) mappings to relations defined by others, specifically those of Quirk et al. (1985), whose relations are presented in Table 2, as well as Badulescu and Moldovan's (2009) relations.

Relation	Examples
POSSESSIVE	my wife's father
SUBJECTIVE	boy's application
OBJECTIVE	the family's support
ORIGIN	the general's letter
DESCRIPTIVE	a women's college
MEASURE	ten days' absense
ATTRIBUTE	the victim's courage
PARTITIVE	the baby's eye
APPOSITION (marginal)	Dublin's fair city

Table 2: The semantic relations proposed by Quirk et al. (1985) for *'s* along with some of their examples.

4.1 Refinement and Inter-annotator Agreement

The semantic relation inventory was refined using an iterative process, with each iteration involv-

[2] Email requests asking for relation definitions and the data were not answered, and, thus, we are unable to provide an informative comparison with their work.

Relation	Example	HDFRE	JB	PTB	Mappings
SUBJECTIVE	Dora's travels	1083	89	3169	Q:SUBJECTIVE, B:AGENT
PRODUCER'S PRODUCT	Ford's Taurus	47	44	1183	Q:ORIGIN, B:MAKE/PRODUCE B:RESULT
OBJECTIVE	Mowgli's capture	380	7	624	Q:OBJECTIVE, B:THEME
CONTROLLER/OWNER/USER	my apartment	882	157	3940	QB:POSSESSIVE
MENTAL EXPERIENCER	Sam's fury	277	22	232	Q:POSSESSIVE, B:EXPERIENCER
RECIPIENT	their bonuses	12	6	382	Q:POSSESSIVE, B:RECIPIENT
MEMBER'S COLLECTION	John's family	144	31	230	QB:POSSESSIVE
PARTITIVE	John's arm	253	582	451	Q:PARTITIVE, B:PART-WHOLE
LOCATION	Libya's people	24	0	955	Q:POSSESSIVE, B:SOURCE/FROM B:LOCATION/SPACE
TEMPORAL	today's rates	0	1	623	Q:POSSESSIVE, B:TEMPORAL
EXTENT	6 hours' drive	8	10	5	QB:MEASURE
KINSHIP	Mary's kid	324	156	264	Q:POSSESSIVE, B:KINSHIP
ATTRIBUTE	picture's vividness	1013	34	1017	Q:ATTRIBUTE, B:PROPERTY
TIME IN STATE	his years in Ohio	145	32	237	QB:POSSESSIVE
POSSESSIVE COMPOUND	the [men's room]	0	0	67	Q:DESCRIPTIVE
ADJECTIVE DETERMINED	his fellow Brit	12	0	33	
OTHER RELATIONAL NOUN	his friend	629	112	1772	QB:POSSESSIVE
OTHER	your Lordship	33	59	146	B:OTHER

Table 3: Possessive semantic relations along with examples, counts, and approximate mappings to other inventories. Q and B represent Quirk et al. (1985) and Badulescu and Moldovan (2009), respectively. HDFRE, JB, PTB: *The History of the Decline and Fall of the Roman Empire*, *The Jungle Book*, and the Penn Treebank, respectively.

ing the annotation of a random set of 50 examples. Each set of examples was extracted such that no two examples had an identical *possessee* word. For a given example, annotators were instructed to select the most appropriate option but could also record a second-best choice to provide additional feedback. Figure 1 presents a screenshot of the HTML-based annotation interface. After the annotation was complete for a given round, agreement and entropy figures were calculated and changes were made to the relation definitions and dataset. The number of refinement rounds was arbitrarily limited to five. To measure agreement, in addition to calculating simple percentage agreement, we computed Fleiss' Kappa (Fleiss, 1971), a measure of agreement that incorporates a correction for agreement due to chance, similar to Cohen's Kappa (Cohen, 1960), but which can be used to measure agreement involving more than two annotations per item. The agreement and entropy figures for these five intermediate annotation rounds are given in Table 4. In all the possessive annotation tables, Annotator A refers to the primary author and the labels B and C refer to two additional annotators.

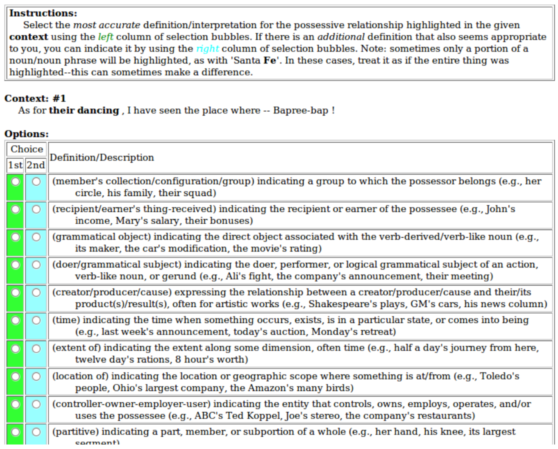

Figure 1: Screenshot of the HTML template page used for annotation.

To calculate a final measure of inter-annotator agreement, we randomly drew 150 examples from the dataset not used in the previous refinement iterations, with 50 examples coming from each of the three original data sources. All three annotators initially agreed on 82 of the 150 examples, leaving 68 examples with at least some disagreement, including 17 for which all three annotators disagreed.

Annotators then engaged in a new task in which they re-annotated these 68 examples, in each case being able to select only from the definitions previously chosen for each example by at least one annotator. No indication of who or how many people had previously selected the definitions was

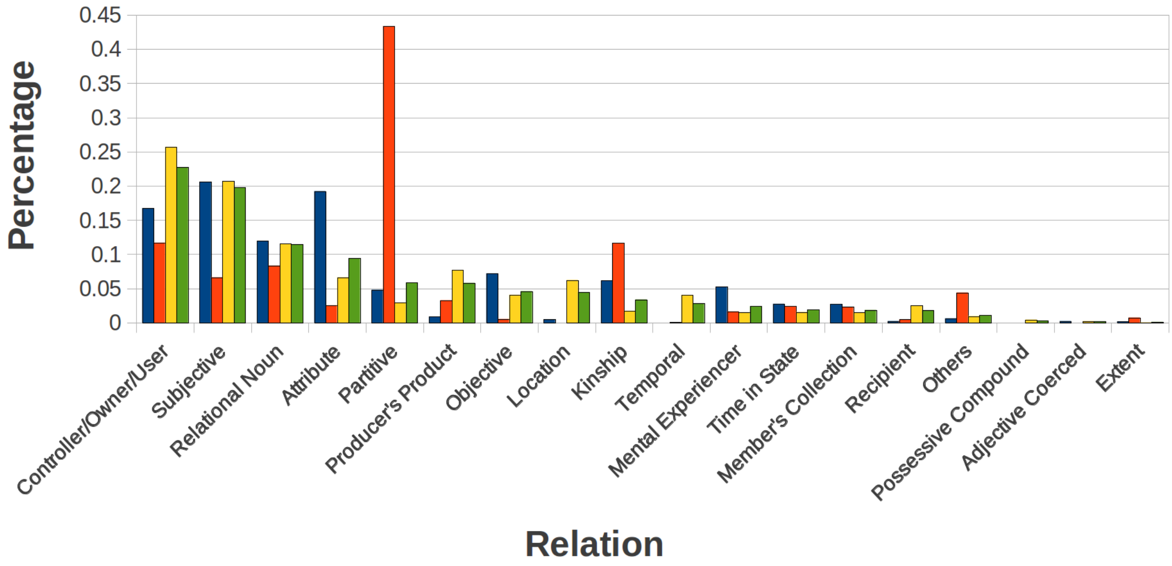

Figure 2: Semantic relation distribution for the dataset presented in this work. HDFRE: History of the Decline and Fall of the Roman Empire; JB: Jungle Book; PTB: Sections 2–21 of the Wall Street Journal portion of the Penn Treebank.

given[3]. Annotators were instructed not to choose a definition simply because they thought they had chosen it before or because they thought someone else had chosen it. After the revision process, all three annotators agreed in 109 cases and all three disagreed in only 6 cases. During the revision process, Annotator A made 8 changes, B made 20 changes, and C made 33 changes. Annotator A likely made the fewest changes because he, as the primary author, spent a significant amount of time thinking about, writing, and re-writing the definitions used for the various iterations. Annotator C's annotation work tended to be less consistent in general than Annotator B's throughout this work as well as in a different task not discussed within this paper, which probably why Annotator C made more changes than Annotator B. Prior to this revision process, the three-way Fleiss' Kappa score was 0.60 but, afterwards, it was at 0.78. The inter-annotator agreement and entropy figures for before and after this revision process, including pairwise scores between individual annotators, are presented in Tables 5 and 6.

4.2 Distribution of Relations

The distribution of semantic relations varies somewhat by the data source. *The Jungle Book*'s distribution is significantly different from the others, with a much larger percentage of PARTITIVE and KINSHIP relations. The Penn Treebank and *The History of the Decline and Fall of the Roman Empire* were substantially more similar, although there are notable differences. For instance, the LOCATION and TEMPORAL relations almost never occur in *The History of the Decline and Fall of the Roman Empire*. Whether these differences are due to variations in genre, time period, and/or other factors would be an interesting topic for future study. The distribution of relations for each data source is presented in Figure 2.

Though it is harder to compare across datasets using different annotation schemes, there are at least a couple notable differences between the distribution of relations for Badulescu and Moldovan's (2009) dataset and the distribution of relations used in this work. One such difference is the much higher percentage of examples labeled as TEMPORAL—11.35% vs only 2.84% in our data. Another difference is a higher incidence of the KINSHIP relation (6.31% vs 3.39%), although it is far less frequent than it is in *The Jungle Book* (11.62%).

4.3 Encountered Ambiguities

One of the problems with creating a list of relations expressed by *'s*-constructions is that some examples can potentially fit into multiple categories. For example, *Joe's resentment* encodes

[3]Of course, if three definitions were present, it could be inferred that all three annotators had initially disagreed.

both SUBJECTIVE relation and MENTAL EXPERIENCER relations and *UK's cities* encodes both PARTITIVE and LOCATION relations. A representative list of these types of issues along with examples designed to illustrate them is presented in Table 7.

5 Experiments

For the automatic classification experiments, we set aside 10% of the data for test purposes, and used the the remaining 90% for training. We used 5-fold cross-validation performed using the training data to tweak the included feature templates and optimize training parameters.

5.1 Learning Approach

The LIBLINEAR (Fan et al., 2008) package was used to train linear Support Vector Machine (SVMs) for all the experiments in the one-against-the-rest style. All training parameters took their default values with the exception of the C parameter, which controls the tradeoff between margin width and training error and which was set to 0.02, the point of highest performance in the cross-validation tuning.

5.2 Feature Generation

For feature generation, we conflated the possessive pronouns 'his', 'her', 'my', and 'your' to 'person.' Similarly, every term matching the case-insensitive regular expression *(corp|co|plc|inc|ag|ltd|llc)\\.?)* was replaced with the word 'corporation.'

All the features used are functions of the following five words.

- The possessor word
- The possessee word
- The syntactic governor of the possessee word
- The set of words between the possessor and possessee word (e.g., *first* in *John's first kiss*)
- The word to the right of the possessee

The following feature templates are used to generate features from the above words. Many of these templates utilize information from WordNet (Fellbaum, 1998).

- WordNet link types (link type list) (e.g., attribute, hypernym, entailment)
- Lexicographer filenames (lexnames)—top level categories used in WordNet (e.g., noun.body, verb.cognition)
- Set of words from the WordNet definitions (gloss terms)
- The list of words connected via WordNet *part-of* links (part words)
- The word's text (the word itself)
- A collection of affix features (e.g., -ion, -er, -ity, -ness, -ism)
- The last {2,3} letters of the word
- List of all possible parts-of-speech in WordNet for the word
- The part-of-speech assigned by the part-of-speech tagger
- WordNet hypernyms
- WordNet synonyms
- Dependent words (all words linked as children in the parse tree)
- Dependency relation to the word's syntactic governor

5.3 Results

The system predicted correct labels for 1,962 of the 2,247 test examples, or 87.4%. The accuracy figures for the test instances from the Penn Treebank, *The Jungle Book*, and *The History of the Decline and Fall of the Roman Empire* were 88.8%, 84.7%, and 80.6%, respectively. The fact that the score for *The Jungle Book* was the lowest is somewhat surprising considering it contains a high percentage of body part and kinship terms, which tend to be straightforward, but this may be because the other sources comprise approximately 94% of the training examples.

Given that human agreement typically represents an upper bound on machine performance in classification tasks, the 87.4% accuracy figure may be somewhat surprising. One explanation is that the examples pulled out for the inter-annotator agreement study each had a unique *possessee* word. For example, "expectations", as in "analyst's expectations", occurs 26 times as the possessee in the dataset, but, for the inter-annotator agreement study, at most one of these examples could be included. More importantly, when the initial relations were being defined, the data were first sorted based upon the possessee and then the possessor in order to create blocks of similar examples. Doing this allowed multiple examples to be assigned to a category more quickly because one can decide upon a category for the whole lot at once and then just extract the few, if any, that belong to other categories. This is likely to be both faster and more consistent than examining each

	Agreement (%)			Fleiss' κ				Entropy		
Iteration	A vs B	A vs C	B vs C	A vs B	A vs C	B vs C	All	A	B	C
1	0.60	0.68	0.54	0.53	0.62	0.46	0.54	3.02	2.98	3.24
2	0.64	0.44	0.50	0.59	0.37	0.45	0.47	3.13	3.40	3.63
3	0.66	0.66	0.72	0.57	0.58	0.66	0.60	2.44	2.47	2.70
4	0.64	0.30	0.38	0.57	0.16	0.28	0.34	2.80	3.29	2.87
5	0.72	0.66	0.60	0.67	0.61	0.54	0.61	3.21	3.12	3.36

Table 4: Intermediate results for the possessives refinement work.

	Agreement (%)			Fleiss' κ				Entropy		
Portion	A vs B	A vs C	B vs C	A vs B	A vs C	B vs C	All	A	B	C
PTB	0.62	0.62	0.54	0.56	0.56	0.46	0.53	3.22	3.17	3.13
HDFRE	0.82	0.78	0.72	0.77	0.71	0.64	0.71	2.73	2.75	2.73
JB	0.74	0.56	0.54	0.70	0.50	0.48	0.56	3.17	3.11	3.17
All	0.73	0.65	0.60	0.69	0.61	0.55	0.62	3.43	3.35	3.51

Table 5: Final possessives annotation agreement figures before revisions.

	Agreement (%)			Fleiss' κ				Entropy		
Source	A vs B	A vs C	B vs C	A vs B	A vs C	B vs C	All	A	B	C
PTB	0.78	0.74	0.74	0.75	0.70	0.70	0.72	3.30	3.11	3.35
HDFRE	0.78	0.76	0.76	0.74	0.72	0.72	0.73	3.03	2.98	3.17
JB	0.92	0.90	0.86	0.90	0.87	0.82	0.86	2.73	2.71	2.65
All	0.83	0.80	0.79	0.80	0.77	0.76	0.78	3.37	3.30	3.48

Table 6: Final possessives annotation agreement figures after revisions.

First Relation	Second Relation	Example
PARTITIVE	CONTROLLER/...	*BoA's Mr. Davis*
PARTITIVE	LOCATION	*UK's cities*
PARTITIVE	OBJECTIVE	*BoA's adviser*
PARTITIVE	OTHER RELATIONAL NOUN	*BoA's chairman*
PARTITIVE	PRODUCER'S PRODUCT	*the lamb's wool*
CONTROLLER/...	PRODUCER'S PRODUCT	*the bird's nest*
CONTROLLER/...	OBJECTIVE	*his assistant*
CONTROLLER/...	LOCATION	*Libya's oil company*
CONTROLLER/...	ATTRIBUTE	*Joe's strength*
CONTROLLER/...	MEMBER'S COLLECTION	*the colonel's unit*
CONTROLLER/...	RECIPIENT	*Joe's trophy*
RECIPIENT	OBJECTIVE	*Joe's reward*
SUBJECTIVE	PRODUCER'S PRODUCT	*Joe's announcement*
SUBJECTIVE	OBJECTIVE	*its change*
SUBJECTIVE	CONTROLLER/...	*Joe's employee*
SUBJECTIVE	LOCATION	*Libya's devolution*
SUBJECTIVE	MENTAL EXPERIENCER	*Joe's resentment*
OBJECTIVE	MENTAL EXPERIENCER	*Joe's concern*
OBJECTIVE	LOCATION	*the town's inhabitants*
KINSHIP	OTHER RELATIONAL NOUN	*his fiancee*

Table 7: Ambiguous/multiclass possessive examples.

example in isolation. This advantage did not exist in the inter-annotator agreement study.

5.4 Feature Ablation Experiments

To evaluate the importance of the different types of features, the same experiment was re-run multiple times, each time including or excluding exactly one feature template. Before each variation, the C parameter was retuned using 5-fold cross validation on the training data. The results for these runs are shown in Table 8.

Based upon the leave-one-out and only-one feature evaluation experiment results, it appears that the possessee word is more important to classification than the possessor word. The possessor word is still valuable though, with it likely being more valuable for certain categories (e.g., TEMPORAL and LOCATION) than others (e.g., KINSHIP). Hypernym and gloss term features proved to be about equally valuable. Curiously, although hypernyms are commonly used as features in NLP classification tasks, gloss terms, which are rarely used for these tasks, are approximately as useful, at least in this particular context. This would be an interesting result to examine in greater detail.

6 Related Work

6.1 Linguistics

Semantic relation inventories for the English *'s*-construction have been around for some time; Taylor (1996) mentions a set of 6 relations enumerated by Poutsma (1914–1916). Curiously, there is not a single dominant semantic relation inventory for possessives. A representative example of semantic relation inventories for *'s*-constructions is the one given by Quirk et al. (1985) (presented earlier in Section 2).

Interestingly, the set of relations expressed by possessives varies by language. For example, Classical Greek permits a *standard of comparison* relation (e.g., "better than Plato") (Nikiforidou, 1991), and, in Japanese, some relations are expressed in the opposite direction (e.g., "blue eye's doll") while others are not (e.g., "Tanaka's face") (Nishiguchi, 2009).

To explain how and why such seemingly different relations as *whole+part* and *cause+effect* are expressed by the same linguistic phenomenon, Nikiforidou (1991) pursues an approach of *metaphorical structuring* in line with the work of Lakoff and Johnson (1980) and Lakoff (1987). She thus proposes a variety of such metaphors as THINGS THAT HAPPEN (TO US) ARE (OUR) POSSESSIONS and CAUSES ARE ORIGINS to explain how the different relations expressed by possessives extend from one another.

Certainly, not all, or even most, of the linguistics literature on English possessives focuses on creating lists of semantic relations. Much of the work covering the semantics of the *'s* construction in English, such as Barker's (1995) work, dwells on the split between cases of *relational* nouns, such as *sister*, that, by their very definition, hold a specific relation to other real or conceptual things, and *non-relational*, or *sortal* nouns (Löbner, 1985), such as *car*.

Vikner and Jensen's (2002) approach for handling these disparate cases is based upon Pustejovsky's (1995) generative lexicon framework. They coerce sortal nouns (e.g., car) into being relational, purporting to create a uniform analysis. They split *lexical possession* into four types: *inherent*, *part-whole*, *agentive*, and *control*, with *agentive* and *control* encompassing many, if not most, of the cases involving sortal nouns.

A variety of other issues related to possessives considered by the linguistics literature include adjectival modifiers that significantly alter interpretation (e.g., favorite and former), double genitives (e.g., book of John's), bare possessives (i.e., cases where the possessee is omitted, as in "Eat at Joe's"), possessive compounds (e.g., driver's license), the syntactic structure of possessives, definitiveness, changes over the course of history, and differences between languages in terms of which relations may be expressed by the genitive. Representative work includes that by Barker (1995), Taylor (1996), Heine (1997), Partee and Borschev (1998), Rosenbach (2002), and Vikner and Jensen (2002).

6.2 Computational Linguistics

Though the relation between nominals in the English possessive construction has received little attention from the NLP community, there is a large body of work that focuses on similar problems involving noun-noun relation interpretation/paraphrasing, including interpreting the relations between the components of noun compounds (Butnariu et al., 2010), disambiguating preposition senses (Litkowski and Hargraves, 2007), or annotating the relation between nominals in more arbitrary constructions within the same sentence (Hendrickx et al., 2009).

Whereas some of these lines of work use fixed inventories of semantic relations (Lauer, 1995; Nastase and Szpakowicz, 2003; Kim and Baldwin, 2005; Girju, 2009; Ó Séaghdha and Copestake, 2009; Tratz and Hovy, 2010), other work allows for a nearly infinite number of interpretations (Butnariu and Veale, 2008; Nakov, 2008). Recent SemEval tasks (Butnariu et al., 2009; Hendrickx et al., 2013) pursue this more open-ended strategy. In these tasks, participating systems recover the implicit predicate between the nouns in noun compounds by creating potentially unique paraphrases for each example. For instance, a system might generate the paraphrase *made of* for the noun com-

Feature Type	Word(s)						Results	
	L	R	C	G	B	N	LOO	OO
Gloss Terms		■					0.867 (0.04)	0.762 (0.08)
Hypernyms		■					0.870 (0.04)	0.760 (0.16)
Synonyms		■					0.873 (0.04)	0.757 (0.32)
Word Itself		■					0.871 (0.04)	0.745 (0.08)
Lexnames		■					0.871 (0.04)	0.514 (0.32)
Last Letters		■					0.870 (0.04)	0.495 (0.64)
Lexnames			■				0.872 (0.04)	0.424 (0.08)
Link types		■					0.874 (0.02)	0.398 (0.64)
Link types			■				0.870 (0.04)	0.338 (0.32)
Word Itself	■						0.870 (0.04)	0.316 (0.16)
Last Letters	■						0.872 (0.02)	0.303 (0.16)
Gloss Terms	■						0.872 (0.02)	0.271 (0.04)
Hypernyms	■						0.875 (0.02)	0.269 (0.08)
Word Itself						■	0.874 (0.02)	0.261 (0.08)
Synonyms	■						0.874 (0.02)	0.260 (0.04)
Lexnames	■						0.874 (0.02)	0.247 (0.04)
Part-of-speech List			■				0.873 (0.02)	0.245 (0.16)
Part-of-speech List		■					0.874 (0.02)	0.243 (0.16)
Dependency		■					0.872 (0.02)	0.241 (0.16)
Part-of-speech List	■						0.874 (0.02)	0.236 (0.32)
Link Types	■						0.874 (0.02)	0.236 (0.64)
Word Itself					■		0.870 (0.02)	0.234 (0.32)
Assigned Part-of-Speech		■					0.874 (0.02)	0.228 (0.08)
Affixes		■					0.873 (0.02)	0.227 (0.16)
Assigned Part-of-Speech					■		0.873 (0.02)	0.194 (0.16)
Hypernyms				■			0.873 (0.02)	0.186 (0.04)
Lexnames				■			0.870 (0.04)	0.170 (0.64)
Text of Dependents		■					0.874 (0.02)	0.156 (0.08)
Parts List		■					0.873 (0.02)	0.141 (0.16)
Affixes	■						0.870 (0.04)	0.114 (0.32)
Affixes			■				0.873 (0.02)	0.105 (0.04)
Parts List	■						0.874 (0.02)	0.103 (0.16)

Table 8: Results for leave-one-out and only-one feature template ablation experiment results for all feature templates sorted by the only-one case. L, R, C, G, B, and N stand for *left word (possessor)*, *right word (possessee)*, *pairwise combination of outputs for possessor and possessee*, *syntactic governor of possessee*, *all tokens between possessor and possessee*, and *the word next to the possessee (on the right)*, respectively. The C parameter value used to train the SVMs is shown in parentheses.

pound *pepperoni pizza*. Computer-generated results are scored against a list of human-generated options in order to rank the participating systems. This approach could be applied to possessives interpretation as well.

Concurrent with the lack of NLP research on the subject is the absence of available annotated datasets for training, evaluation, and analysis. The NomBank project (Meyers et al., 2004) provides coarse annotations for some of the possessive constructions in the Penn Treebank, but only those that meet their criteria.

7 Conclusion

In this paper, we present a semantic relation inventory for *'s* possessives consisting of 17 relations expressed by the English *'s* construction, the largest available manually-annotated collection of possessives, and an effective method for automatically assigning the relations to unseen examples. We explain our methodology for building this inventory and dataset and report a strong level of inter-annotator agreement, reaching 0.78 Kappa overall. The resulting dataset is quite large, at 21,938 instances, and crosses multiple domains, including news, fiction, and historical non-fiction. It is the only large fully-annotated publicly-available collection of possessive examples that we are aware of. The straightforward SVM-based automatic classification system achieves 87.4% accuracy—the highest automatic possessive interpretation accuracy figured reported to date. These high results suggest that SVMs are a good choice for automatic possessive interpretation systems, in contrast to Moldovan and Badulescu (2005) findings. The data and software presented in this paper are available for download at http://www.isi.edu/publications/licensed-sw/fanseparser/index.html.

8 Future Work

Going forward, we would like to examine the various ambiguities of possessives described in Section 4.3. Instead of trying to find the one-best interpretation for a given possessive example, we would like to produce a list of *all* appropriate intepretations.

Another avenue for future research is to study variation in possessive use across genres, including scientific and technical genres. Similarly, one could automatically process large volumes of text from various time periods to investigate changes in the use of the possessive over time.

Acknowledgments

We would like to thank Charles Zheng and Sarah Benzel for all their annotation work and valuable feedback.

References

Adriana Badulescu and Dan Moldovan. 2009. A Semantic Scattering Model for the Automatic Interpretation of English Genitives. *Natural Language Engineering*, 15:215–239.

Chris Barker. 1995. *Possessive Descriptions*. CSLI Publications, Stanford, CA, USA.

Cristina Butnariu and Tony Veale. 2008. A Concept-Centered Approach to Noun-Compound Interpretation. In *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 81–88.

Cristina Butnariu, Su Nam Kim, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz, and Tony Veale. 2009. SemEval-2010 Task 9: The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions. In *DEW '09: Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pages 100–105.

Cristina Butnariu, Su Nam Kim, Preslav Nakov, Diarmuid. Ó Séaghdha, Stan Szpakowicz, and Tony Veale. 2010. SemEval-2010 Task 9: The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 39–44.

Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research*, 9:1871–1874.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Edward Gibbon. 1776. *The History of the Decline and Fall of the Roman Empire*, volume I of *The History of the Decline and Fall of the Roman Empire*. Printed for W. Strahan and T. Cadell.

Roxanna Girju. 2009. The Syntax and Semantics of Prepositions in the Task of Automatic Interpretation of Nominal Phrases and Compounds: A Cross-linguistic Study. *Computational Linguistics - Special Issue on Prepositions in Application*, 35(2):185–228.

Bernd Heine. 1997. *Possession: Cognitive Sources, Forces, and Grammaticalization*. Cambridge University Press, United Kingdom.

Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. 2009. Semeval-2010 task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*, pages 94–99.

Iris Hendrickx, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz, and Tony Veale. 2013. Task Description: SemEval-2013 Task 4: Free Paraphrases of Noun Compounds. `http://www.cs.york.ac.uk/semeval-2013/task4/`. [Online; accessed 1-May-2013].

Su Nam Kim and Timothy Baldwin. 2005. Automatic Interpretation of Noun Compounds using WordNet::Similarity. *Natural Language Processing–IJCNLP 2005*, pages 945–956.

Rudyard Kipling. 1894. *The Jungle Book*. Macmillan, London, UK.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live by*. The University of Chicago Press, Chicago, USA.

George Lakoff. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. The University of Chicago Press, Chicago, USA.

Mark Lauer. 1995. Corpus Statistics Meet the Noun Compound: Some Empirical Results. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*, pages 47–54.

Ken Litkowski and Orin Hargraves. 2007. SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 24–29.

Sebastian Löbner. 1985. Definites. *Journal of Semantics*, 4(4):279.

Mitchell P. Marcus, Mary A. Marcinkiewicz, and Beatrice Santorini. 1993. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):330.

Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young, and Ralph Grishman. 2004. The NomBank Project: An Interim Report. In *Proceedings of the NAACL/HLT Workshop on Frontiers in Corpus Annotation*.

Dan Moldovan and Adriana Badulescu. 2005. A Semantic Scattering Model for the Automatic Interpretation of Genitives. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 891–898.

Preslav Nakov. 2008. Noun Compound Interpretation Using Paraphrasing Verbs: Feasibility Study. In *Proceedings of the 13th International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, pages 103–117.

Vivi Nastase and Stan Szpakowicz. 2003. Exploring Noun-Modifier Semantic Relations. In *Fifth International Workshop on Computational Semantics (IWCS-5)*, pages 285–301.

Kiki Nikiforidou. 1991. The Meanings of the Genitive: A Case Study in the Semantic Structure and Semantic Change. *Cognitive Linguistics*, 2(2):149–206.

Sumiyo Nishiguchi. 2009. Qualia-Based Lexical Knowledge for the Disambiguation of the Japanese Postposition No. In *Proceedings of the Eighth International Conference on Computational Semantics*.

Diarmuid Ó Séaghdha and Ann Copestake. 2009. Using Lexical and Relational Similarity to Classify Semantic Relations. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 621–629.

Barbara H. Partee and Vladimir Borschev. 1998. Integrating Lexical and Formal Semantics: Genitives, Relational Nouns, and Type-Shifting. In *Proceedings of the Second Tbilisi Symposium on Language, Logic, and Computation*, pages 229–241.

James Pustejovsky. 1995. *The Generative Lexicon*. MIT Press, Cambridge, MA, USA.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman Inc., New York.

Anette Rosenbach. 2002. *Genitive Variation in English: Conceptual Factors in Synchronic and Diachronic Studies*. Topics in English linguistics. Mouton de Gruyter.

Sidney Siegel and N. John Castellan. 1988. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill.

John R. Taylor. 1996. *Possessives in English*. Oxford University Press, New York.

Stephen Tratz and Eduard Hovy. 2010. A Taxonomy, Dataset, and Classifier for Automatic Noun Compound Interpretation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 678–687.

Stephen Tratz and Eduard Hovy. 2011. A Fast, Accurate, Non-Projective, Semantically-Enriched Parser. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1257–1268.

Carl Vikner and Per Anker Jensen. 2002. A Semantic Analysis of the English Genitive. Interation of Lexical and Formal Semantics. *Studia Linguistica*, 56(2):191–226.