Natural Language Engineering **15** (2): 215–239. © 2008 Cambridge University Press
doi:10.1017/S1351324908004798 Printed in the United Kingdom

A Semantic Scattering model for the automatic interpretation of English genitives

ADRIANA BADULESCU and DAN MOLDOVAN
Lymba Corporation, 1701 N. Collins Blvd., Suite 3000, Richardson, TX 75080, USA
e-mail: {adriana,moldovan}@lymba.com

(*Received 14 March 2006; Revised 9 May 2008; first published online 29 August 2008*)

Abstract

An important problem in knowledge discovery from text is the automatic extraction of semantic relations. This paper addresses the automatic classification of the semantic relations expressed by English genitives. A learning model is introduced based on the statistical analysis of the distribution of genitives' semantic relations in a corpus. The semantic and contextual features of the genitive's noun phrase constituents play a key role in the identification of the semantic relation. The algorithm was trained and tested on a corpus of approximately 20,000 sentences and achieved an f-measure of 79.80 per cent for of-genitives, far better than the 40.60 per cent obtained using a Decision Trees algorithm, the 50.55 per cent obtained using a Naive Bayes algorithm, or the 72.13 per cent obtained using a Support Vector Machines algorithm on the same corpus using the same features. The results were similar for s-genitives: 78.45 per cent using Semantic Scattering, 47.00 per cent using Decision Trees, 43.70 per cent using Naive Bayes, and 70.32 per cent using a Support Vector Machines algorithm. The results demonstrate the importance of word sense disambiguation and semantic generalization/specialization for this task. They also demonstrate that different patterns (in our case the two types of genitive constructions) encode different semantic information and should be treated differently in the sense that different models should be built for different patterns.

1 Introduction

1.1 Problem description

The identification of semantic relations in open text is at the core of Natural Language Processing and many of its applications. Detecting semantic relations is useful for the syntactic and semantic analysis of text and thus plays an important role in automatic text understanding and generation.

Furthermore, semantic relations represent the core elements in the organization of lexical semantic knowledge bases used for inferences. Recently, there has been a renewed interest in text semantics fueled in part by the complexity of some major research initiatives in question answering, text summarization, text understanding, and others, launched in the United States and abroad.

Two of the most frequently used linguistic constructions that encode a large set of semantic relations are the s-genitives, e.g. *man's brother*, and the of-genitives, e.g. *dress of silk*.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Consider, for example, the following text:

"The Riverside County District Attorney's Office filed a misdemeanor vehicular manslaughter charge against Ramon Segura in connection with the death of Alyssa Castro of Thousand Palms in September 2003. The child's parents were dropping her off at a day care center in the 30-800 block of Landau Boulevard in Cathedral City. The child's mother had moved the child from a car safety seat to an area near the open passenger-side door of the car [1]."

There are seven semantic relations expressed by genitives: (1) *Riverside County District Attorney's Office* is an s-genitive encoding a possession relation, (2) *death of Alyssa Castro* is an of-genitive encoding a theme relation, (3) *Alyssa Castro of Thousand Palms* is an of-genitive encoding a source/from relation, (4) *child's parents* is an s-genitive encoding a kinship relation, (5) *30-800 block of Landau Boulevard* is an of-genitive encoding a part-whole and location relation, (6) *child's mother* is an s-genitive encoding a kinship relation, and (7) *passenger-side door of the car* is an of-genitive encoding a part-whole relation.

The interpretation of these phrase-level constructions is paramount for various applications that make use of lexical semantics.

This paper provides a detailed corpus analysis of genitive constructions and a model for their automatic interpretation in English texts.

The next subsection discusses the genitives' semantics and contains a summary of previous work on genitives. Section 2 presents a detailed analysis of the set of semantic relations encoded by the genitive constructions, and provides statistical evidence of their distribution in a very large corpus. In Section 3, we propose a general model that gives a comprehensive semantic characterization of the genitive constructions, and Section 4 describes the algorithm for their automatic interpretation in open text. The results obtained are summarized in Section 5, followed by discussion and conclusions.

1.2 Semantics of genitives

In English, there are two kinds of genitives. In general, in one, the modifier is morphologically linked to the possessive clitic *'s* and precedes the head noun (s-genitive, i.e. NP_{modif} *'s* NP_{head} where NP_{head} is an N-bar meaning that it does not have specifiers or determiners), and in the second one the modifier is syntactically marked by the preposition *of* and follows the head noun (of-genitive, i.e. NP_{head} *of* NP_{modif}).

Although, the genitive constructions have been studied for a long time in linguistics, their semantic investigation proved to be very difficult, as the meanings of the two constructions are difficult to pin down. There are many factors that contribute to the genitives' semantic behavior, such as the type of the genitive, the semantics of the constituent nouns, the order of the constituents, the syntactic role, the surrounding context, and others.

[1] "Civil trial pending in infant's car death," *The Desert Sun*, October 18, 2004, http://www.thedesertsun.com/news/stories2004/local/20041018012715.shtml/.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

According to Partee and Borschev (1999) "the terminology surrounding *possessives* and *genitives* is confusing, since the correspondences among morphological forms, syntactic positions, grammatical relations, and semantic interpretations are complex and subject to debate."

A characteristic of genitives is that they are very productive, as the construction can be given various semantic interpretations. For example, *Ernest's book* can mean the book Ernest owns, the book Ernest wrote, the book Ernest is reading, or the book Ernest is very fond of. Each of these interpretations is possible in the right context.

For a long time, the two types of genitives have been considered as being identical, encoding only alternative ways of expressing the same information (Jespersen, 1949; Strang, 1962; Altenberg, 1982; Quirk, Greenbaum, Leech and Svartvik 1985; Chomsky, 1986). The difference between them consists in the position of the head constituent. While in of-genitives the head constituent is initial, in s-genitives the head is final. The use of different types of genitives is driven by the fact that in English the given (active) information always precedes the new (inactive) information. Therefore, if the head of the compound is the given information then an of-genitive is used (e.g. *the car of John*). Otherwise, an s-genitive is used (e.g. *Mary's beauty*).

However, this proves to be wrong, as the two types of genitives differ in their semantics (Langacker, 1990, 1992, 1993; Vikner and Jensen, 1999; Stefanowitsch, 2001; Huddleston and Pullum, 2002). For instance, the part-whole relation expressed by the pair *dress-silk* can be encoded by the of-genitive *dress of silk*, but not by the s-genitive **silk's dress*. The researchers focused on the analysis of a vast variety of relations that can be encoded by both constructions and claim that for each type of genitive representation there is a core set of semantic relations that cannot be encoded by the other (Langacker, 1990, 1992, 1993; Vikner and Jensen, 1999; Stefanowitsch, 2001).

Moreover, researchers (Williams, 1982; Barker, 1995; Vikner and Jensen, 1999) have acknowledged the underspecified character of the genitive constructions and their distinctive interpretations based on lexical and pragmatic information.

The lexical interpretation can be detected only from the lexical semantics of the constituent nouns, NP_{modif} and NP_{head}, and most of the time this information can be extracted from lexicons. One such example is a construction in which one of the nouns in the genitive can be used only in relation to a class of other nouns and in particular contexts. These nouns are called relational, as they evoke a particular semantic relation by themselves. For instance, in the example *Mary's spouse*, the noun *spouse* evokes a kinship relation. In this case, although the relation is implicit, it is made available only by the semantics of the two constituent nouns.

On the other hand, the construction of pragmatic interpretations depends essentially on pragmatic and discourse knowledge. This corresponds to those situations where the constituents are stand-alone nouns that do not evoke, by themselves, a particular semantic relation. World and pragmatic knowledge can be used for finding the most likely interpretation or ranking the possible interpretation. For example, as we previously mentioned, the genitive *Ernest's book* can have a large number of interpretations that are all equally possible in the right context. However, if Ernest is the famous writer Ernest Hemingway, then the interpretations as the book written by Ernest become more likely.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Given the importance of the semantic relations encoded by the genitive constructions, the disambiguation of these relations has been handled in linguistics by two main approaches.

The first approach is based on a prototype analysis that takes one of the semantic relations as basic and defines all the other possible interpretations as extensions of this basic prototype. This semantic approach defined by Taylor (1996) and adopted by other researchers (Nikiforidou, 1991; Stefanowitsch, 2001) is based on the principle of inheritance with overrides. He observed that possession is the semantic relation used in English texts with very high frequency and can be defined by the following properties:

- the possessor is a human being,
- the possessee is a specific concrete thing,
- the possessor owns the possessee,
- possessor and possessee need to be in close spatial proximity, and
- the relation is a long-term one.

The other semantic relations are analyzed as derivative members of the prototype relation, where the members share some of the prototype's properties. For example, in the genitive construction *Mary's brother* (kinship), the referent of brother is not a concrete thing, and we cannot say that Mary owns her brother. However, all the other properties defined earlier for possession hold.

Although it received considerable attention in the linguistics community, this approach has difficulties in defining all the characteristics of the relations using a single prototype. For example, it is unclear how a property relation, like *Mary's coldness*, could be explained by the possession prototype, since neither *Mary* nor *coldness* share any of the prototype's properties. In addition, there are possession relations that are not covered by Taylor's properties. For instance, *the dog's favorite bone* which is a possession relation, does not satisfy the first property since the possessee (the dog) is not a human being.

The second approach proposed is a lexical and semantic analysis of the genitives that is based on well-defined constraints that help to distinguish a particular semantic relation from all the others. (Langacker 1992, 1993, 1995) claimed that the analysis of genitives deals with the semantics of the modifier ("reference-point" constituent) and of the head ("target" constituent). The general idea is that any attempt to interpret genitive constructions has to deal with the semantic analysis of the two noun constituents.

This last semantic approach for the interpretation of English genitives is very important and useful for the automatic detection of semantic relations and is the one we adopt.

1.3 Previous work

There has been much interest recently in the discovery of semantic relations from open text using symbolic and statistical techniques. This includes the Senseval 3 and CoNLL competitions on automatic labeling of semantic roles (Litkowski, 2004;

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Carreras and Marquez, 2004, 2005), the interpretation of noun compound semantics (Lapata, 2002; Rosario, Heart and Fillmore, 2002; Girju, Moldovan, Tatu and Antohe, 2005), the SemEval nominal relations competition (Girju *et al.*, 2007), and many others. However, not much work has been done to automatically interpret the genitive constructions.

In 1999, Berland and Charniak (1999) applied statistical methods over a very large corpus (North American News Corpus) to find part-whole relations. Following Hearst's method for the automatic acquisition of hypernymy relations (Hearst, 1998), they used the genitive construction to detect part-whole relations based on a list of six seeds representing "*whole*" objects (i.e. *book, building, car, hospital, plant,* and *school*). Their system's output was an ordered list of possible "*parts*" according to some statistical metrics (Dunning's log-likelihood metric and Johnson's significant-difference metric). They filtered the words ending in *"ing," "ness,"* and *"ity"* and presented the results only for two specific patterns (*"NN's NN"'* and *"NN of DT NN"*). The accuracy obtained for the first 50 parts was 55.00 per cent and for the first 20 parts was 70.00 per cent.

In 2003, Girju, Badulescu, and Moldovan (2003) detected the part-whole relations for some of the most frequent part-whole patterns (including the genitives) using Iterative Semantic Specialization, a learning model that searches for constraints in the WordNet noun hierarchies. For their work, they used the SemCor 1.7 corpus (part of Brown corpus) and the LA Times (LAT) articles from the TREC-9 collection. The results were 83.57 per cent precision and 72 per cent recall for genitives and "have" patterns for the part-whole relation. In 2006, in Girju, Badulescu, and Moldovan (2006), they improved the method and used it for a larger variety of patterns obtaining an f-measure of 87.26 per cent for genitives, 73.59 per cent for noun compounds, and 75.97 per cent for preposition cluster (patterns of the type *NP prep NP*).

In 2004, Moldovan, Badulescu, Tatu, Antohe, and Girju (2004) tried to use different machine learning techniques for interpretation of different constructions like genitives, noun compounds, adjectival phrases, and adjectival clauses with promising results. For their work, they used eXtended WordNet (XWN) 2.0 glosses and Wall Street Journal (WSJ) articles from the TREC-9 collection. They even compared their results with the first version of our Semantic Scattering software that outperformed the other machine learning techniques for genitives.

In 2004, Badulescu (2004) introduced new per-relation methods (for extracting one relation at a time) and per-pattern methods (for extracting all the relations encoded by a pattern). For training and evaluating her systems and experiments, she used the LAT articles and WSJ articles from the TREC-9 collection. She achieved an f-measure of 83.05 per cent for part-whole and 62.37 per cent for possession using the per-relation method and 79.85 per cent for genitives and 78.75 per cent for the noun compounds using the per-pattern method.

In 2005, Moldovan and Badulescu (2005) used an improved version of the semantic scattering software achieving an average of 78.63 per cent f-measure for genitives on LAT articles.

In 2005, Girju *et al.* (2005) built a system for noun compounds interpretation using features previously used in noun compound interpretation and parenthetization. For

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

their work, they also used the XWN 2.0 glosses and WSJ articles from the TREC-9 collection. They also built a simplified version of our model (following (Moldovan *et al*., 2004) to use as a benchmark for their research. Their version of the Semantic Scattering, however, did not perform as well as the original Moldovan *et al.* (2004) model.

In 2007, researchers worldwide (Girju *et al*., 2007) used predefined or automatic methods (mostly machine learning techniques) for detecting a set of seven relations (cause-effect, instrument-agency, product-producer, origin-entity, theme-tool, part-whole, and content-container) (one relation at a time) on a web corpus obtaining an average 60.15 per cent f-measure that was lower than a 64.8 per cent all true baseline (for a system that always returns the target relation). The systems were trained and tested on a web corpus constructed using nominal queries. A small percentage of the sentences contained genitives. The top scores (f-measures) were 58 per cent for the systems that did not use word senses or the queries, 72.4 per cent for the systems that used only word senses, 58.4 per cent for the systems that used only the queries, and 60.6 per cent for the systems that used both word senses and queries.

Given the importance of the semantic relations encoded by the genitive constructions, the disambiguation of these relations has long been studied (Williams, 1982; Nikiforidou, 1991; Taylor, 1996; Vikner and Jensen, 1999; Stefanowitsch, 2001; and others).

2 Genitives' corpus analysis

2.1 The data

In order to construct a general model on genitives, we analyzed the syntactic and semantic behavior of both constructions on a large corpus of examples (representative for the collection) selected randomly from an open domain text collection, the LAT articles from the TREC-9 collection. This analysis is justified by our desire to answer the following questions: "*What are the semantic relations encoded by the genitives?*" and "*What is their distribution in a large corpus?*"

A set of 20,000 sentences was randomly selected from the LAT collection. In this set, there were 1,040 of-genitives and 960 s-genitives. From these, 80 per cent were used for training and 20 per cent for testing (10 per cent for development and 10 per cent for testing).

Each genitive instance was tagged with the corresponding semantic relation by two annotators, based on a list of the 35 most frequently used semantic relations presented by Badulescu (2004) which are shown in Table 1. The genitives' noun components were manually disambiguated with the corresponding WordNet 2.0 senses (Fellbaum, 1998).

For the concepts or senses of concepts that were not in WordNet, we used their Named Entities. For extracting the named entities, we used an in-house Named Entity Recognizer (Moldovan, Clark and Bowden, 2007) that has an 89.53 per cent f-measure on MUC6 data that labels events (events, activities, actions), living things (humans, animals, plants), groups (groups, organizations, institutions, industries),

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

locations (municipalities/towns, provinces/states, countries, continents, planets, attractions, airports, addresses, universities, and other geographic points), products (products, weapons, instruments, chemical elements, drugs), dates (dates, times, time periods), quantities (quantities, monetary values, prices, numbers, percentages, currencies, phone numbers), diseases, properties (colors, flavors, etc.), and other names. The entities were uniquely mapped into a corresponding WordNet concept (e.g. *town* to *municipality#1*) and that WordNet concept was used to represent the entity.

2.2 Interannotator agreement

The annotators, two graduate students in Computational Semantics, were given the genitives and the sentences in which they occurred. Whenever the annotators found an example encoding a semantic relation other than those provided or without any semantic relation, they had to tag it as "other." In addition to the type of relation, the annotators were asked to provide the correct WordNet 2.0 senses of the two nouns and information about the *order* of the arguments of the relation in the genitive construction.

For s-genitives, the order is always the same: the head is always the first argument of the relation while the modifier is always the second argument of the relation. For instance, in *John's hand* (part-whole) the "part" concept—the first argument of the part-whole relation (*hand*)—is the head while the "whole" concept—the second argument of the part-whole relation (*John*)—is the modifier of the construction. However, for of-genitives, the order is not always the same. For instance, in *the car of the owner* (possession) the possessee—first argument (*the car*) is the head and the possessor—second argument (*owner*) is the modifier, while in *the owner of the car* the possessee is the modifier and the possessor is the head and the order is reversed. Approximately one-third of the training examples had the nouns in reverse order.

Most of the time, one genitive instance was tagged with one semantic relation, but there were also situations in which an example could belong to more than one relation in the same context. For example, the genitive *city of USA* was tagged as a part-whole relation and as a location relation. There were 21 such cases in the training corpus.

The judges' agreement was measured using the Kappa statistic (Siegel and Castellan, 1988), one of the most frequently used measures of interannotator agreement for classification tasks: $\kappa = \frac{Pr(A)-Pr(E)}{1-Pr(E)}$, where $Pr(A)$ is the proportion of times the raters agree, and $Pr(E)$ is the probability of agreement by chance.

The κ coefficient is 1 if there is a total agreement among the annotators, and 0 if there is no agreement other than that expected to occur by chance.

The average κ coefficient is close to 0.82 for both of- and s-genitives, showing a good level of agreement for the training and testing data on the set of 35 relations, taking into consideration the difficulty of the task. This can be explained by the instructions the annotators received prior to annotation and by their expertise in Lexical Semantics.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Table 1. *The distribution of the semantic relations in the annotated corpus of 20,000 sentences*

Number	Semantic relations	Frequency of-genitives	Frequency s-genitives	Examples
1	possession	36	220	*Mary's book*
2	kinship	25	61	*Mary's brother*
3	property	109	75	*John's coldness*
4	agent	11	123	*Investigation of the crew*
5	temporal	5	109	*Last year's exhibition*
6	depiction-depicted	29	7	*A picture of my nice*
7	part-whole	329	114	*The girl's mouth*
8	hypernymy (is-a)	0	0	N/A
9	entail	0	0	N/A
10	cause	10	3	*Death of cancer*
11	make/produce	11	57	*Maker of computer*
12	instrument	0	0	N/A
13	location/space	32	46	*University of Texas*
14	purpose	0	0	N/A
15	source/from	56	33	*President of Bolivia*
16	topic	66	5	*Museum of art*
17	manner	0	0	N/A
18	means	0	0	N/A
19	accompaniment	10	4	*Solution of the problem*
20	experiencer	1	2	*Victim of lung disease*
21	recipient	44	32	*Josephine's reward*
22	frequency	0	0	N/A
23	influence	0	0	N/A
24	associated with	5	2	*Contractors of shipyard*
25	measure	113	1	*Hundred of dollars*
26	synonymy	0	0	N/A
27	antonymy	0	0	N/A
28	prob. of existence	0	0	N/A
29	possibility	0	0	N/A
30	certainty	0	0	N/A
31	theme	112	38	*Acquisition of the holding*
32	result	7	2	*Result of the review*
33	stimulus	0	0	N/A
34	extent	0	0	N/A
35	predicate	0	0	N/A
36	other	29	26	*State of emergency*

2.3 Distribution of semantic relations

Table 1 shows the distribution of the semantic relations in the annotated corpus.

In the case of the of-genitives, there were 19 relations found from the total of 35 relations considered, and the missing relations (N/A) are hypernymy, entailment, instrument, purpose, manner, means, frequency, influence, synonymy, antonymy, probability, possibility, certainty, stimulus, extent, and predicate. The

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

most frequently occurring relations were possession, kinship, property, part-whole, location, source, theme, and measure.

There were other relations (107 for of-genitives) that do not belong to the predefined list of 35 relations, such as *state of emergency*. These examples were clustered in different undefined subsets based on their semantics. The largest subsets did not cover more than 3 per cent of the other set of examples. This observation shows that the set of 35 semantic relations from Table 1 is representative for genitives.

Table 1 also shows the semantic preferences of each genitive form. For example, possession, kinship, and some kinds of part-whole relations are most of the time encoded by the s-genitive, while some specific part-whole relations, such as *dress of silk* and *array of flowers*, can only be encoded by the of-genitive. This simple analysis leads to the important conclusion that the two constructions must be treated separately as their semantic content is different. This observation is also consistent with other work in Linguistics on the grammatical variation of the English genitives (Stefanowitsch, 2001).

The distribution of the relations is different for the two construction types. For of-genitives the part-whole is the most frequently used relation, followed by theme and measure, while for s-genitives, the possession is the most frequent, followed by agent and part-whole.

3 The model

3.1 Problem formulation

Given a genitive, the goal is to develop a procedure for the automatic labeling of the semantic relation it conveys. The semantic relation derives from the semantics of the noun phrases participating in genitives as well as the surrounding context.

Semantic classification of syntactic patterns in general can be formulated as a learning problem. This is a multiclass classification problem since the output can be one of the semantic relations in the set. We cast this as a supervised learning problem where input/output pairs are available as training data.

An important (first) step is to map the characteristics of each genitive construction into a feature vector. We define x_i to be the feature vector of an instance i and X to be the space of all instances, i.e. $x_i \in X$.

The multiclass classification is performed by a function that maps the feature space X into a semantic space S: $F : X \rightarrow S$, where S is the set of semantic relations from Table 1, i.e. $r_k \in S$.

Let T be the training set of examples or instances $T = (x_1 r_1, x_2 r_2, \ldots, x_n r_n) \subseteq (X \times S)^n$ where n is the number of examples x, each accompanied by its semantic relation label r. The problem is to decide which semantic relation r to assign to a new, unseen example x_{n+1}. In order to classify a given set of examples (members of X), one needs some kind of measure of the similarity (density, distance, etc.) or the difference between any two given members of X.

3.2 Feature space

An essential aspect of our approach is the word sense disambiguation (WSD) of the noun. Using a state-of-the-art open-text WSD system with 65 per cent accuracy

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

for nouns (Novischi, Moldovan, Parker, Badulescu and Hauser, 2004), each word is mapped into its corresponding WordNet 2.0 sense. The disambiguation process takes into account surrounding words, and it is through this process that *context* gets to play a role in labeling the genitives' semantics.

So far, we have identified and experimented with the following NP features:

(1) **Semantic class of head noun** specifies the WordNet sense (synset) of the head noun and implicitly points to all its hypernyms. It is extracted automatically via a WSD module. The genitive semantics is influenced heavily by the meaning of the noun constituents. For example, *child's mother* is a kinship relation whereas *child's toy* is a possession relation.

(2) **Semantic class of modifier noun** specifies the WordNet synset of the modifier noun. The following examples show that the semantic interpretation of a genitive is also influenced by the semantic of the modifier noun; *Mary's apartment* is a possession relation, and *apartment of New York* is a location relation.

The positive and negative genitive examples of the training corpus are pairs of concepts of the following format:

$$< modifier_semclass\#WNsense\,;head_semclass\#WNsense\,;target >,$$

where `target` is a set of at least one of the 36 semantic relations. The *modifier_semclass* and *head_semclass* concepts are WordNet semantic classes tagged with their corresponding WordNet senses.

3.3 Semantic Scattering learning model

For every pair of <*modifier–head*> noun genitives, let us define with f_i^m and f_j^h the WordNet 2.0 senses of the modifier and the head, respectively. For convenience we replace the tuple $< f_i^m, f_j^h >$ with f_{ij}.

The Semantic Scattering model is based on the following observations:

Observation 1. f_i^m and f_j^h can be regarded as nodes on some paths that link the senses of the most specific noun concepts with the top of the noun hierarchies.

Observation 2. The closer the pair of noun senses f_{ij} is to the bottom of noun hierarchies the fewer the semantic relations associated with it; the more general the f_{ij} is the more semantic relations there are.

The probability of a semantic relation r given feature pair f_{ij} is $P(r|f_{ij}) = \frac{n(r,f_{ij})}{n(f_{ij})}$ and is defined as the ratio between the number of occurrences of a relation r in the presence of feature pair f_{ij} over the number of occurrences of feature pair f_{ij} in the corpus.

The most probable relation $\hat{r}$ is $\hat{r} = \text{argmax}_{r \in R} P(r|f_{ij})$.

From the training corpus, one can measure the quantities $n(r,f_{ij})$ and $n(f_{ij})$.

Depending on the level of abstraction of f_{ij} two cases are possible.

Case 1. The feature pair f_{ij} is specific enough such that there is only one semantic relation r for which $P(r|f_{ij}) = 1$ and 0 for all the other semantic relations.

Case 2. The feature pair f_{ij} is general enough such that there are at least two semantic relations for which $P(r|f_{ij}) \neq 0$. In this case, the most appropriate $\hat{r}$ is $\hat{r} = \text{argmax}_{r \in R} P(r|f_{ij})$.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

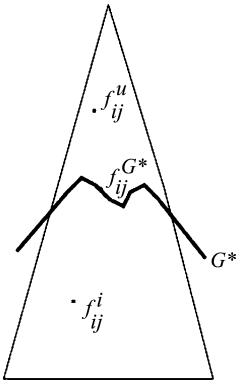


Fig. 1. Conceptual view of the noun hierarchies separated by the boundary G^*.

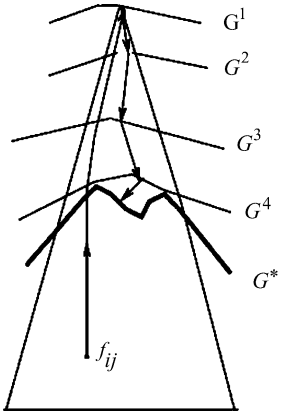


Fig. 2. Boundary G^* is found through an iterative process called "Semantic Scattering".

Definition. A boundary G^* in the WordNet noun hierarchies is a set of synset pairs such that
(1) for any feature pair on the boundary, denoted $f_{ij}^{G^*} \in G^*, f_{ij}^{G^*}$ maps uniquely into only one relation r,
(2) for any $f_{ij}^{u} \succ f_{ij}^{G^*}, f_{ij}^{u}$ maps into more than one relation r, and
(3) for any $f_{ij}^{l} \prec f_{ij}^{G^*}, f_{ij}^{l}$ maps uniquely into a semantic relation r.

Here relations $\succ$ and $\prec$ mean "semantically more general" and "semantically more specific," respectively. This is illustrated in Figure 1.

Observation 3. We have noticed that there are more concept pairs under the boundary G^* than above, i.e. $|\{f_{ij}^{l}\}| \gg |\{f_{ij}^{u}\}|$.

3.4 Boundary detection algorithm

An approximation to boundary G^* is found using the training set through an iterative process called *Semantic Scattering*. We start with the most general boundary corresponding to the nine WordNet noun hierarchies and then specialize it based on the training data until a good approximation is reached (Figure 2).

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Table 2. *The conditional probabilities for different relations for the feature* entity#1-entity#1

No.	Relation (r)	$P(r\|f_{ij})$
1	possession	0.048
2	kinship	0.120
3	property	0.006
6	depiction-depicted	0.032
7	part-whole	0.430
11	make/produce	0.016
13	location/space	0.035
15	source/from	0.285
16	topic	0.012
19	accompaniment	0.004
21	recipient	0.010
24	associated with	0.001
25	measure	0.001
Others		0.000

The numbers in the first column identify the semantic relations (as in Table 1)

Step 1. Create an initial boundary

The initial boundary denoted by G^1 is formed from combinations of the nine WordNet hierarchies: *abstraction#6*, *act#2*, *entity#1*, *event#1*, *group#1*, *possession #2*, *phenomenon#1*, *psychological_feature#1*, and *state#4*. To each training example a corresponding feature $f_{ij} = < f_i^m, f_j^h >$ is first determined, after which we replace the most general corresponding feature consisting of top WordNet hierarchy concepts denoted by f_{ij}^1. For instance, the example *apartment of the woman* corresponds to the general feature *entity#1-entity#1* and the possession relation, *husband of the woman* corresponds to *entity#1-entity#1* and the kinship relation, and *hand of the woman* corresponds to *entity#1-entity#1* and the part-whole relation. At this high-level G^1, each feature pair f_{ij}^1 corresponds to a number of semantic relations. For each feature, one can determine the most probable relation using the equation $\hat{r} = \text{argmax}_{r \in R} P(r|f_{ij})$. For instance, feature *entity#1-entity#1* corresponds to 13 relations, the most probable one being the part-whole relation as indicated by Table 2.

Step 2. Specialize the boundary

2.1 Constructing a lower boundary

This step consists of specializing the semantic classes of the ambiguous features. A feature f_{ij}^k on boundary G^k is ambiguous if it corresponds to more then one relation and its most relevant relation has a conditional probability less than 0.9 (the threshold obtained from experiments). To eliminate nonimportant specializations, we specialize only the ambiguous classes that occur in more than 1 per cent of the training examples.

The specialization procedure consists of first identifying features f_{ij}^k with more than one semantic relation, then replacing these features with their hyponyms. Thus,

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Table 3. *Statistics for the semantic class features by level of specialization for* of-genitives

Boundary	G^1	G^2	G^3
Number of modifier features	9	31	74
Number head features	9	34	66
Number of feature pairs	63 out of 81	216 out of 1054	314 out of 4,884
Number of features with one relation	26	153	281
Average number of relations per feature	3	1.46	1.14

Table 4. *Statistics for the semantic class features by level of specialization for* s-genitives

Boundary	G^1	G^2	G^3
Number of modifier features	9	37	91
Number head features	9	24	36
No. of feature pairs	41 of 81	157 out of 888	247 out of 3,276
Number of features with one relation	14	99	200
Average number of relations per feature	3.59	1.78	1.36

one feature breaks down into several new specialized features. The net effect is that the semantic relations that were attached to f_{ij}^k will be "scattered" across the new specialized features. This process continues until each feature has only one semantic relation attached. Each iteration creates a new boundary, as shown in Figure 2. Tables 3 and 4 show statistics of semantic features f_{ij}^k for each level of specialization G^k. Note the average number of relations per feature decreases asymptotically to 1 as k increases.

For instance, the feature *entity#1-entity#1* is ambiguous and we need to specialize the semantic class *entity* with its hyponyms (*thing#12, object#1, causal_agent#1*, etc.). After specialization, features are unambiguous or ambiguous. For instance, the feature *thing#12-causal_agent#1* is unambiguous since it encodes the part-whole relation with probability 1. On the other hand, the feature *object#1-causal_agent#1* is ambiguous since it encodes three different relations: the possession relation with 0.5 probability (*apartment of woman*), the depiction relation with 0.4 probability (*photo of the woman*), and the make-produce relation with 0.1 probability (*book of Hemingway*).

The specialization of ambiguous features facilitates the mapping and classification of seen and unseen examples at run time.

2.2 Testing the new boundary

The new boundary is more specific than the previous boundary and it is closer to the ideal boundary. However, we do not know how well it behaves on unseen examples and we are looking for a boundary that classifies the unseen examples with

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

a high accuracy. We test the boundary on unseen examples. For that, we used 10 per cent of the annotated examples (different from the 10 per cent of the examples used for testing) and compute the accuracy (f-measure) of the new boundary on them.

If the accuracy is larger than the previous boundary's accuracy, we are converging toward the best approximation of the boundary and thus we should repeat Step 2 for the new boundary.

If the accuracy is lower than the previous boundary's accuracy, the new boundary is too specific and the previous boundary is a better approximation of the ideal boundary.

The Semantic Scattering model can use pragmatic world knowledge captured by the semantics of the constituents. For instance, with enough examples, the boundary can capture the information that when the head is *work#2* and the modifier is *author#1* the most likely relation is make-produce and can return the correct relation for *Hemingway's book*.

For the automatic detection of the semantic relations encoded by genitives, the boundary constructed by the Semantic Scattering model is more appropriate than a "tree cut," like the one used for verb disambiguation (Li and Abe, 1998; McCarthy, 2001) and constructed using the Minimum Description Length model (Rissanen, 1978).

Semantic Scattering builds only one general model that captures the preferences of an open class of head-modifier pairs toward different relations without building specified "tree cuts" for each head, modifier, or head-modifier pair and without restricting the interpretation to the head or modifier nouns or head-modifier pairs that have such a "tree cut."

4 Semantic relations classification algorithm

The ideal boundary G^* is used for classifying the semantic relations encoded by genitives. The algorithm consists of the following steps:

Step 1. Process the sentence.
Perform WSD and syntactic parsing of the sentence containing the genitive.

Step 2. Identify the head and modifier noun concepts.

Step 3. Identify the feature pair.
Using the results from WSD and WordNet noun hierarchies, map the head and modifier concepts into the corresponding classes from the boundary and identify the feature pair f_{ij} for the two classes.

If there is no such feature (due to the appearance of a new combination of classes in the unseen data), identify the feature that is closer to the corresponding pair of classes: $f_{ij}^k = \text{argmin}\sqrt{(d(f_i,f_i^k))^2 + (d(f_j,f_j^k))^2}$ where $d(f_m,f_n)$ is the number of WordNet hypernymy links between the classes f_m and f_n.

If there is more than one such feature (due to WordNet multiple inheritance) then return all of them.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Step 4. Find the semantic relation.

Using the feature f_{ij}, determine the semantic relation that corresponds to that feature on the boundary.

If there is more than one such relation, return the most probable one ($\text{argmax}_{r \in R} P(r|f_{ij})$).

If there is no such relation, mark it as other.

If there is more than one feature, then return the most probable relation overall.

5 Results

For testing, we considered 20 per cent of the annotated examples. We used half of the examples for detecting the boundary G^* and half for testing the system.

We define the *precision*, *recall*, and *f-measure* performance metrics in this context:

$$\text{Precision} = \frac{\text{Number of correctly retrieved relations}}{\text{Number of relations retrieved}}$$

$$\text{Recall} = \frac{\text{Number of correctly retrieved relations}}{\text{Number of correct relations}}$$

$$\text{f-measure} = \frac{2}{1/\text{Precision} + 1/\text{Recall}}$$

5.1 G^* boundary detection

The algorithm ran iteratively performing boundary specializations on the WordNet is-a noun hierarchies in order to eliminate the ambiguities of the training examples. Boundary G^1 corresponds to the semantic classes of the nine WordNet noun hierarchies and boundaries G^2 and G^3 to their subsequent immediate hyponyms. For both s- and of-genitives, boundary G^2 was more accurate than boundary G^1 and therefore we repeated Step 2. However, boundary G^3 was less accurate than boundary G^2 and thus boundary G^2 is the best approximation of the ideal boundary.

Tables 5 and 6 show the results obtained for different boundaries for of- and s-genitives (G^1, G^2, and G^3), respectively, on the development testing examples and the *Results* on the testing examples (using the best boundary approximation G^2).

5.2 Semantic Relations Classification

Tables 5, 6, and 7 show the results of the application of the Semantic Relations Classification algorithm on 10 per cent of the examples. Overall, the Semantic Scattering system achieves 81.82 per cent precision, 77.88 per cent recall, and 79.80 per cent f-measure for of-genitives and 83.53 per cent precision, 73.96 per cent recall, and 78.45 per cent f-measure for s-genitives, giving a total of 82.61 per cent precision, 76.00 per cent recall, and 79.17 per cent f-measure for genitives.

A per-relation breakdown of the results shows very good performance on the relations with the largest number of training examples for both of-genitives (84.71 per cent for part-whole, 88.89 per cent for measure, 63.64 per cent for theme, and

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Table 5. *The results obtained for different boundaries for* of-genitives

Boundary	G^1	G^2	G^3	Results
Number of correctly retrieved relations	52	74	66	81
Number of relations retrieved	80	99	91	99
Number of correct relations	104	104	104	104
Precision (%)	65.00	74.75	72.53	81.82
Recall (%)	50.00	71.15	63.46	77.88
f-measure (%)	57.50	72.95	68.00	79.80

Table 6. *The results obtained for different boundaries for* s-genitives

Boundary	G^1	G^2	G^3	Results
Number of correctly retrieved relations	19	68	38	71
Number of relations retrieved	93	85	72	85
Number of correct relations	96	96	96	96
Precision (%)	20.43	80.00	52.78	83.53
Recall (%)	19.79	70.83	39.58	73.96
f-measure (%)	20.11	75.42	46.18	78.45

73.68 per cent for property) and s-genitives (81.82 per cent for possession, 72.72 per cent for agent, 83.33 per cent for part-whole, and 66.67 per cent for temporal) while the performance is lower for the relations with fewer training examples.

5.3 Experiments with WSD

Tables 8 and 9 show the results obtained when classifying the 36 relations (the 36th relation being other) for of- and s-genitives for different semantic models that use different types of word senses. The results are presented for the *Final* Semantic Scattering system that uses G^2 as the best approximation of the G^* together with two baselines. The *Baseline1* system obtained the results without WSD features, i.e.

Table 7. *Overall results for the semantic interpretation of genitives*

Results	of-genitives	s-genitives	Genitives
Number of correctly retrieved relations	81	71	152
Number of relations retrieved	99	85	184
Number of correct relations	104	96	200
Precision (%)	81.82	83.53	82.61
Recall (%)	77.88	73.96	76.00
f-measure (%)	79.80	78.45	79.17

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Table 8. *The contribution of WSD in the semantic interpretation of* of-genitives

Baselines	Baseline1 (no WSD)	Baseline2 (manual WSD)	Final (WSD system)
Number of correctly retrieved relations	59	83	81
Number of relations retrieved	75	103	99
Number of correct relations	104	104	104
Precision (%)	76.62	80.58	81.82
Recall (%)	56.73	79.81	77.88
f-measure (%)	65.92	80.19	79.80

Table 9. *The contribution of WSD in the semantic interpretation of* s-genitives

Baselines	Baseline1 (no WSD)	Baseline2 (manual WSD)	Final (WSD system)
Number of correctly retrieved relations	27	73	71
Number of relations retrieved	66	88	85
Number of correct relations	96	96	96
Precision (%)	40.91	82.95	83.53
Recall (%)	28.13	76.04	73.96
f-measure (%)	33.34	79.35	78.45

using only the default sense number 1 for the concept pairs. The *Baseline2* system uses word senses assigned (manually) by human annotators.

WSD is very important for our system as indicated by the Baseline systems' performance. The impact of WSD system on the f-measure (final results minus Baseline1) is 13.88 per cent for of-genitives and 45.11 per cent for s-genitives, while the impact of manual WSD (Baseline2 minus final results) is only 0.39 per cent for of-genitives and 0.9 per cent for s-genitives showing the important contribution of the WSD system to the final results.

5.4 Experiments with different corpora

In order to test the generality of our algorithm, we applied the algorithm for of-genitives on two different corpora: LAT articles from TREC-9 collection and WSJ articles from the Penn Treebank 2 collection. For LAT we used the testing examples and for WSJ we used the 1,396 examples extracted from 20,000 sentences from the WSJ collection.

Table 10 presents the results obtained by the Semantic Scattering method for the LAT and WSJ corpora. We obtained 79.80 per cent f-measure for LAT and 77.61 per cent f-measure for WSJ. Therefore, the results are general enough.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Table 10. *Comparison between the results obtained when using the Semantic Scattering system on LA Times and Wall Street Journal corpora for* of-genitives

Results	LA Times	Wall Street Journal
Number of correctly retrieved relations	81	1,085
Number of relations retrieved	99	1,400
Number of correct relations	104	1,396
Precision (%)	81.82	77.50
Recall (%)	77.88	77.72
f-measure (%)	79.80%	77.61

These experiments indicate that the system shows the same strong performance for different corpora, for different domains, and for different corpus sizes (there are 10 per cent more examples in the WSJ experiments than in the LAT one).

The similar performance of the system on different corpora and for different corpora sizes, also, shows the significance of the results.

5.5 Experiments with different patterns

In order to test the generality of our model, we used the model over *noun compounds*, a more complex construction whose interpretation has proven to be very difficult, and compare the results with the best results in noun compounds interpretation.

For decades, many researchers have tackled the problem and built specific systems (some very advanced) for noun compound interpretation in open domain using different numbers of predicates/relations. They used different supervised or unsupervised, lexical, semantic, syntactic, or probabilistic systems, built and tested on corpora of various sizes (e.g. Grolier Encyclopedia, British National Corpus, Altavista web search results, eXtended WordNet glosses, English Gigaword Corpus, Wall Street Journal, and LA Times articles) with accuracies for commonly cited and recognized results ranging from 40 per cent (Lauer, 1995) to 66.78 per cent (Girju *et al*., 2005).

For a meaningful comparison with the best results on noun compound interpretation (Girju *et al*., 2005), we built a set of 2,120 examples from WSJ articles from the TREC-9 collection, annotated it with the 35 relations of Moldovan *et al*. (2004) and "no relation," and split it in training and testing. Using the 1,908 training examples, we built a Semantic Scattering system for noun compounds and evaluated it on the 212 testing examples. We obtained 64.62 per cent accuracy, 63.13 per cent precision, 60.121 per cent recall, and 61.58 per cent f-measure. The results for noun compounds are not as high as for genitives (Table 10) due to the complexity of the construction and annotation difficulties. However, they compare well with the results of Girju *et al*. (2005), showing a good generality of the model.

5.6 Experiments with clustering

We performed another experiment with the clustering of the examples. According to the semantic analysis of genitives, the of- and s-genitives are not equivalent and

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Table 11. *The contribution of per-construction models to the semantic interpretation of* genitives

Baselines	of-genitives	s-genitives	Combined models	Combined examples
Number of correctly retrieved relations	81	71	152	149
Number of relations retrieved	99	85	184	188
Number of correct relations	104	96	200	200
Precision (%)	81.82	83.53	82.61	79.26
Recall (%)	77.88	73.96	76.00	74.50
f-measure (%)	79.80	78.45	79.17	76.81

therefore we wanted to treat them differently and build separate models for each. However, one may wonder what would happen if only one model was created for all genitives. In these experiments, we wanted to see how the differences between the two genitives would affect a combined model.

For a more accurate comparison, we used a *Combined Models* system that uses both of- and s-genitives models and a *Combined Examples* system that uses one model built from the combined of- and s-genitives examples. For the Combined Examples model, we started from the observation that in an s-genitive the head follows the *"'s"* (NP_{modif} *'s* NP_{head}) and in an of-genitive the head precedes the *"of"* (NP_{head} *of* NP_{modif}). Then, we combined all the examples NP_{head}–NP_{modif} (from both constructions) and built a model using Semantic Scattering.

Table 11 shows the results obtained using the of-genitives system, the s-genitives system, the Combined Models system, and the Combined Examples model. As we can see, the performance of the Combined Examples system is lower than the one of the individual genitive models and the Combined Models systems, as expected considering the divergences between the two types of genitives. However, it still performed well, reflecting the soundness of the Semantic Scattering approach.

5.7 Comparison with other models

To evaluate our model, we conducted experiments with other frequently used machine learning models, on the same dataset, using the same features.

Table 12 shows a comparison between the results obtained with the Semantic Scattering algorithm and Decision Trees (C5.0; Quinlan, 2002), the Naive Bayes model (jBNC—Bayesian network classifier toolbox; Friedman, Geiger and Goldszimidt, 1997), and Support Vector Machines (libSVM with the best-performing RBF kernel $e^{\gamma|x_i - x_j|}$; Change and Lin, 2004) on the same training/evaluation data using two sets of features. *Feature set 1* consisted of the lemmas of the concepts with their corresponding senses. *Feature set 2* consisted of WordNet semantic classes of the concepts. For SVM, to build an accurate vector space, we used a customized Feature set 2 that incorporates all the concepts from the WordNet path from the concept to the WordNet semantic class.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Table 12. *The performance of four learning models on the same testing corpus; *marks the results for the customized set of features*

Model	Feature set 1		Feature set 2	
	of-genitives	s-genitives	of-genitives	s-genitives
Semantic Scattering (SS) (%)	79.80	78.45	79.80	78.45
Decision Trees (DT) (%)	40.60	47.00	31.45	23.51
Naive Bayes (NB) (%)	42.31	43.70	50.55	35.77
Support Vector Machines (SVM) (%)	31.45	23.51	72.13*	70.32*

Table 13. *The error types encountered on the testing corpus*

Error type	of-genitives	s-genitives
Missing features (%)	28.57	29.17
Generality of semantic classes (%)	28.57	20.83
WSD system (%)	19.05	29.17
The reversed order of constituents (%)	14.29	12.50
NERD system (%)	4.76	8.33
Missing WordNet sense (%)	4.76	0.00

The results show that the Semantic Scattering model outperformed all the other models with an absolute gain of 39.2 per cent f-score for decision trees, 29.25 for NB, and 7.62 for SVM for of-genitives and 31.45 for DT, 34.75 for NB, and 8.13 for SVN for s-genitives, respectively.

5.8 Error analysis

An important way of improving the performance of a system is to perform a detailed error analysis of the results. We analyzed the various error sources encountered in our experiments and summarized the results in Table 13.

There are six sources of errors as follows:

WSD system. Our semantic interpretation approach is based on the WordNet semantic classes of the two disambiguated concepts that represent the *modifier* and the *head*, respectively, in genitive constructions. Thus, if the WSD system fails to annotate the noun concepts with the correct senses, the system will generate wrong semantic classes which in turn will most likely lead to errors in classification.

Let us consider the concept *cat*. This noun has eight senses in WordNet corresponding to different types of entities (humans, animals, and objects) or acts. In a genitive construction, different senses determine different semantic relations.

If the concept *cat* is in a modifier position for s-genitives (e.g. *X's cat*) we can have different relations. The most probable relation is the possession relation since animals (senses 1 and 7) and objects (senses 4, 5, and 6) can be possessed. However, other relations can occur. For instance, if *cat* represents a human (senses 2 and 3),

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Table 14. *The semantic relations encoded by the genitive* Mary's cat *for different senses of the concept* cat *in WordNet*

Sense number	Meaning	Semantic relation
1	Feline mammal usually having thick soft fur and being unable to roar; domestic cats; wildcats	possession
2	An informal term for a youth or man	kinship or accompaniment
3	A spiteful woman gossip	kinship
4	The leaves of the shrub Cathaedulis which are chewed like tobacco or used to make tea	possession
5	A whip with nine knotted cords	possession
6	A large vehicle that is driven by caterpillar tracks	possession
7	Any of several large cats typically able to roar and living in the wild	possession
8	A method of examining body organs by scanning them with X-rays and using a computer to construct a series of cross-sectional scans along a single axis	theme or agent

then the relations can be either kinship (if the two persons are closely related) or accompaniment (if the two persons are involved in the same event) if X is a human. In addition, if the *cat* represents a process (the CAT computerized tomography process) then the relation may be theme or agent corresponding to the role played by *X* in the CAT: the patient who gets or the doctor who performs or takes the CAT scan.

Table 14 presents the relations encoded by the genitive *Mary's cat* for different senses of the concept *cat*. Some senses correspond to the same semantic class (e.g. human, animal, and objects) like in the case of senses 1 and 7 which correspond to animals, in which case an error created by the WSD system will not affect the final result as the error is eliminated in the specialization phase. However, when the concept is assigned a sense from a different class, the sense of *cat* is mapped erroneously into a semantic class that encodes different relation, and the error might propagate to the final result.

Missing features. Missing feature pairs in the training corpus may affect the specialization of the boundary and consequently misclassification of the testing examples. For instance, the genitive *acquisition#1 of insurance_company#1* found in the testing corpus was classified as other relation since in the training corpus there was no feature pair *action#1-social_group#1* to which the example could be generalized.

Since we rely heavily on the WSD, any error in this module has a great impact on the classification of genitives.

Generality of the semantic classes. Another cause of error is intrinsic to the model itself. Since the specialization process stops in order not to overshoot the real boundary, some semantic classes are too general and produce errors. For instance, *photo#1 of soccer_ball#1* (a photo depicting a soccer ball) is a depiction-depicted relation. The corresponding semantic feature pair for this genitive is

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

artifact#1-artifact#1. Since the specialization process stopped due to the decrease in the system's accuracy, this feature was not specialized and thus the depiction-depicted relation was discriminated over the relations that maximize the $P(r|f_{ij})$ for the part-whole relation.

The reversed order of constituents. Another cause of error is the incorrect order of the modifier and head nouns in the genitives. As shown in Section 2.1, one-third of the examples are in reverse order while two-thirds are in the modifier–head order. Therefore, the class of examples that are in reversed order contributes only one-third of the examples to the boundary and have a smaller weight and accuracy. For instance, the genitive *chain#1 of hotels#1* that encodes a part-whole relation was misclassified because the "part" concept usually appears in the genitive before the "whole" concept and the model did not captured the reversed order.

Name Entity Recognizer Dictionary (NERD) module. Another source of errors is caused by the NERD module used by the WSD system. The NERD system identifies people, organizations, and other information extraction categories and annotates them accordingly. However, in open-domain applications NERD does not always provide the correct annotation. For example, in the phrase *friend of York*, it tags *York* as a name of a person with sense#1. However, the system will read *York#1*, look in WordNet and sees that the first sense of York is the House of York, the English royal house that reigned from 1,461 to 1,485. Consequently, the system will consider *York#1* a *group* instead of an *entity*, yielding an erroneous result.

Missing WordNet sense. Another class of errors is generated by the fact that some senses of some concepts are not found in WordNet. For example, the concept *import* is not listed in WordNet as denoting the act of importing commodities from a foreign country. Thus the relation *import of sweater* (recipient) is misclassified.

The same problem occurs for the concept *participant* for which WordNet lists only senses corresponding to people and not to other entities such as countries (e.g. a country can be one of the participants at a NATO meeting).

6 Discussion and conclusions

The classification of genitives is an example of a learning problem where a tailored model outperforms other generally applicable models.

This paper presents a model for the semantic classification of genitives. A set of 35 semantic relations was identified and we provided statistical evidence that for genitives some relations are more frequent than others and that the two types of genitive constructions have different semantic information and should be treated differently.

The model relies on the statistical distribution of the semantic classes of noun constituents on a corpus and the semantic and contextual features of the noun constituents.

The algorithm was trained and tested on a corpus of approximately 20,000 sentences containing 2,249 of-genitives and 1,006 s-genitives and achieved an f-measure of 79.80 per cent for of-genitives and 78.45 per cent for s-genitives.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Our experiments with other popular machine learning models built from the same training examples using similar features and tested on the same testing examples assessed that Semantic Scattering outperforms Support Vector Machines (that yield a lower 72.1 per cent for of-genitives and 70.32 per cent for s-genitives), Naive Bayes (that achieves only 50.55 per cent for of-genitives and 43.70 per cent for s-genitives), and Decision Trees (that yield an even lower 40.60 per cent for of-genitives and 47.00 per cent for s-genitives).

Our experiments with WSD and different boundaries also revealed the importance of WSD and WordNet generalization/specialization. We have directly compared and contrasted the results of our system with a variety of baselines and have demonstrated superior results.

Our experiments with corpora also showed the independence of our model from a specific text collection, showing similar good performance on both LAT and WSJ texts.

The remainder of our experiments also show that the two genitive constructions are not necessarily alternative ways of encoding the same sets of relations and that the systems perform well for other patterns such as the very ambiguous noun compounds and therefore can be used for different patterns.

We have also identified the sources of errors which when addressed may bring further improvements.

In addition, we believe that the performance of the system will improve with the number of examples and larger corpora will achieve better results. Also, the accuracy could improve through the use of more complex semantic, syntactic, or pragmatic features such as different properties of the constituent concepts (e.g. functional and animacy), semantic roles of the constituents (nominalizations), syntactic position in sentence (Huddleston and Pullum, 2002), contexts, novelty, semantic preferences, etc.

Also, the method can be improved by using a different boundary "closeness" measure (instead of the Euclidian distance) like lexical chains (Hirst and St-Onge, 1998; Moldovan and Novischi, 2002), similarity measures (Jiang and Conrath, 1997; Leacock and Chodorow, 1998; Maki, McKinley and Thompson, 2004; Cilibrasi and Vitanyi, 2007), or topic signatures (Agirre, Alfonseca and de Lacalle, 2004).

References

Agirre, E., Alfonseca, E., and de Lacalle, O. L. 2004. Approximating hierarchy-based similarity for WordNet Nominal Synsets using topic signatures. In *Proceedings of GWC 2004*, Brno, Czech Republic.

Altenberg, B. 1982. Binominal NPs in a thematic perspective: genitive vs. of-constructions in 17th century English. *Scandinavian Symposium on Syntax Variation*, Stockholm Studies in English, 52, Stockholm: Almqvist & Wiksell.

Badulescu, A. 2004. *Classification of Semantic Relations Between Nouns*. PhD Dissertation, University of Texas at Dallas, Richardson, Texas.

Barker, C. 1995. *Possessive Descriptions*. Stanford: CSLI Publications.

Berland, M., and Charniak, E. 1999. Finding parts in very large corpora. In *Proceeding of ACL*, 1999, College Park, Maryland.

Carreras, X., and Marquez, L. 2004. Introduction to the CoNLL-2004 shared task: semantic role labeling. In *Proceedings of CoNLL-2004*, Boston, Massachusetts.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Carreras, X., and Marquez, L. 2005. Introduction to the CoNLL-2005 shared task: semantic role labeling. In *Proceedings of CoNLL-2005*, Ann Arbor, Michigan.

Chang, C.-C., and Lin, C.-J. 2004. Libsvm: a library for support vector machines. http://www.csie.ntu.edu.tw/cjlin/papers/libsvm.pdf (accessed December 29, 2004).

Chomsky, N. 1986. *Knowledge of Language. Its Nature, Origin and Use*. New York: Praeger.

Cilibrasi, R. L., and Vitanyi, P. M. B. 2007. The google similarity distance. In IEEE/ACM Transactions on Knowledge and Data Engineering, **19**(3):370–383.

Evens, M. 1980. Lexical-semantic relations: a comparative survey. Edmonton, Canada: Linguistic Research, Inc.

Fellbaum, C. 1998. *WordNet—An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Friedman, N., Geiger, D., and Goldszmidt, M. 1997. Bayesian network classifiers. *Machine Learning* **29**(2/3):131–163.

Girju, R., Badulescu, A., and Moldovan, D. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of the Human Language Technology Conference (HLT-NAACL 2003)*, Edmonton, Canada.

Girju, R., Badulescu, A., and Moldovan, D. 2006. Automatic discovery of part-whole relations. *Computational Linguistics* **32**(2):83–135.

Girju, R., Moldovan, D., Tatu, M., and Antohe, D. 2005. On the semantics of noun compounds. *Computer Speech & Language* **19**(4):479–496.

Girju, R., Nakov, P., Nastase, V., Szpakowicz, S., Turney, P., and Yuret, D. 2007. SemEval-2007 task 04: classification of semantic relations between nominals. In *Proceedings of SemEval-2007 Workshop at ACL 2007*, Prague, Czech Republic.

Hearst, M. 1998. Automated discovery of WordNet relations. In C. Fellbaum (ed.), *An Electronic Lexical Database and Some of its Applications*. Cambridge, MA: MIT Press.

Hirst, G., and St-Onge, D. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. In Christiane Fellbaum (ed.), *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Huddleston, R., and Pullum, G. K. 2002. *The Cambridge Grammar of the English Language*. Cambridge, UK: Cambridge University Press.

Jespersen, O. 1949. A modern English grammar on historical principles. In *Syntax* **7**, Copenhagen: Munksgaard.

Jiang, J. J., and Conrath, D. W. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of ROCLING 1997*, Taipei, Taiwan.

Langacker, R. 1990. *Concept, Image, and Symbol. The Cognitive Basis of Grammar*. New York: Berlin and Mouton.

Langacker, R. 1992. The symbolic nature of cognitive grammar: the meaning of of and of-periphrasis. In M. Putz (ed.), *Thirty Years of Linguistic Evolution*, Amsterdam: John Benjamins.

Langacker, R. 1993. Reference-point constructions. *Cognitive Linguistics* **4**:1–38.

Langacker, R. 1995. Possession and possessive constructions. In J. Taylor and R. MacLaury (eds.), *Language and the Cognitive Construal of the World*, Berlin: Mouton de Gruyter.

Lapata, M. 2002. The disambiguation of nominalisations. *Computational Linguistics* **28**(3):357–388.

Lauer, M. 1995. *Designing Statistical Language Learners: Experiments on Noun Compounds*. PhD Thesis, Macquarie University, Australia.

Leacock, C., and Chodorow, M. 1998. Combining local context and WordNet similarity for word sense identification. In C. Fellbaum (ed.), *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Li, H., and Abe, N. 1998. Generalizing case frames using a thesaurus and the MDL principle. *Computational Linguistics* **24**(2):217–224.

Light, M., and Greiff, W. 2002. Statistical models for the induction and use of selectional preferences. *Cognitive Science* **87**:1–13.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press

Litkowski, K. 2004. Senseval-3 task. Automatic labeling of semantic roles. In *Proceedings of Senseval 3*, Barcelona, Spain.

Maki, W. S., McKinley, L. N., and Thompson A. G. 2004. Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments & Computers, Web-Based Archive of Norms, Stimuli, and Data* **36**(3):421–431.

McCarthy, D. 2001. *Lexical Acqusition at the Syntax–Semantics Interface: Diathesis Alternations, Subcategorization Frames and Selectional Preferences*. PhD Dissertation, Univesity of Sussex.

Moldovan, D., and Badulescu, A. 2005. A Semantic Scattering model for the automatic interpretation of genitives. In *Proceedings of the Human Language Technology Conference (HLT-NAACL) 2006*, Vancouver, Canada.

Moldovan, D., Badulescu, A., Tatu, M., Antohe, D., and Girju, R. 2004. Models for the semantic classification of noun phrases. In *Proceedings of the Human Language Technology Conference (HLT-NAACL) 2004, Computational Lexical Semantics Workshop*, Boston, Massachusetts.

Moldovan, D., Clark, C., and Bowden, M. 2007. Lymba's PowerAnswer 4 in TREC 2007. In *Proceedings of TREC 2007 Conference*, Gaithersburg, Maryland.

Moldovan, D., and Novischi, A. 2002. Lexical chains for question answering. In *Proceedings of COLING 2002*, Taipei, Taiwan.

Nikiforidou, K. 1991. The meanings of the genitive: a case study in the semantic structure and semantic change. *Cognitive Linguistics* **2**(149):149–205.

Novischi, A., Moldovan, D., Parker, P., Badulescu, A., and Hauser, B. 2004. LCC's WSD systems for senseval 3. In *Proceedings of Senseval 3*, Barcelona, Spain.

Partee, B., and Borschev, V. 1999. Possessives, favorite, and coercion. In *Proceedings of ESCOL99*, Ithaca, New York.

Quinlan, R. 2002. Data mining tools see5 and c5.0. http://www.rulequest.com/see5-info.html (accessed December 29, 2004).

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. 1985. *A Comprehensive Grammar of English Language*. Harlow, England: Longman.

Rissanen, J. 1978. Modeling by shortest data description. *Automatica* **14**:149–205.

Rosario, B., Hearst, M., and Fillmore, C. 2002. The descendent of hierarchy and selection in relational semantics. In *Proceedings of ACL 2002*, Philadelphia, Pennsylvania.

Siegel, S., and Castellan, N. J. 1988. *Non Parametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Stefanowitsch, A. 2001. Constructional semantics as a limit to grammatical alternation: two genitives of English. In G. Rohdenburg and B. Mondorf (eds.), *Determinants of Grammatical Variation in English*, Berlin: Mouton de Gruyter.

Strang, B. 1962. *Modern English Structure*. London: Edward Arnold.

Taylor, J. 1996. *Possessives in English. An Exploration in Cognitive Grammar*. Oxford: Clarendon Press.

Vikner, C., and Jensen, P. A. 1999. *A Semantic Analysis of the English Genitive: Interaction of Lexical and Formal Semantics*. Denmark: Ms. Copenhagen and Kolding.

Williams, E. 1982. The NP cycle. *Linguistic Inquiry* **13**:277–295.

https://doi.org/10.1017/S1351324908004798 Published online by Cambridge University Press