Identifying Concept Attributes Using a Classifier

Massimo Poesio
University of Essex
Computer Science /
Language and Computation
poesio at essex.ac.uk

Abdulrahman Almuhareb
University of Essex
Computer Science /
Language and Computation
aalmuh at essex.ac.uk

Abstract

We developed a novel classification of concept attributes and two supervised classifiers using this classification to identify concept attributes from candidate attributes extracted from the Web. Our binary (attribute / non-attribute) classifier achieves an accuracy of 81.82% whereas our 5-way classifier achieves 80.35%.

1 Introduction

The assumption that concept **attributes** and, more in general, **features**[1] are an important aspect of conceptual representation is widespread in all disciplines involved with conceptual representations, from Artificial Intelligence / Knowledge Representation (starting with at least (Woods, 1975) and down to (Baader et al, 2003)), Linguistics (e.g., in the theories of the lexicon based on typed feature structures and/or Pustejovsky's Generative Lexicon theory: (Pustejovsky 1995)) and Psychology (Murphy 2002, Vinson et al 2003). This being the case, it is surprising how little attention has been devoted to this aspect of lexical representation in work on large-scale lexical semantics in Computational Linguistics. The most extensive resource at our disposal, WordNet (Fellbaum, 1998) contains very little information that would be considered as being about 'attributes'—only information about parts, not about qualities such as **height**, or even to the values of such attributes in the adjective network—and this information is still very sparse. On the other hand, the only work on the extraction of lexical semantic relations we are aware of has concentrated on the type of relations found in WordNet: hyponymy (Hearst, 1998; Caraballo, 1999) and meronymy (Berland and Charniak, 1999; Poesio et al, 2002).[2]

The work discussed here could be perhaps best described as an example of **empirical ontology**: using linguistics and philosophical ideas to improve the results of empirical work on lexical / ontology acquisition, and vice versa, using findings from empirical analysis to question some of the assumptions of theoretical work on ontology and the lexicon. Specifically, we discuss work on the acquisition of (nominal) concept attributes whose goal is twofold: on the one hand, to clarify the notion of 'attribute' and its role in lexical semantics, if any; on the other, to develop methods to acquire such information automatically (e.g., to supplement WordNet).

The structure of the paper is as follows. After a short review of relevant literature on extracting semantic relations and on attributes in the lexicon, we discuss our classification of attributes, followed by the features we used to classify them. We then discuss our training methods and the results we achieved.

[1] The term **attribute** is used informally here to indicate the type of relational information about concepts that is expressed using so-called **roles** in Description Logics (Baader et al, 2003)—i.e., excluding **IS-A** style information (that cars are vehicles, for instance). It is meant to be a more restrictive term than the term **feature**, often used to indicate any property of concepts, particularly in Psychology. We are carrying out a systematic analysis of the sets of features used in work such as (Vinson et al, 2003) (see Discussion).

[2] In work on the acquisition of lexical information about verbs there has been some work on the acquisition of thematic roles, (e.g., Merlo and Stevenson, 2001).

Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition, pages 18–27,
Ann Arbor, June 2005.

2 Background

2.1 Using Patterns to Extract Semantic Relations

The work discussed here belongs to a line of research attempting to acquire information about lexical and other semantic relations other than similarity / synonymy by identifying **syntactic constructions** that are often (but not always!) used to express such relations. The earliest work of this type we are aware of is the work by Hearst (1998) on acquiring information about hyponymy (= **IS-A** links) by searching for instances of patterns such as

`NP {, NP}* or other NP`

(as in, e.g., *bruises …. broken bones and other INJURIES*). A similar approach was used by Berland and Charniak (1999) and Poesio et al (2002) to extract information about **part-of** relations using patterns such as

`the N of the N is ….`

(as in *the wheel of the CAR is*) and by Girju and Moldovan (2002) and Sanchez-Graillet and Poesio (2004) to extract causal relations. In previous work (Almuhareb and Poesio, 2004) we used this same approach to extract attributes, using the pattern

`"the * of the C [is|was]"`

(suggested by, e.g., (Woods, 1975) as a test for 'attributehood') to search for attributes of concept C in the Web, using the Google API. Although the information extracted this way proved a useful addition to our lexical representations from a clustering perspective, from the point of view of lexicon building this approach results in too many false positives, as very few syntactic constructions are used to express exclusively one type of semantic relation. For example, the 'attributes' of **deer** extracted using the text pattern above include "the *majority* of the deer," "the *lake* of the deer," and "the *picture* of the deer." Girju and Moldovan (2002) addressed the problem of false positives for causal relations by developing WordNet-based *filters* to remove unlikely candidates. In this work, we developed a semantic filter for attributes based on a linguistic theory of attributes which does not rely on WordNet except as a source of morphological information (see below).

2.2 Two Theories of Attributes

The earliest attempt to classify attributes and other properties of substances we are aware of goes back to Aristotle, e.g., in *Categories*,[3] but our classification of attributes was inspired primarily by the work of Pustejovsky (1995) and Guarino (e.g., (1992)). According to Pustejovsky's *Generative Lexicon* theory (1995), an integral part of a lexical entry is its **Qualia Structure**, which consists of four 'roles':[4] the **Formal Role**, specifying what type of object it is: e.g., in the case of a book, that it has a shape, a color, etc.; the **Constitutive Role**, specifying the stuff and parts that it consists of (e.g., in the case of a book, that it is made of paper, it has chapters and an index, etc.); the **Telic Role**, specifying the purpose of the object (e.g., in the case of a book, *reading*); and the **Agentive Role**, specifying how the object was created (e.g., in the case of a book, by *writing*).

Guarino (1992) argues that there are two types of attributes: **relational** and **non-relational**. Relational attributes include **qualities** such as *color* and *position*, and **relational social roles** such as *son* and *spouse*. Non-relational attributes include **parts** such as *wheel* and *engine*. Activities are not viewed as attributes in Guarino's classification.

3 Attribute Extraction and Classification

The goal of this work is to identify genuine attributes by classifying candidate attributes collected using text patterns as discussed in (Almuhareb and Poesio, 2004) according to a scheme inspired by those proposed by Guarino and Pustejovsky.

The scheme we used to classify the training data in the experiment discussed below consists of six categories:

- **Qualities**: Analogous to Guarino's qualities and Pustejovsky's formal 'role'. (E.g., "the *color* of the car".)
- **Parts:** Related to Guarino's non-relational attributes and Pustejovsky's constitutive 'roles'. (E.g., "the *hood* of the car").
- **Related-Objects:** A new category introduced to cover the numerous physical objects which are 'related' to an object but are not part of it—e.g., "the *track* of the deer".

[3] E.g., http://plato.stanford.edu/entries/substance. Thanks to one of the referees for drawing our attention to this.

[4] 'Facets' would be perhaps a more appropriate term to avoid confusions with the use of the term 'role' in Knowledge Representation.

- **Activities:** These include both the types of activities which are part of Pustejovsky's telic 'role' and those which would be included in his agentive 'role'. (E.g., "the *repairing* of the car".)
- **Related-Agents:** For the activities in which the concept in question is acted upon, the agent of the activity: e.g., "the *writer* of the book", "the *driver* of the car".
- **Non-Attributes:** This category covers the cases in which the construction "the N of the N" expresses other semantic relations, as in: "the *last* of the deer", "the *majority* of the deer," "the *lake* of the deer," and "in the *case* of the deer".

We will quickly add that (i) we do not view this classification as definitive—in fact, we already collapsed the classes 'part' and 'related objects' in the experiments discussed below—and (ii) not all of these distinctions are very easy even for human judges to do. For example, *design*, as an attribute of a *car*, can be judged to be a quality if we think of it as taking values such as *modern* and *standard*; on the other hand, *design* might also be viewed as an activity in other contexts discussing the designing process. Another type of difficulty is that a given attribute may express different things for different objects. For example, *introduction* is a part of a *book*, and an activity for a *product*. An additional difficulty results from the strong similarity between parts and related-objects. For example, "key" is a related-object to a *car* but it is not part of it. We will return to this issue and to agreement on this classification scheme when discussing the experiment.

One difference from previous work is that we use additional linguistic constructions to extract candidate attributes. The construction "the X of the Y is" used in our previous work is only one example of genitive construction. Quirk *et al* (1985) list eight types of genitives in English, four of which are useful for our purposes:

- *Possessive Genitive*: used to express qualities, parts, related-objects, and related-agents.
- *Genitive of Measure*: used to express qualities.
- *Subjective & Objective Genitives*: used to express activities.

We used all of these constructions in the work discussed here.

4 Information Used to Classify Attributes

Our attribute classifier uses four types of information: *morphological information*, an *attribute model*, a *question model*, and an *attributive-usage model*. In this section we discuss how this information is automatically computed.

4.1 Morphological Information

Our use of morphological information is based on the noun classification scheme proposed by Dixon (1991). According to Dixon, derivational morphology provides some information about attribute type. Parts are concrete objects and almost all of them are expressed using basic noun roots (i.e., not derived from adjectives or verbs). Most of qualities and properties are either basic noun roots or derived from adjectives. Finally, activities are mostly nouns derived from verbs. Although these rules only have a heuristic value, we found that morphologically based heuristics did provide useful cues when used in combination with the other types of information discussed below.

As we are not aware of any publicly available software performing automatic derivational morphology, we developed our own (and very basic) heuristic methods. The techniques we used involve using information from WordNet, suffix-checking, and a POS tagger.

WordNet was used to find nouns that are derived from verbs and to filter out words that are not in the noun database. Nouns in WordNet are linked to their derivationally related verbs, but there is no indication about which is derived from which. We use a heuristic based on length to decide this: the system checks if the noun contains more letters than the most similar related verb. If this is the case, then the noun is judged to be derived from the verb. If the same word is used both as a noun and as a verb, then we check the usage familiarity of the word, which can also be found in WordNet. If the word is used more as a verb and the verbal usage is not rare, then again the system treats the noun as derived from the verb.

To find nouns that are derived from adjectives we used simple heuristics based on suffix-checking. (This was also done by Berland and Charniak (1999).) All words that end with "*ity*" or "*ness*" are considered to be derived from adjectives. A noun not found to be derived from a verb or an adjective is assumed to be a basic noun root.

In addition to derivational morphology, we used the Brill tagger (Brill, 1995) to filter out adjectives and other types of words that can occasionally be used as nouns such as *better*, *first*, and *whole* before training. Only nouns, base form verbs, and gerund form verbs were kept in the candidate attribute list.

4.2 Clustering Attributes

Attributes are themselves concepts, at least in the sense that they have their own attributes: for example, a part of a car, such as a wheel, has its own parts (the tyre) its qualities (weight, diameter) etc. This observation suggests that it should be possible to find similar attributes in an unsupervised fashion by looking at their attributes, just as we did earlier for concepts (Almuhareb and Poesio, 2004). In order to do this, we used our text patterns for finding attributes to collect from the Web up to 500 pattern instances for each of the candidate attributes. The collected data were used to build a vectorial representation of attributes as done in (Almuhareb and Poesio, 2004). We then used CLUTO (Karypis, 2002) to cluster attributes using these vectorial representations. In a first round of experiments we found that the classes 'parts' and 'related objects' were difficult to differentiate, and therefore we merged them. The final model clusters candidate attributes into five classes: activities, parts & related-objects, qualities, related-agents, and non-attributes. This classification was used as one of the input features in our supervised classifier for attributes.

We also developed a measure to identify particularly distinctive 'attributes of attributes'—attributes which have a strong tendency to occur primarily with attributes (or any concept) of a given class—which has proven to work pretty well. This measure, which we call *Uniqueness*, actually is the product of two factors: the degree of uniqueness proper, i.e., the probability $P(class_i \mid attribute_j)$ that an attribute (or, in fact, any other noun) will belong to class i given than it has attribute j; and a measure of 'definitional power' –the probability $P(attribute_j \mid class_i)$ that a concept belonging to a given class will have a certain attribute. Using MLE to estimate these probabilities, the degree of uniqueness of $attributes_j$ of $class_i$ is computed as follows:

$$Uniqueness_{i,j} = \frac{C(class_i, attribute_j)^2}{n_i \times C(attribute_j)}$$

where $\boldsymbol{n_i}$ is the number of concepts in $class_i$. $\boldsymbol{C}$ is a count function that counts concepts that are associated with the given attribute. Uniqueness ranges from 0 to 1.

Table 1 shows the 10 most distinctive attributes for each of the five attribute classes, as determined by the Uniqueness measure just introduced, for the 1,155 candidate attributes in the training data for the experiment discussed below.

Class	Top 10 Distinctive Attributes
Related-Agent (0.39)	identity, hands, duty, consent, responsibility, part, attention, voice, death, job
Part & Related-Object (0.40)	inside, shape, top, outside, surface, bottom, center, front, size, interior
Activity (0.29)	time, result, process, results, timing, date, effect, beginning, cause, purpose
Quality (0.23)	measure, basis, determination, question, extent, issue, measurement, light, result, increase
Non-Attribute (0.18)	content, value, rest, nature, meaning, format, interpretation, essence, size, source

Table 1: Top 10 distinctive attributes of the five classes of candidate attributes. Average distinctiveness (uniqueness) for the top 10 attributes is shown between parentheses

Most of the top 10 attributes of related-agents, parts & related-objects, and activities are genuinely distinctive attributes for such classes. Thus, attributes of related-agents reflect the 'intentionality' aspect typical of members of this class: *identity*, *duty*, and *responsibility*. Attributes of parts are common attributes of physical objects (e.g., *inside*, *shape*). Most attributes of activities have to do with temporal properties and causal structure: e.g., *beginning*, *cause*. The 'distinctive' attributes of the

quality class are less distinctive, but four such attributes (*measure*, *extent*, *measurement*, and *increase*) are related to values since many of the qualities can have different values (e.g., *small* and *large* for the quality *size*). There are however several attributes in common between these classes of attributes, emphasizing yet again how some of these distinctions at least are not completely clear cut: e.g., *result*, in common between activities and qualities (two classes which are sometimes difficult to distinguish). Finally, as one would expect, the attributes of the non-attribute class are not really distinctive: their average uniqueness score is the lowest. This is because 'non-attribute' is a heterogeneous class.

4.3 The Question Model

Certain types of attributes can only be used when asking certain types of questions. For example, it is possible to ask "*What is the color of the car?*" but not "**When is the color of the car?*".

We created a text pattern for each type of question and used these patterns to search the Web and collect counts of occurrences of particular questions. An example of such patterns would be:

- *"what is|are the A of the"*

where *A* is the candidate attribute under investigation. Patterns for *who*, *when*, *where*, and *how* are similar.

After collecting occurrence frequencies for all the candidate attributes, we transform these counts into weights using the *t*-test weighting function as done for all of our counts, using the following formula from Manning and Schuetze (1999):

$$t_{i,j} \approx \frac{\frac{C(question_i, attribute_j)}{N} - \frac{C(question_i) \times C(attribute_j)}{N^2}}{\sqrt{\frac{C(question_i, attribute_j)}{N^2}}}$$

where ***N*** is the total number of relations, and ***C*** is a count function.

Table 2 shows the 10 most frequent attributes for each question type. This data was collected using a more restricted form of the question patterns and a varying number of instances for each type of questions. The restricted form includes a question mark at the end of the phrase and was used to improve the precision. For example, the *what*-pattern would be *"what is the * of the *?"*.

Question	Top 10 Attributes
what	purpose, name, nature, role, cost, function, significance, size, source, status
who	author, owner, head, leader, president, sponsor, god, lord, father, king
where	rest, location, house, fury, word, edge, center, end, ark, voice
how	quality, rest, pace, level, length, morale, performance, content, organization, cleanliness
when	end, day, time, beginning, date, onset, running, birthday, fast, opening

Table 2: Frequent attributes for each question type

Instances of the *what*-pattern are frequent in the Web: the Google count was more than 2,000,000 for a query issued in mid 2004. The *who*-pattern is next in terms of occurrence, with about 350,000 instances. The *when*-pattern is the most infrequent pattern, about 5,300 instances.

The counts broadly reflected our intuitions about the use of such questions. W*hat*-questions are mainly used with qualities, whereas *who*-questions are used with related-agents. Attributes occurring with *when*-questions have some temporal aspects; attributes occurring with *how*-questions are mostly qualities and activities, and attributes in *where*-questions are of different types but some are related to locations. Parts usually do not occur with these types of questions.

4.4 Attributive Use

Finally, we exploited the fact that certain types of attributes are used more in language as concepts rather than as attributes. For instance, it is more common to encounter the phrase *"the size of the *"* than *"the * of the size"*. On the other hand, it is more common to encounter the phrase *"the * of the window"* than *"the window of the *"*. Generally speaking, parts, related-objects, and related-agents are more likely to have more attributes than qualities and activities. We used the two patterns "the * of the A" and "the A of the *" to collect Google counts for all of the candidate attributes. These counts were also weighted using the *t*-test as in the question model.

Table 3 illustrates the attributive and conceptual usage for each attribute class using a training data of 1,155 attributes. The usage averages confirm the initial assumption.

Attribute Class	Average *T*-Test Score	
	Conceptual	Attributive
Parts & Related-Objects	18.81	3.00
Non-Attributes	13.29	11.07
Related-Agents	12.15	2.54
Activities	3.22	5.08
Qualities	0.23	17.09

Table 3: Conceptual and attributive usage averages for each attribute class

5 The Experiment

We trained two classifiers: a 2-way classifier that simply classifies candidate attributes into attributes and non-attributes, and a 5-way classifier that classifies candidate attributes into activities, parts & related-objects, qualities, related-agents, and non-attributes. These classifiers were trained using decision trees algorithm (J48) from WEKA (Witten and Frank, 1999).

Feature	election	abdomen	acidity	creator	problem
Cluster Id	1	2	4	0	3
What	0.00	0.00	0.00	0.00	3.80
When	2.62	0.00	0.00	0.00	0.00
Where	0.78	0.94	0.00	0.00	0.00
Who	0.00	0.00	0.00	30.28	0.00
How	2.05	0.00	1.54	0.00	2.61
Conceptual	38.16	20.15	0.00	0.00	135.40
Attributive	0.00	0.00	10.22	1.60	0.00
Morph	DV	BN	DA	DV	BN
Attribute Class (Output)	Activity	Part	Quality	Related Agent	Non-Attribute

Table 4: Five examples of training instances. The values for **morph** are as follows: DV: derived from verb; BN: basic noun; DA: derived from adjective

Our training and testing material was acquired as follows. We started from the 24,178 candidate attributes collected for the concepts in the balanced concept dataset we recently developed (Almuhareb and Poesio, 2005). We threw out every candidate attribute with a Google frequency less than 20; this reduced the number of candidate attributes to 4,728. We then removed words other than nouns and gerunds as discussed above, obtaining 4,296 candidate attributes.

The four types of input features for this filtered set of candidate attributes were computed as discussed in the previous section. The best results were obtained using all of these features. A training set of 1,155 candidate attributes was selected and hand-classified (see below for agreement figures). We tried to include enough samples for each attribute class in the training set. Table 4 shows the input features for five different training examples, one for each attribute class.

6 Evaluation

For a qualitative idea of the behavior of our classifier, the best attributes for some concepts are listed in Appendix A. We concentrate here on quantitative analyses.

6.1 Classifier Evaluation 1: Cross-Validation

Our two classifiers were evaluated, first of all, using 10-fold cross-validation. The 2-way classifier correctly classified 81.82% of the candidate attributes (the baseline accuracy is 80.61%). The 5-way classifier correctly classified 80.35% of the attributes (the baseline accuracy is 23.55%). The precision / recall results are shown in Table 5.

Attribute Class	P	R	F
2-Way Classifier			
Attribute	0.854	0.934	0.892
Non-Attribute	0.551	0.335	0.417
5-Way Classifier			
Related-Agent	0.930	0.970	0.950
Part & Related-Object	0.842	0.882	0.862
Activity	0.822	0.878	0.849
Quality	0.799	0.821	0.810
Non-Attribute	0.602	0.487	0.538

Table 5: Cross-validation results for the two attribute classifiers

As it can be seen from Table 5, both classifiers achieve good *F* values for all classes except for the non-attribute class: *F*-measures range from 81% to 95%. With the 2-way classifier, the valid attribute class has an *F*-measure of 89.2%. With the 5-way classifier, **related-agent** is the most accurate class (*F* = 95%) followed by **part & related-object**, **activity**, and **quality** (86.2%, 84.9%, and 81.0%,

respectively). With **non-attribute,** however, we find an *F* of 41.7% in the 2-way classification, and 53.8% in the 5-way classification. This suggests that the best strategy for lexicon building would be to use these classifiers to 'find' attributes rather than 'filter' non-attributes.

6.2 Classifier Evaluation 2: Human Judges

Next, we evaluated the accuracy of the attribute classifiers against two human judges (the authors). We randomly selected a concept from each of the 21 classes in the balanced dataset. Next, we used the classifiers to classify the 20 best candidate attributes of each concept, as determined by their *t*-test scores. Then, the judges decided if the assigned classes are correct or not. For the 5-way classifier, the judges also assigned the correct class if the automatic assigned class is incorrect.

After a preliminary examination we decided not to consider two troublesome concepts: *constructor* and *future*. The reason for eliminating c*onstructor* is that we discovered it is ambiguous: in addition to the sense of 'a person who builds things', we discovered that *constructor* is used widely in the Web as a name for a fundamental *method* in object oriented programming languages such as Java. Most of the best candidate attributes (e.g., *call*, *arguments*, *code*, and *version*) related to the latter sense, that doesn't exist in WordNet. Our system is currently not able to do word sense discrimination, but we are currently working on this issue. The reason for ignoring the concept *future* was that this word is most commonly used as a modifier in phrases such as: "*the car of the future*", and "*the office of the future"*, and that all of the best candidate attributes occurred in this type of construction. This reduced the number of evaluated concepts to 19.

According to the judges, the 2-way classifier was on average able to correctly assign attribute classes for 82.57% of the candidate attributes. This is very close to its performance in evaluation 1. The results using the *F*-measure reveal similar results too. Table 6 shows the results of the two classifiers based on the precision and recall measures.

According to the judges, the 5-way classifier correctly classified 68.72% on average. This performance is good but not as good as its performance in evaluation 1 (80.35%). The decrease in the performance was also shown in the *F*-measure. The *F*-measure ranges from 0.712 to 0.839 excluding the non-attribute class.

Attribute Class	**P**	**R**	**F**
2-Way Classifier			
Attribute	0.928	0.872	0.899
Non-Attribute	0.311	0.459	0.369
5-Way Classifier			
Related-Agent	0.813	0.868	0.839
Part & Related-Object	0.814	0.753	0.781
Activity	0.870	0.602	0.712
Quality	0.821	0.658	0.730
Non-Attribute	0.308	0.632	0.414

Table 6: Evaluation against human judges results for the two classifiers

An important question when using human judges is the degree of agreement among them. The *K*-statistic was used to measure this agreement. The values of *K* are shown in Table 7. In the 2-way classification, the judges agreed on 89.84% of the cases. On the other hand, the *K*-statistic for this classification task is 0.452. This indicates that part of this strong agreement is because that the majority of the candidate attributes are valid attributes. It also shows the difficulty of identifying non-attributes even for human judges. In the 5-way classification, the two judges have a high level of agreement; Kappa statistic is 0.749. The judges and the 5-way classifier agreed on 63.71% of the cases.

Description	**2-Way**	**5-Way**
Human Judges	89.84%	80.69%
Human Judges (Kappa)	0.452	0.749
Human Judges & Classifier	78.36%	63.71%

Table 7: Level of agreement between the human judges and the classifiers

6.3 Re-Clustering the Balanced Dataset

Finally, we looked at whether using the classifiers results in a better lexical description for the purposes of clustering (Almuhareb and Poesio, 2004). In Table 8 we show the results obtained using the output of the 2-way classifier to re-cluster the 402 concepts of our balanced dataset, comparing these results with those obtained using all attributes (first column) and all attributes that remain after frequency cutoff and POS filtering (column 2). The results are based on the CLUTO evaluation meas-

ures: *Purity* (which measures the degree of cohesion of the clusters obtained) and *Entropy*. The purity and entropy formulas are shown in Table 9.

Description	All Candidate Attributes	Filtered Candidate Attributes	2-Way Attributes
Purity	0.657	0.672	0.693
Entropy	0.335	0.319	0.302
Vector Size	24,178	4,296	3,824

Table 8: Results of re-clustering concepts using different sets of attributes

Clustering the concepts using only filtered candidate attributes improved the clustering purity from 0.657 to 0.672. This improvement in purity is not significant. However, clustering using only the attributes sanctioned by the 2-way classifier improved the purity further to 0.693, and this improvement in purity from the initial purity was significant ($t = 2.646$, $df = 801$, $p < 0.05$).

	Entropy	Purity
Single Cluster	$E(S_r) = -\frac{1}{\log q}\sum_{i=1}^{q}\frac{n_r^i}{n_r}\log\frac{n_r^i}{n_r}$	$P(S_r) = \frac{1}{n_r}\max_i(n_r^i)$
Overall	$Entropy = \sum_{r=1}^{k}\frac{n_r}{n}E(S_r)$	$Purity = \sum_{r=1}^{k}\frac{n_r}{n}P(S_r)$

Table 9: Entropy and Purity in CLUTO.
S_r is a cluster, n_r is the size of the cluster, q is the number of classes, n^i_r is the number of concepts from the ith class that were assigned to the rth cluster, n is the number of concepts, and k is the number of clusters.

7 Discussion and Conclusions

The lexicon does not simply contain information about synonymy and hyponymy relations; it also contains information about the attributes of the concepts expressed by senses, as in Qualia structures. In previous work, we developed techniques for mining candidate attributes from the Web; in this paper we presented a method for improving the quality of attributes thus extracted, based on a classification for attributes derived from work in linguistics and philosophy, and a classifier that automatically tags candidate attributes with such classes. Both the 2-way and the 5-way classifiers achieve good precision and recall. Our work also reveals, however, that the notion of attribute is not fully understood. On the one hand, that attribute judgments are not always easy for humans even given a scheme; on the other hand, the results for certain types of attributes, especially activities and qualities, could certainly be improved. We also found that whereas attributes of physical objects are relatively easy to classify, the attributes of other types of concepts are harder –particularly with activities. (See the Appendix for examples.) Our longer term goal is thus to further clarify the notion of attribute, possibly refining our classification scheme, in collaboration with linguists, philosophers, and psycholinguists. One comparison we are particularly interested in pursuing at the moment is that with feature lists used by psychologist, for whom knowledge representation is entirely concept-based, and virtually every property of a concept counts as an attribute, including properties that would be viewed as **IS-A** links and what would be considered a value. Is it possible to make a principled, yet cognitively based distinction?

Acknowledgments

Abdulrahman Almuhareb is supported by King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia. We wish to thank the anonymous referees for many helpful suggestions.

References

Almuhareb, A. and Poesio, M. (2004). Attribute-Based and Value-Based Clustering: An Evaluation. In *Proc. of EMNLP*. Barcelona, July.

Almuhareb, A. and Poesio, M. (2005). Concept Learning and Categorization from the Web. In *Proc. of CogSci*. Italy, July.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D. and Patel-Schneider, P. (Editors). (2003). *The Description Logic Handbook*. Cambridge University Press.

Berland, M. and Charniak, E. (1999). Finding parts in very large corpora. In *Proc. of the 37th ACL*, (pp. 57–64). University of Maryland.

Brill, E. (1995). Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging. *Computational Linguistics*.

Caraballo, S. A. (1999). Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proc. of the 37th ACL*.

Dixon, R. M. W. (1991). *A New Approach to English Grammar, on Semantic Principles*. Clarendon Press, Oxford.

Fellbaum, C. (Editor). (1998). *WordNet: An electronic lexical database*. The MIT Press.

Girju, R. and Moldovan, D. (2002). Mining answers for causal questions. In *Proc. AAAI*.

Guarino, N. (1992). Concepts, attributes and arbitrary relations: some linguistic and ontological criteria for structuring knowledge base. *Data and Knowledge Engineering*, 8, (pp. 249–261).

Hearst, M. A. (1998). Automated discovery of WordNet relations. In Fellbaum, C. (Editor). *WordNet: An Electronic Lexical Database*. MIT Press.

Karypis, G. (2002). *CLUTO: A clustering toolkit. Technical Report 02-017*. University of Minnesota. At http://www-users.cs.umn.edu/~karypis/cluto/.

Manning, C. D. and Schuetze H. (1999). *Foundations of Statistical NLP*. MIT Press.

Merlo, P. and Stevenson, S. (2001). Automatic Verb Classification Based on Statistical Distributions of Argument Structure. *Computational Linguistics*. 27: 3, 373-408.

Murphy, G. L. (2002). *The Big Book of Concepts*. The MIT Press.

Poesio, M., Ishikawa, T., Schulte im Walde, S. and Vieira, R. (2002). Acquiring lexical knowledge for anaphora resolution. In *Proc. Of LREC*.

Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Sanchez-Graillet, O. and Poesio, M. (2004). Building Bayesian Networks from text. In *Proc. of LREC*, Lisbon, May.

Vinson, D. P., Vigliocco, G., Cappa, S., and Siri, S. (2003). The breakdown of semantic knowledge: insights from a statistical model of meaning representation. *Brain and Language*, 86(3), 347-365(19).

Witten, I. H. and Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann.

Woods, W. A. (1975). What's in a link: Foundations for semantic networks. In Daniel G. Bobrow and Alan M. Collins, editors, *Representation and Understanding: Studies in Cognitive Science*, (pp. 35-82). Academic Press, New York.

Appendix A. 5-Way Automatic Classification of the Best Candidate Attributes of Some Concepts

Car

Class	Best Attributes
Activity	acceleration, performance, styling, construction, propulsion, insurance, stance, ride, movement
Part & Related-Object	front, body, mass, underside, hood, roof, nose, graphics, side, trunk, engine, boot, frame, bottom, backseat, chassis, wheelbase, silhouette, floor, battery, windshield, seat, undercarriage, tank, window, steering, drive, finish
Quality	speed, weight, handling, velocity, color, condition, width, look, colour, feel, momentum, heritage, shape, appearance, ownership, make, convenience, age, quality, reliability
Related-Agent	driver, owner, buyer, sponsor, occupant, seller
Non-Attribute	rest, price, design, balance, motion, lure, control, use, future, cost, inertia, model, wheel, style, position, setup, sale, supply, safety

Camel

Class	Best Attributes
Activity	introduction, selling, argument, exhaustion
Part & Related-Object	nose, hump, furniture, saddle, hair, flesh, neck, milk, head, reins, foot, eye, hooves, humps, ass, feet, hoof, flanks, bones, ears, bag, skin, haunches, stomach, legs, urine, meat, penis, load, breast, backside, testicles, rope, corpse, house, nostrils, foam, bell, sight, butt, fur, bodies, toe, hoofs, heads, knees, pancreas, mouth, coat, uterus, necks, chin, udders
Quality	origins, gait, domestication, usefulness, pace, fleetness, smell, existence, appeal, birth, awkwardness
Related-Agent	ghost
Non-Attribute	gift, rhythm, physiology, battle, case, example, dance, manner, description

Cancer

Class	Best Attributes
Activity	growth, development, removal, treatment, recurrence, diagnosis, pain, spreading, metastasis, detection, eradication, elimination, production, discovery, remission, advance, excision, prevention, evolution, disappearance, anxiety
Part & Related-Object	location, site, lump, nature, root, cells, margin, formation, margins, roots, world, region
Quality	extent, size, seriousness, progression, severity, aggressiveness, cause, progress, symptoms, effects, risk, incidence, staging, biology, onset, characteristics, histology, ability, status, appearance, thickness, sensitivity, causes, prevalence, responsiveness, ravages, frequency, aetiology, circumstances, rarity, outcome, behavior, genetics
Related-Agent	club, patient
Non-Attribute	stage, spread, grade, origin, course, power, return, area, response, presence, type, particulars, occurrence, prognosis, pathogenesis, source, news, cure, pathology, properties, genesis, boundaries, drama, stages, chapter

Family

Class	Best Attributes
Activity	disintegration, protection, decline, destruction, breakup, abolition, participation, reunification, reconciliation, dissolution, composition, restoration
Part & Related-Object	head, institution, support, flower, core, fabric, culture, dimension, food, lineage, cornerstone, community
Quality	breakdown, importance, honor, structure, sociology, integrity, unity, sanctity, health, privacy, survival, definition, influence, honour, involvement, continuity, stability, size, preservation, upbringing, centrality, ancestry, solidarity, hallmark, status, functioning, primacy, autonomy
Related-Agent	father, baby, member, mother, members, patriarch, breadwinner, matriarch, man, foundation, founder, heir, daughter
Non-Attribute	rest, role, income, history, concept, welfare, pedigree, genealogy, presence, context, origin, bond, tradition, taxonomy, system, wealth, lifestyle, surname, crisis, ideology, rights, economics, safety