Integrating General-purpose and Corpus-based Verb Classification

Roberto Basili*
University of Roma Tor Vergata

Paola Velardi†
University of Ancona

Maria Teresa Pazienza*
University of Roma Tor Vergata

1. Introduction

A long-standing debate in the computational linguistic community is about the generality of lexical taxonomies. Many linguists (Nirenburg 1995; Hirst 1995) stress that taxonomies that are not language neutral, at least at the intermediate and high level, have little hope of success. On the other hand, lexicon builders who have experience of designing taxonomies for real applications claim that in sublanguages there exist very domain-dependent similarity relations. Given our experience and results, we are inclined to take the second position, but we are indeed sensitive to the theoretical motivations of the first.

The problem is that the similarity relations suggested by the thematic structures of words[1] in sentences are highly domain dependent, and it is difficult, though perhaps not impossible, to find common invariants across sublanguages when this model of word similarity is adopted. On the other hand, conceptual, or **compositional** models of similarity are much more difficult to understand and formalize on a systematic basis, because of the difficulty of defining a commonly agreed upon set of semantic primitives into which words may be decomposed.

It may be possible, however, and highly interesting, to integrate the results of a purely inductive method, such as the conceptual clustering system CIAULA (Basili, Pazienza, and Velardi 1993c, 1996a), and a hand-encoded, domain-general classification, such as, for example, WordNet. The purpose of one such experiment, which we describe in this paper, is to find some points of contact between psychologically motivated models, as WordNet, and data-driven models, as CIAULA.[2]

2. Detecting Verb Similarities with a Sublanguage

To analyze verb similarities, we used CIAULA, a conceptual clustering algorithm for word classification, which we applied to the task of verb categorization. We will not provide details of the specific algorithm used (they may be found in works referred to above), we will simply summarize the main features of the system.

CIAULA is an unsupervised learning algorithm for incremental concept formation, based on an augmented version of the well-known COBWEB (Fisher 1987).

The input observations (instances) for our concept-formation algorithm are **verb observations** in sentences, represented by their generalized thematic structures, acquired semiautomatically from corpora by ARIOSTO_LEX (Basili, Pazienza, and Velardi 1993a, 1993b, 1996b).

In CIAULA, the thematic roles of a verb v in a sentence are represented by a feature-vector:

$$v/(R_{i_t} : Cat_{i_t}) \qquad i_t \in I, j_t \in J \quad t = 1, 2, \ldots, n \tag{1}$$

In (1), R_{i_t} are the **conceptual relations**, or thematic roles (*agentive, instrument,* etc.) and Cat_{i_t} are the **conceptual types** of the words to which v is related semantically. Conceptual relations are assigned semiautomatically. Conceptual types are assigned either manually (in Italian, since no on-line resources are available) or automatically, using WordNet.

For example, the following sentence in an English remote sensing domain (hereafter RSD):

> ... *the satellite produced information with high accuracy*

originates the instance:

```
(2)    produce / (MANNER: PROPERTY,
                  AFFECTED: ABSTRACTION,
                  INSTRUMENT: INSTRUMENTALITY )
```

Semantic similarity is strongly suggested by the observation of verb configurations, in which words of the same conceptual type play the same roles.

Distinguishing features of CIAULA are:

1. *Treatment of ambiguity and polysemy.* In contrast to classical conceptual clustering algorithms, CIAULA accounts for multiple instances of the same concept, that is, of the same verb.

2. *Identification of most representative clusters.* The method identifies the **basic-level categories** (Rosch 1978) for an acquired hierarchy of verbs, i.e., those bringing the most predictive information about their members.

The results of repeated experiments showed that CIAULA is able, upon an appropriate setting of the model parameters, to detect similarity relations in the thematic structure of verbs, and to provide a probabilistic and semantic description of the acquired clusters. We observed that, by tuning the model parameters to obtain clusters of "very similar" instances, a percentage varying between 30% and 60% of verbs (depending upon the number of input observations) belong to singletons, that is, are not "similar enough" to any other verb. If we relax these constraints, we obtain larger clusters and fewer singletons, but verbs in a cluster are less semantically close to each other.

In Figure 1 we show two basic-level categories obtained with different values of the model parameters, for the RSD. For each cluster member, the local and global membership is shown. Members of Class 1,870 are verbs that take an *abstraction* (ABS) as a direct object. Class 1,603 was generated in a different run, in which we imposed a tighter similarity among the cluster members. The verbs in this category take with

```
Class:1870  Card:  6
PROTOTYPE (i.e., Predicted Thematic Roles):
          --  (AFFECTED)  --> [ABS]
Verbs (and their local and global degree of membership):
    generate      (0.500 - 0.250)
    customize     (0.500 - 1.000)
    provide       (0.500 - 0.167)
    use           (1.000 - 0.071)
    show          (0.500 - 0.333)
    vary          (0.500 - 0.500)
Class: 1603  Card:  3
PROTOTYPE (i.e., Predicted Thematic Roles):
 --  (AGENTIVE)  -- > [ABS]
 --     (MANNER)  -- > [PR,CO]
    Verbs (and their local and global degree of membership):
    illustrate    (0.500 - 1.000)
    calculate     (0.600 - 0.333)
    deal          (1.000 - 0.333)
```

Figure 1
Two CIAULA clusters obtained with different model parameters.

highest probability an ABSTRACTION as *agentive*, and a *manner* modifier that may be a PROPERTY (PR) or a COGNITIVE_PROCESS (CO). Some examples are: *the data/$_{ABS}$ illustrate the problem/$_{(ABS,CO)}$ with accuracy/$_{PR}$ ...*, *the algorithm/$_{ABS}$ efficiently/$_{PR}$ calculates* ....

The similarity relations detected by CIAULA cannot be used *tout court* as a taxonomy in a NLP system. However, they can be used to *tune* a general-purpose taxonomy to a specific domain, by reducing sense ambiguity and identifying new, domain-specific senses. For example, the predicted thematic structure for the cluster 1,603 shows that the verb *to deal* has a more specific use than in general language. In the RSD, *algorithms/$_{ABS}$ deal with parameters/$_{PR}$ ....*

Cross-linguistic experiments (Basili, Pazienza, and Velardi 1996b) showed that similarity relations are different in different domains, which raises the issue of detecting **language invariants**, that is, a language-neutral ontology (at least at the highest levels). Our contribution to this long-standing issue will be empirical, rather than methodological.

3. Analysing the Relations between Corpus-induced and Human-deduced Categories

In this section, we propose a method to analyze the relations between a domain-general ontology, such as WordNet, derived by linguists seeking language-neutral principles, and our example-driven clusters, derived by CIAULA. The purpose of this analysis is not to validate CIAULA with WordNet, nor to augment WordNet with CIAULA. Rather, our purpose is to identify commonalities and discrepancies, and to investigate the possibility of profitably integrating the two approaches.

One of the motivations for using WordNet is that, in WordNet, verb meaning is represented in terms of semantic *relations*, rather than semantic *primitives* (Miller and Fellbaum 1991). Hence, in principle, WordNet and CIAULA adopt the same relational

Let C be a cluster automatically derived by CIAULA (Basili et al. 1996a). A cluster is simply a set of verbs modeled by means of a graded membership function, i.e., the local membership of verbs, $\mu(v, C)$ (Basili, Pazienza, and Velardi 1993c). Let $S(v)$ be the set of senses of the verb v.

For each sense s in $S(v)$, the set of WordNet hyperonims of s is defined. Let $syns(v)$ denote this set, i.e.,

$$sysn(v) = \{syns \mid \exists s \in S(v), s \; is\_a \; syns\}$$

The is_a relation denote the transitive closure of WordNet IS_A: $syns(v)$ is the set of possible (ambiguous) WordNet hyperonims of the verb v, through its senses (i.e., $S(v)$).

Let $syns(C)$ denote the set of all hyperonims of at least one verb v in C, i.e.,

$$syns(C) = \cup_{v \in C} \; sysn(v)$$

Let $V(syns, C)$ be the set of verbs of a given cluster C that are hyponyms of $syns$. Formally,

$$V(sysn, C) = \{v \in C, syns \; \varepsilon \; syns(v)\}$$

The preference score g is a real-valued function defined by:

$$g(syns) = \sum_{v \in V(syns, C)} \frac{1}{|S(v)|}$$

where $| \; |$ denotes cardinality.
The best Wordnet label $syns$ for the cluster C is the one that maximizes g.

Figure 2
Labeling algorithm of CIAULA clusters.

approach to describe verbs and detect similarities. To investigate the commonalities between CIAULA and WordNet we decided to automatically select the best WordNet concept as a label to assign to each acquired CIAULA cluster.

Let v_i be the members of a CIAULA basic-level cluster C, $S(v_i)$ the synsets for each v_i and $h(S(v_i))$, or h_i, the set of supertypes (hyperonims) of $S(v_i)$. If $gr(h_i)$ is the number of incoming IS_A arcs for a supertype, that is the number of synsets of v_i that point to h_i, an intuitive algorithm would be to select as the best supertype for a cluster the one that maximizes $gr(h_i)$ values.

Things, however, are more complex. First, we must apply some normalization in order to reduce the noise caused by the more ambiguous verbs. Second, we must balance the effect of verbs that have more than one synset pointing to the same h_i. In fact, a supertype could gain evidence only because several senses of the same verb point to it. Finally, the algorithm must avoid the selection of excessively general categories, like *create, make*. Figure 2 describes the algorithm more formally.

During a first experiment, we ran the tagging algorithm using unrestricted sets of verbs first clustered by CIAULA. Because of the relatively sparse examples, the over generality of WordNet and the over specificity of CIAULA produced limited interactions. In some cases, CIAULA clusters received a "pertinent" WordNet sense label, but in some cases they did not. A "good" example in a legal domain in Italian
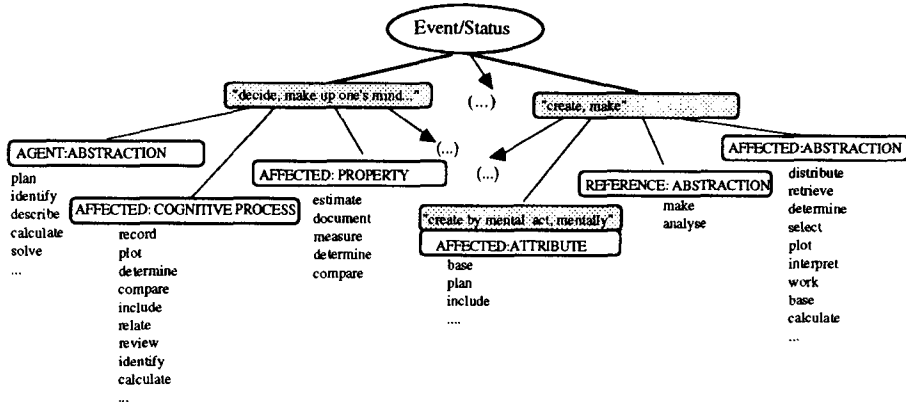
Figure 3
A portion of the acquired verb taxonomy.

(hereafter LD) is the class: *evaluate, regulate, assign, determine, examine, resolve, maintain* that received the label: *judge, form an opinion of, pass judgement on*. A "bad" choice in the same domain, is the class: *indicate, establish, foresee, determine* that received the overly general label: *create, make*. On average, we were satisfied with one-half of the tags assigned to clusters.

This evaluation was performed by inspection, hence it is purely empirical. Given a set of examples of verb uses, we can more or less easily tell whether a conceptual definition in terms of thematic structures (as provided by CIAULA) is appropriate or not. But it is much more difficult to say whether, for example, the CIAULA class *measure, propose, derive, evaluate, discover, classify, describe, calculate* is appropriately described by the WordNet label *communicate, transmit thoughts, transmit feelings*.

In WordNet, there are limited definitions of the conceptual labels used, or hyperonims. In general, lists of words or phrases are used in place of a single label, so that the reader may have an idea of what is really meant. But the higher the node, the deeper the "meaning" behind a hyperonim, the harder is the human task of evaluating the appropriateness of a classification.

Looking more in detail, the problem with misclassifications is twofold. In some cases, the problem is the overambiguity and the very fine-grained concept labels adopted in WordNet. Especially with large CIAULA clusters, the number of synsets becomes too large, and the algorithm does not gain enough evidence of any significantly promising pattern in the hierarchy. The second problem is the overspecificity of CIAULA. For example, verbs in the second cluster of the previous example (i.e., the "bad" choice), have been clustered because they occurred in patterns like: *the law/$_{DOCUMENT}$ indicates (establish, determine. . .) the deadline/$_{TEMPORAL\_ENTITY}$ for the presentation. . . .* It is unlikely that the linguists who developed WordNet had in mind such a narrow use of these verbs when classifying them.

WordNet labels for CIAULA classes are somewhat overly general. Different CIAULA clusters received the same WordNet label, and this was used as a hint to further structure the induced classification. An example is shown in Figure 3. The resulting taxonomy is built under a default node.

In a second set of experiments, we ran CIAULA on a more homogeneous set of verbs. Rather than inducing verb categories from scratch, we augmented the semantic bias of CIAULA by preclassifying all the verbs in the RSD using the 15 WordNet semantic domains for verbs, which are: *bodily care, change, cognition, communication,*

competition, consumption, contact, creation, emotion, motion, perception, possession, social interaction, stative, and *weather.* We then fed CIAULA with groups of verbs belonging to each of these categories. Of course, many of these verbs are ambiguous, but we used a probabilistic method (Basili et al. 1995) to select the observations of each verb that are genuine examples of a semantic domain.

This experiment produced rather appropriate classifications. Table 1 shows the labels assigned to some basic-level clusters generated by CIAULA for the RSD verbs belonging to the semantic category *cognition.* Figure 4 shows an excerpt of the CIAULA clusters of Table 1, with the prototypical description of each cluster.

In Table 1 the fourth column (Overlap Score) is the ratio between verbs in a cluster (column 1) that belong to the WordNet synset of column 3, and the cardinality of the cluster. In fact the best synset for a cluster does not necessarily cover all the cluster members. In the full experiment, 67% of the clusters have a score ≥ 0.5, indicating a good overlap between CIAULA and WordNet. Worst clusters, as far as the overlap score is concerned, are those in which there are very high-level and ambiguous verbs, like *make.* These verbs usually produce noise, because of WordNet ambiguity and of the spurious (for the category) input examples fed to CIAULA.

There are instead clusters with a low overlap score that seem very appropriate if one looks at the usage patterns in the corpus. For examples, the verbs of cluster 2,725 in Table 1 are highly characterized (i.e., have high local membership values) by the fact that they take as object some physical PROPERTY (PR) of a NATURAL_OBJECT. If we consider the prototypical descriptions of clusters globally, we observe recurrent patterns of use of the clustered verbs. Verbs of *cognition* in the RSD are strongly characterized by a MENTAL OBJECT (MO), or COGNITIVE_PROCESS (C0), or ABSTRACTION (ABS), in the position of direct object (AFFECTED). Frequently, the object of a *cognition* verb is a physical PROPERTY or a NATURAL_OBJECT, and the analysis is performed with some INSTRUMENTALITY (INS) (... *cloud parameters are derived from satellite...*).

In order to analyze the correspondence/divergence between human-coded verb classes and data-driven clusters we can compare the argument structure proposed for the synsets in WordNet and the intentional description of the classes, i.e., the prototypical semantic patterns of CIAULA clusters (Figure 4).

The WordNet argument structure for verbs, however, simply provides a qualitative description of the possible phrasal patterns in which verbs in a given synset can be used. For example the sense *record, enter, put down, make a record of* of the verb *to record* (line 1 in Table 1), is described by

(3)   {Somebody, Something} records {something, somebody}
        Somebody *records* that CLAUSE

As shown, the information available is mainly syntactic, with the exception of the ANIMATE/INANIMATE distinction for the arguments.

The classification of CIAULA assigns the verb *to record* to four classes: (1) *record, enter, put down, make a record of,* (2) *decide, make up one's mind, decide upon, determine,* (3) *create, make,* and (4) *investigate, look into,* as shown in Table 1. The different classes are characterized by the following semantic patterns, as shown in Figure 4:

(i)   class 4170 *to record* /(AFFECTED: PROPERTY)

(ii)   class 3637 *to record* /(AGENTIVE: COGNITIVE_PROCESS)

(iii)   class 3518 *to record* /(LOCATION: PLACE)

Table 1
Excerpt of CIAULA clusters for *cognition* verbs in the RSD.

Class #	Ciaula Clusters (cognition verbs)	WordNet Labels (synsets)	Overlap Scores
137	represent, lie	symbolize, stand for, express indirectly, represent	1.00
397	base, study, estimate, document, calculate, explore	analyze, analyse, study, examine	0.33
429	review, compare, include	judge, form an opinion of, pass judgment on	0.66
562	estimate, increase, soil, process, perform, observe, approach	change, alter	0.42
684	determine, compute	decide, make up one's mind, decide upon, determine	1.00
914	divide, transform, make	create, make	0.66
1,196	include, base, deal, involve, mind, relate, measure, review, compare	study, think about, contemplate	0.18
1,224	derive, describe, retrieve, document, review, compute, measure	get, acquire, enter upon, come upon, luck into	0.28
1,374	calculate, provide	think, cogitate, cerebrate	1.00
1,587	calculate, relate, focus	think, cogitate, cerebrate	1.00
1,941	scan, compare, propose, estimate, compute, analyse, study, experiment, base, evaluate	analyze, analyse, study, examine	0.30
2,049	assess, plan, review	judge, form an opinion of, pass judgment on	0.66
2,055	determine, provide, plan, retrieve, view, show	create, make	0.33
2,102	provide, make	create, make	1.00
2,147	include, propose, derive	include	0.33
2,383	account, stand, situate, estimate, study, make	create, make	0.33
2,491	select, locate, analyse	find, regain	0.66
2,725	infer, derive, measure, select, estimate, locate, compare, calculate	reason, reason out, conclude, arrive at	0.50
2,797	research, base, record	investigate, look into	0.66
3,518	collect, develop, map, list, record, make	create, make	0.50
3,637	determine, record, compare, measure, decide	make up one's mind, decide upon, determine	0.75
3,758	scan, survey	look at, take a look at, examine, examine by sight	1.00
4,080	research, locate	investigate, look into	1.00
4,170	evaluate, plot, record, base	record, enter, put down, make a record of	0.50

(iv)   class 2797 *to record* /(AFFECTED: ARTIFACT)

Some differences between the pattern in (3) and any of the feature vectors (*i–iv*) are:

- Most of the syntactic relations expressed in (3) are accounted for in the semantic patterns that CIAULA detects as prototypical for the verb *to*

```
CLASS: 2383    Card: 7                    CLASS: 2055    Card: 7
   stand       (0.500 - 1.000)              provide     (0.500 - 0.111)
   situate     (0.500 - 1.000)              plan        (0.500 - 0.100)
   estimate    (0.500 - 0.037)              retrieve    (0.500 - 0.167)
   study       (0.500 - 0.036)              view        (0.500 - 0.167)
   account     (1.000 - 1.000)              determine   (1.000 - 0.105)
   make        (0.500 - 0.111)              show        (0.500 - 0.500)
   PROT:  --   (REFERENCE)  --> [DOC]       PROT:  --   (AFFECTED) -- > [ABS]

CLASS: 2725    Card: 15                   CLASS: 2102    Card: 3
select         (0.800 - 0.077)              make        (0.500 - 0.111)
estimate       (0.800 - 0.037)              provide     (1.000 - 0.222)
locate         (0.200 - 0.143)              PROT:  --   (MANNER) -- > [CO]
measure        (0.400 - 0.077)
infer          (1.000 - 0.556)            CLASS: 914     Card: 5
compare        (0.200 - 0.048)              transform   (0.333 - 1.000)
derive         (0.600 - 0.103)              divide      (1.000 - 1.000)
calculate      (0.200 - 0.043)              make        (0.333 - 0.111)
PROT:  --      (AFFECTED)  --> [PR]         PROT:  --   (LOCATION)  --> [LOC]

CLASS: 2797    Card: 5                    CLASS: 3637    Card: 5
  base         (0.333 - 0.019)              record      (0.500 - 0.043)
  record       (0.333 - 0.043)              determine   (1.000 - 0.105)
  research     (1.000 - 0.125)              compare     (0.500 - 0.048)
  PROT:  --    (AFFECTED) -- > [ART]        measure     (0.500 - 0.038)
                                            PROT:  --   (AGENT)  --> [CO]
CLASS: 3518    Card: 7
  develop      (0.500 - 0.250)            CLASS: 4170    Card: 5
  map          (0.500 - 0.500)              plot        (0.500 - 0.091)
  list         (0.500 - 1.000)              record      (0.500 - 0.043)
  collect      (1.000 - 0.667)              base        (0.500 - 0.019)
  record       (0.500 - 0.043)              evaluate    (1.000 - 0.286)
  make         (0.500 - 0.111)              PROT:  --   (AFFECTED) -- > [PR]
  PROT:  --    (LOCATION)  -- > [LOC]
```

Figure 4
An excerpt of clusters of Table 1, with extensional and intensional descriptions.

record. Within the general properties of taking an (INANIMATE or ANIMATE) entity as subject and object, however, CIAULA specifies the semantics of the object and subject typical of the domain. For example (see cluster 4,170) in this domain, the activity of *recording, evaluating, plotting* a PROPERTY (e.g., *sea surface temperature, wind speed,...*) is significant;

- In some cases the (weak) semantic expectations on the argument structure in WordNet are violated. For example, in pattern (*ii*), a COGNITIVE_PROCESS rather than *somebody* is the agentive of *to record* (e.g., *the algorithm/$_{CO}$ records the changes...*) and of the other members of the class labeled *decide, make up one's mind, decide upon, determine* (cluster 3,637).

- Some relations are not predicted by WordNet, as for example pattern (*iii*). Locative relations are treated in WordNet as lexical adjuncts of the verb *to record*. However, they seem very relevant in the sublanguage (as for example in sentences like: ...*pollutants are recorded and analyzed in surface waters, temperature is recorded (collected) in the bay area*...). It seems that (some) lexical adjuncts may play an important role in the definition of domain-specific senses of verbs.

It appears that much information relevant for the lexical encoding of verbs is domain specific and is completely missing in a general-purpose classification like Word-Net. Therefore, semantically driven NL interpreters may profitably be augmented with the information obtained by merging these different sources.

A further interesting issue is related to identical tags assigned to different clusters. Verbs in these classes should express similar acts or events. An analysis of the prototypical patterns that CIAULA assigns to these classes suggests that despite the shared WordNet tags, verbs in these classes are very different.

For example, the classes 2,383, 2,055, 3,518, 2,102, 914 are all labeled *create, make*, a very general synset in WordNet. Their patterns are very different, and show almost no overlap. The main motivation for this divergence between WordNet tagging and the meanings of CIAULA clusters is twofold. On the one hand, CIAULA clusters are very fine grained, as they are built from single observations of verb uses. On the other hand, WordNet is often missing most of these precise (technical) uses of verbs. As a result, the labeling algorithm of Figure 2 is forced to generalize over many levels, with a consequent loss of information. The argument structure of the reached synset is thus too generic. It is worth noticing that even in these cases, we still achieve useful information, since the WordNet argument structure can be further specified by domain-specific semantic constraints. The class 2,055, for example, may be described by an extended argument structure like

(4)   {Somebody, Something} *makes* {X},
        where X is an ABSTRACTION

4. Conclusions

The results illustrated in this paper are very interesting, though not conclusive. Verbs are vessels for human creativity in language communication, and so much is left to further studies. We discovered thematic features that are apparently more "basic" than others, with respect to a given semantic domain (*cognition*) and a given sublanguage (RSD). We could specify features that were described at a very general level in Word-Net, and detect semantic restrictions specific to the sublanguage, not accounted for in WordNet. These results suggest that, with appropriate customization, it is still possible to exploit the information in general-purpose on-line thesauri that would be otherwise almost unusable in real NLP applications. As proposed in this paper, an appropriate process of lexical tuning can significantly reduce the overgenerality (excessive ambiguity) and underspecificity (weak constraints on verb argument structures) that is typical of general-purpose resources.

References

Basili, R., Della Rocca M., Pazienza, P. Velardi. 1995. Contexts and categories: Tuning a general purpose verb classification to sublanguages. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, Tzigov Chark, Bulgaria, September.

Basili, R., M. T. Pazienza, and P. Velardi. 1993a. What can be acquired from raw texts? *Journal of Machine Translation*, 8.

Basili, R., M. T. Pazienza, and P. Velardi. 1993b. Acquisition of selectional patterns in sublanguages. *Journal of Machine Translation*, 8:147–173.

Basili, R., M. T. Pazienza, and P. Velardi. 1993c. Hierarchical clustering of verbs. ACL-SIGLEX Workshop on Lexical Acquisition, Columbus, OH, June.

Basili, R., M. T. Pazienza, and P. Velardi. 1996a. A context driven conceptual clustering method for verb classification. In B. Boguraev, and J. Pustejovsky, editors, MIT Press.

Basili, R., M. T. Pazienza, and P. Velardi. 1996b. An empirical symbolic approach to natural language processing. *Artificial Intelligence Journal*. To appear.

Fisher, D. 1987. Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 2.

Hirst, G. 1995. Working notes. In *AAAI 1995 Fall Symposium on Representation and Acquisition of Lexical Knowledge: Polisemy, Ambiguity and Generativity*, Stanford, CA, March.

Miller, G., and C. Fellbaum. 1991. Semantic networks of English. *Cognition*, 41:197–229.

Nirenburg, S. 1995. Apologiae Ontologiae, Working Notes. In *AAAI 1995 Fall Symposium on Representation and Acquisition of Lexical Knowledge: Polisemy, Ambiguity and Generativity*. Stanford, CA, March.

Rosch, E. 1978. Principle of categorization. In *Cognition and Categorization*, Erlbaum.