

HIERARCHICAL CLUSTERING OF VERBS

Roberto Basili (*)
Maria Teresa Pazienza (*)
Paola Velardi (**)

(*) Universita' di Roma Tor Vergata, Italy
(**) Universita' di Ancona, Italy

Abstract.

In this paper we present an unsupervised learning algorithm for incremental concept formation, based on an augmented version of COBWEB. The algorithm is applied to the task of acquiring a verb taxonomy through the systematic observation of verb usages in corpora.

Using a Machine Learning methodology for a Natural language problem required adjustments on both sides. In fact, concept formation algorithms assume the input information as being stable, unambiguous and complete. At the opposite, linguistic data are ambiguous, incomplete, and possibly erroneous.

A NL processor is used to extract semi-automatically from corpora the thematic roles of verbs and derive a feature-vector representation of verb instances. In order to account for multiple instances of the same verb, the measure of category utility, defined in COBWEB, has been augmented with the notion of memory inertia. Memory inertia models the influence that previously classified instances of a given verb have on the classification of subsequent instances of the same verb. Finally, a method is defined to identify the *basic-level* classes of an acquired hierarchy, i.e. those bringing the most predictive information about their members.

1. Introduction

The design of word-sense taxonomies is acknowledged as one of the most difficult (and frustrating) tasks in NLP systems.

The decision to assign a word to a category is far from being straightforward

(Nirenburg and Raskin (1987)) and often the lexicon builders do not use consistent classification principia.

Automatic approaches to the acquisition of word taxonomies have generally made use of machine readable dictionaries (MRD), for the typical definitory nature of MRD texts. For example, in Byrd et al., (1987) and other similar studies the category of a word is acquired from the first few words of a dictionary definition. Besides the well known problems of inconsistency and circularity of definitions, an inherent difficulty with this approach is that *verbs* can hardly be defined in terms of *genus* and *differentiae*. Verb semantics resides in the nature of the event they describe, that is better expressed by the roles played by its arguments in a sentence. Psycholinguistic studies on verb semantics outline the relevance of thematic roles, especially in categorisation activities Keil, (1989), Jackendoff (1983) and indicate the argument structure of verbs as playing a central role in language acquisition Pinker (1989). In NLP, representing verb semantics with their thematic roles is a consolidated practice, even though theoretical researches (Pustejovski (1991)) propose more rich and formal representation frameworks.

More recent papers Hindle (1990), Pereira and Tishby (1992) proposed to cluster nouns on the basis of a metric derived from the distribution of subject, verb and object in the texts. Both papers use as a source of information large corpora, but differ in the type of statistical approach used to determine word similarity. These studies, though valuable, leave several open problems:

- 1) A metric of conceptual closeness based on mere syntactic similarity is questionable, particularly if applied to verbs. In fact, the argument structure of verbs is variegated and poorly overlapping. Furthermore, subject and object relations do not fully characterize many verbs.
- 2) Many events accumulate statistical evidence only in very large corpora, even though in Pereira and Tishby (1992) the adopted notion of distributional similarity in part avoids this problem.
- 3) The description of a word is an "agglomerate" of its occurrences in the corpus, and it is not possible to discriminate different senses.
- 4) None of the aforementioned studies provide a method to describe and evaluate the derived categories.

As a result, the acquired classifications seem of little use for a large-scale NLP system, and even for a linguist that is in charge of deriving the taxonomy.

Our research is an attempt to overcome in part the aforementioned limitations. We present a corpus-driven unsupervised learning algorithm based on a modified version of COBWEB Fisher (1987), Gennari et al. (1989). The algorithm learns verb classifications through the systematic observation of verb usages in sentences. The algorithm has been tested on two domains with very different linguistic styles, a commercial and a legal corpus of about 500,000 words each.

In section 2 we highlight the advantages that concept formation algorithms, like COBWEB, have over "agglomerate" statistical approaches. However, using a Machine Learning methodology for a Natural Language Processing problem required adjustments on both sides. Raw texts representing instances of verb usages have been processed to fit the feature-vector like representation needed for concept formation algorithms. The NL processor used for this task is briefly summarized in section 2.1. Similarly, it was necessary to adapt COBWEB to the linguistic nature of the classification activity, since, for

example, the algorithm does not discriminate different instances of the same entity, i.e. polysemic verbs, nor identical instances of different entities, i.e. verbs with the same pattern of use. These modifications are discussed in sections 2.1 through 2.3. Finally, in section 3 we present a method to identify the basic-level categories of a classification, i.e. those that are repository of most of the lexical information about their members.

Class descriptions and basic-level categories, as derived by our clustering algorithm, are in our view greatly helpful at addressing the intuition of a linguist towards the relevant taxonomic relations in a given language domain.

2. CIAULA¹: An algorithm to acquire word clusters

Incremental example-based learning algorithms, like COBWEB Fisher (1987), seem more adequate than other Machine Learning and Statistical methods to the task of acquiring word taxonomies from corpora. COBWEB has several desirable features:

- a) *Incrementality*, since whenever new data are available, the system updates its classification;
- b) A *formal description* of the acquired clusters;
- c) The notion of *category utility*, used to select among competing classifications.

b) and c) are particularly relevant to our linguistic problem, as remarked in the Introduction.

On the other side, applying COBWEB to verb classification is not straightforward. First, there is a knowledge representation problem, that is common to most Machine Learning algorithms: Input instances must be pre-coded (manually) using a feature-

¹ Ciaula stands for Concept formation Algorithm Used for Language Acquisition, and has been inspired by the tale "*Ciaula scopre la luna*" by Luigi Pirandello (1922).

vector like representation. This limited the use of such algorithms in many real world problems. In the specific case we are analyzing, a manual codification of verb instances is not realistic on a large scale.

Second, the algorithm does not distinguish multiple usages of the same verb, nor different verbs that are found with the same pattern of use, since different instances with the same feature vector are taken as identical. The motivation is that concept formation algorithms as COBWEB assume the input information as being stable, unambiguous, and complete. At the opposite, our data do not exhibit a stable behaviour, they are ambiguous, incomplete, and possibly misleading, since errors in codification of verb instances may well be possible.

In the following sections we will discuss the methods by which we attempted to overcome these obstacles.

2.1 Representing verb instances

This section describes the formal representation of verb instances and verb clusters in CIAULA.

Verb usages input to the clustering algorithm are represented by their thematic roles, acquired semi-automatically from corpora by a process that has been described in Basili, (1992a), (1992b), (in press). In short, sentences including verbs are processed as follows:

First, a (general-purpose) morphologic and a partial syntactic analyzer Basili, (1992b) extracts from the sentences in the corpus all the *elementary syntactic relations* (esl) in which a word participates. Syntactic relations are word pairs and triples augmented with a syntactic information, e.g. for the verb *to carry*: N_V(company,carry) V_N(carry,food) V_N(carry,goods) V_prep_N(carry,with,truck), etc.

Each syntactic relation is stored with its frequency of occurrence in the corpus. Ambiguous relations are weighted by a $1/k$ factor, where k is the number of competing esl in a sentence.

Second, the verb arguments are tagged by hand using 10-12 "naive" conceptual types (semantic tags), such as: ACT, PLACE, HUMAN_ENTITY, GOOD, etc. Conceptual types are not the same for every domain, even though the commercial and legal domains have many common types.

Syntactic relations between words are validated in terms of semantic relations between word classes using a set of semi-automatically acquired selectional rules Basili, (1992a). For example, V_prep_N(carry,with,truck) is accepted as an instance of the high-level selectional rule [ACT]-->(INSTRUMENT)-

>[MACHINE]. The relation: [carry]->(INSTRUMENT)->[truck] is acquired as part of the argument structure of the verb *to carry*. In other published papers we demonstrated that the use of semantic tags greatly increase the statistical stability of the data, and add predictive power to the acquired information on word usages, at the price of a limited manual work (the semantic tagging).

For the purpose of this paper, the interesting aspect is that single instances of verb usages (local² meanings) are validated on the basis of a global analysis of the corpus. This considerably reduces (though does not eliminate) the presence of erroneous instances.

The detected thematic roles of a verb v in a sentence are represented by the feature-vector:

$$(1) v / (\mathcal{R}_{i_t} : \text{Cat}_{j_t}) \quad i_t \in I, j_t \in J \quad t=1,2,\dots,n$$

where \mathcal{R}_{i_t} are the thematic roles (AGENT, INSTRUMENT etc.)³ and Cat_{j_t} are the conceptual types of the words to which v is related semantically. For example, the

² i.e. meanings that are completely described within a single sentence of the corpus

³ The roles used are an extension of Sowa's *conceptual relations* [Sowa 1984]. Details on the set of conceptual relations used and a corpus-based method to select a domain-appropriate set, are provided in other papers.

following sentence in the commercial domain:

"... la ditta produce beni di consumo con macchinari elettromeccanici.."

"... the company produces goods with electromechanical machines.."

originates the instance:

produce/(AGENT:HUMAN_ENTITY,
OBJECT:GOODS,
INSTRUMENT:MACHINE)

Configurations in which words of the same conceptual type play the same roles are strong suggestion of semantic similarity between the related events. The categorisation process must capture this similarity among local meanings of verbs.

The representation of verb clusters follows the scheme adopted in COBWEB. Each target class is represented by the probability that its members (i.e. verbs) are seen with a set of typical roles. Given the set $\{\mathcal{R}_i\}_{i \in I}$ of thematic roles and the set $\{\text{Cat}_j\}_{j \in J}$ of conceptual types, a target class \mathcal{C} for our clustering system is given by the following

$$(2) \quad \mathcal{C} = \langle c_{\mathcal{C}}, [x]_{ij}, V_{\mathcal{C}}, S_{\mathcal{C}} \rangle$$

or equivalently by

$$(2)' \quad \langle c, [x]_{ij}, V, S \rangle$$

A class is represented in COBWEB by the matrix $[x]_{ij}$, showing the distribution of probability among relations (\mathcal{R}_i) and conceptual types (Cat_j). The additional parameters $V_{\mathcal{C}}$ and $c_{\mathcal{C}}$ are introduced to account for multiple instances of the same verb in a class. $c_{\mathcal{C}}$ is the cardinality (i.e. the number of different instance members of \mathcal{C}), and $V_{\mathcal{C}}$ is the set of pairs $\langle v, v\# \rangle$ such that it exists at least one instance

$$v / (\mathcal{R}_i; \text{Cat}_j)$$

classified in \mathcal{C} , and $v\#$ is the number of such instances.

Finally, $S_{\mathcal{C}}$ is the set of \mathcal{C} subtypes. The

definitions of the *empty class* (3.1) and of the *top node* of the taxonomy (3.2) follows from (2)

$$(3.1) \quad \langle 0, [x]_{ij}, \{\emptyset\}, \{\emptyset\} \rangle$$

with $x_{ij}=0$ for each i, j

$$(3.2) \quad \langle N_{\text{tot}}, [x]_{ij}, V, S \rangle$$

where N_{tot} is the number of available instances in the corpus, V is the set of verbs with their absolute occurrences.

An excerpt of a class acquired from the legal domain is showed in Fig. 1. The semantic types used in this domain are listed in the figure.

Special type of classes are those in which only a verb has been classified, that we will call singleton classes. A *singleton class* is a class $\mathcal{C} = \langle c, [x]_{ij}, V, S \rangle$ for which $\text{card}(V)=1$. It will be denoted by $\{v\}$ where v is the only member of (whatever its occurrences) \mathcal{C} . For a singleton class it is clearly true that $S = \{\emptyset\}$. Note that a singleton class is different from an instance because any number of instances of the verb v can be classified in $\{v\}$.

2.2 Measuring the utility of a classification

As remarked in the introduction, a useful property of concept formation algorithms, with respect to agglomerate statistical approaches, is the use of formal methods that guide the classification choices.

Quantitative approaches to model human choices in categorisation have been adopted in psychological models of conceptual development. In her seminal work, Rosch (1976) introduced a metrics of preference, the category cue validity, expressed by the sum of expectations of observing some feature in the class members. This value is maximum for the so-called *basic level* categories. A later development, used in COBWEB, introduces the notion of category utility, derived from the application of the Bayes law to the expression of the predictive power of a given classification. Given a classification

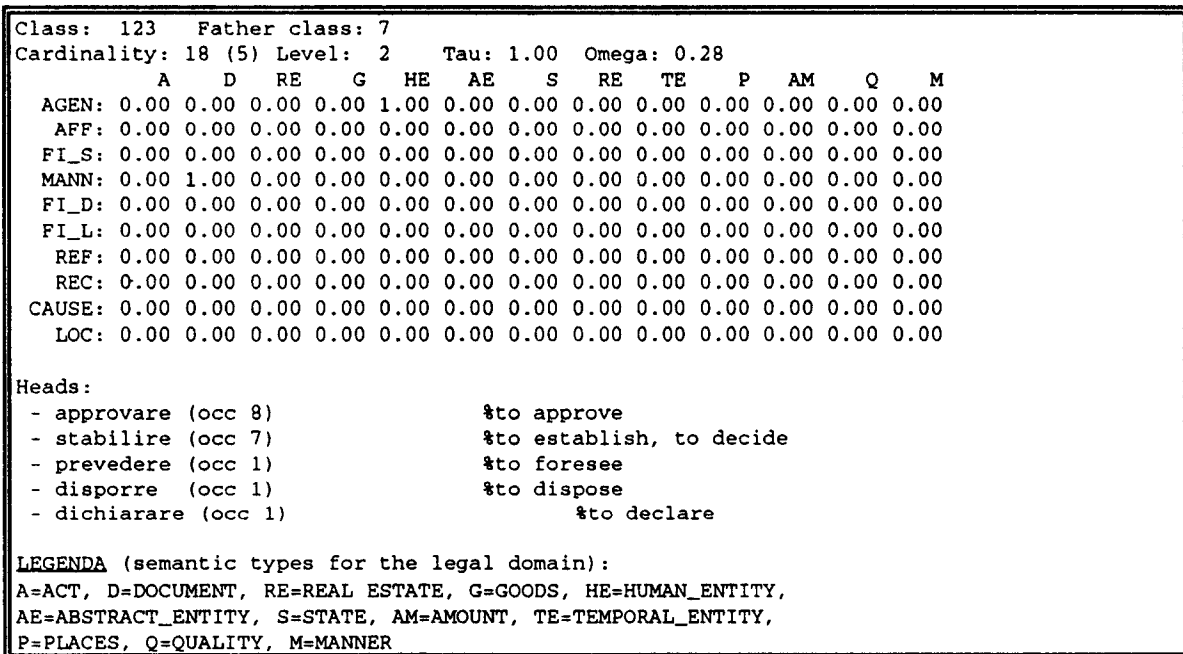
into K classes, the *category utility* is given by:

$$(4) \sum_{k=1}^K \text{prob}(C_k) \sum_{ij} \text{prob}(\text{attr}_i=\text{val}_j | C_k)^2$$

In COBWEB, a hill climbing algorithm is defined to maximize the category utility of a resulting classification. The following expression is used to discriminate among conflicting clusters:

$$(5) \frac{\sum_{k=1}^K \text{prob}(C_k) \sum_{ij} \text{prob}(\text{attr}_i=\text{val}_j | C_k)^2 - \sum_{ij} \text{prob}(\text{attr}_i=\text{val}_j)^2}{K}$$

The clusters that maximize the above quantity provide the system with the capability of deriving the best predictive taxonomy with respect to the set of i attributes and j values. This evaluation maximizes infra-class similarity and intra-class dissimilarity.



- Fig 1. Example of cluster produced by the system -

The notion of category utility adopted in COBWEB, however, does not fully cope with our linguistic problem. As remarked in the previous section, multiple instances of the same entity are not considered in COBWEB. In order to account for multiple instances of a verb, we introduced the notion of *mnemonic inertia*. The mnemonic inertia models an inertial trend attracting a new instance of an already classified verb in the class where it was previously classified.

Given the incoming instance

$$v / (R_i:Cat_j)$$

and a current classification in the set of classes \mathcal{C}_k , for each k the mnemonic inertia is modelled by:

$$(6) \quad \mu_k(v) = \#_v / c_k$$

where $\#_v$ is the number of instances of the verb v already classified in \mathcal{C}_k and c_k is the cardinality of \mathcal{C}_k .

(6) expresses a fuzzy membership of v to the class \mathcal{C}_k . The more instances of v are classified into \mathcal{C}_k , the more future observations of v will be attracted by \mathcal{C}_k . A suitable combination of the mnemonic

inertia and the category utility provides our system with generalization capabilities along with the "conservative" policy of leaving different verb instances separate. The desired effect within the data is that slightly different usages of a verb are classified in the same cluster, while remarkable differences result in different classifications.

The global measure of category utility, used by the CIAULA algorithm during classification, can now be defined. Let $v / (R_i:Cat_j)$ be the incoming instance, \mathcal{C}_k be the set of classes, and let $cu(v,k)$ be the category utility as defined in (5), the measure μ , given by

$$(7) \mu = vcu(v,k) + (1-v)\mu_k(v) \quad v \in [0,1]$$

expresses the global utility of the classification obtained by assigning the instance v to the class \mathcal{C}_k . (7) is a distance metrics among instances and classes.

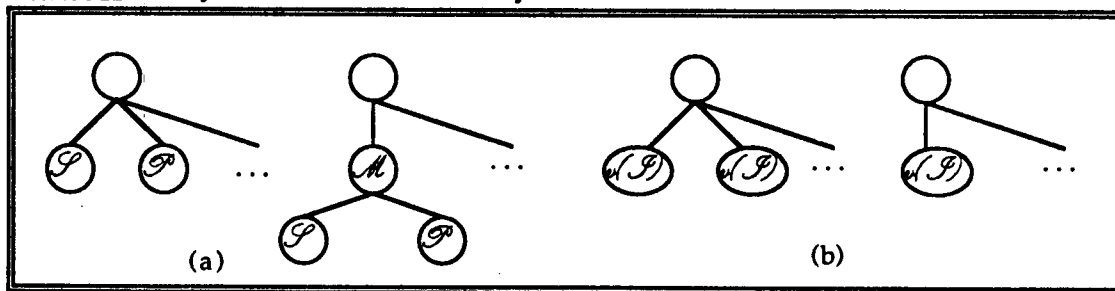
2.3 The incremental clustering algorithm.

The algorithm for the incremental clustering of verb instances follows the approach used in COBWEB. Given a new incoming instance I and a current valid classification $\{\mathcal{C}_k\}_{k \in K}$, the system evaluates the utility

of the new classification obtained by inserting I in each class. The maximum utility value corresponds to the best predictive configuration of classes. A further attempt is made to change the current configuration (introducing a new class, merging the two best candidate for the classification or splitting the best classes in the set of its son) to improve the predictivity. The main difference with respect to COBWEB, due to the linguistic nature of the problem at hand, concern the procedure to evaluate the utility of a temporary classification and the MERGE operator, as it applies to singleton classes. The description of the algorithm is given in Appendix 1. Auxiliary procedures are omitted for brevity.

According to (7), the procedure $G_UTILITY(x, I, \mathcal{C}, \mathcal{C}_k, v)$ evaluates the utility of the classification as a combination of the category utility and the inertial factor introduced in (6). Current values experimented for v are 0.90-0.75.

Figure 2 shows the difference between the standard MERGE operation, identical to that used in COBWEB, and the elementary MERGE between two singleton classes, as defined in CIAULA.



- Fig. 2: Merge (a) vs. Elementary Merge (b) -

3. Experimental Results.

The algorithm has been experimented on two corpora of about 500,000 words each, a legal and a commercial domain, that exhibit very different linguistic styles and verb usages. Only verbs for which at least

65 instances in each corpus have been considered, in order to further reduce parsing errors. Notice however that the use of semantic tags in corpus parsing reduces considerably the noise, with respect to other corpus-based approaches.

In the first experiment, CIAULA classifies 3325 examples of 371 verbs, from the legal corpus. In the second, it receives 1296 examples of 41 verbs from the commercial corpus. Upon a careful analysis of the clusters obtained from each domain, the resulting classifications were judged quite expressive, and semantically biased from the target linguistic domains, a part from some noise due to wrong semantic interpretation of elementary syntactic structures Basili et al., (1992a). However, the granularity of the description of the final taxonomy is too fine, to be usefully imported in the type hierarchy of a NLP system. Furthermore, the order of presentation of the different examples strongly influences the final result⁴. In order to derive reliable results we must find some invariant with respect to the presentation order. An additional requirement is to define some objective measure of the quality of the acquired classification, other than the personal judgement of the authors.

In this section we define a measure of the class *informative power*, able to capture the most relevant levels of the hierarchy. The idea is to extract from the hierarchy the *basic level classes*, or classes that are repository of the most relevant lexical information about their members. We define basic level classes of the classification those bringing most predictive and stable information with respect to the presentation order.

The notion of basic level classes has been introduced in Rosch (1978). She experimentally demonstrated that some conceptual categories are more meaningful than others as for the quantity of information they bring about their members. Membership to such classes implies a grater number of attributes to be inherited by instances of the domain. These classes appear at the intermediate levels of a taxonomy: for example within the vague notion of *animal*, classes such *dog* or *cat*

⁴ This is an inherent problem with concept formation algorithms

seem to concentrate the major part of information about their members, with respect for example to the class of *mammals* Lakoff (1987).

But what is a basic-level class for verbs? A formal definition for these more representative classes, able to guide the intuition of the linguist in the categorisation activity has been attempted, and will be discussed in the next section.

3.1. Basic level categories of verbs.

The information conveyed by the derived clusters, $\mathcal{C} = \langle c, [x]_{ij}, V, S \rangle$, is in the distributions of the matrices $[x]_{ij}$, and in the set V . Two examples may be helpful at distinguishing classes that are more selective, from other more vague clusters.

Let \mathcal{C}_1 be a singleton class, with $\mathcal{C}_1 = \langle 1, [x_1], V_1, \{\emptyset\} \rangle$. This clearly implies that $[x_1]$ is binary. This class is highly *typical*, as it is strongly characterized by its only instance, but it has no generalization power. Given, for example, a class $\mathcal{C}_2 = \langle 10, [x_2], V_2, S \rangle$ for which the cardinality of a V_2 is 10, and let $[x_2]$ be such that for each couple $\langle i, j \rangle$ for which $x_{2ij} \neq 0$, it follows $x_{2ij} = 1/10$. This class is scarcely typical but has a *strong generalization power*, as it clusters verbs that show no overlaps between the thematic roles they are represented by. We can say that typicality is signaled by high values of roles-types probabilities (i.e. $x_{ij} = \text{prob}(\mathcal{R}_i : \text{Cat}_j | \mathcal{C})$), while the generalization power ω of a class $\mathcal{C} = \langle c, [x]_{ij}, V, S \rangle$, is related to the following quantity:

$$(8) \quad \omega = \text{card}(V)/c$$

To quantify the typicality of a class $\mathcal{C} = \langle c, [x]_{ij}, V, S \rangle$, the following definitions are useful. Given a threshold $\alpha \in [0, 1]$, the *typicality* of \mathcal{C} is given by:

$$(9) \tau_{\mathcal{C}} = \sum_{\langle i,j \rangle \in T_{\mathcal{C}}} x_{ij} / \text{card}(T_{\mathcal{C}})$$

where $T_{\mathcal{C}}$ is the *typicality set* of \mathcal{C} , i.e. $\{\langle i,j \rangle \mid x_{ij} > \alpha\}$.

DEF (Basic-level verb category). Given two thresholds $\gamma, \delta \in [0, 1]$, $\mathcal{C} = \langle c, [x]_{ij}, V, S \rangle$ is a *basic-level category* for the related taxonomy iff:

$$(10.1) \omega < \gamma \quad (\text{generalization power})$$

$$(10.2) \tau_{\mathcal{C}} > \delta \quad (\text{typicality})$$

Like all the classes derived by the algorithm of section 2.3, each basic-level category $\mathcal{C} = \langle c, [x]_{ij}, V, S \rangle$ determines two fuzzy membership values of the verb v included in V . The *local membership* of v to \mathcal{C} , $\mu_1^{\mathcal{C}}(v)$, is defined by:

$$(11) \mu_1^{\mathcal{C}}(v) = \#_v / \max\{\# \mid \langle w, \#_w \rangle \in V\}$$

The *global membership* of v to \mathcal{C} , $\mu_2^{\mathcal{C}}(v)$, is:

$$(12) \mu_2^{\mathcal{C}}(v) = \#_v / n_v,$$

where n_v is the number of different instances of v in the learning set. (11) depends on the contribution of v to the distribution of probabilities $[x]_{ij}$, i.e. it measures the adherence of v to the prototype. (12) determines how typical is the classification of v in \mathcal{C} , with respect to all the observations of v in the corpus. Low values of the global membership are useful at identifying instances of v that are likely to be originated by parsing errors.

Given a classification \mathcal{F} of extended sets of linguistic instances, the definition (10) identifies all the basic-level classes. Repeated experiment over the two corpora demonstrated that these classes are substantially invariant with respect to the presentation order of the instances.

The values $\gamma=0.6$ and $\delta=0.75$ have been empirically selected as producing the most stable results in both corpora.

4 Discussion

The Appendix 2 shows all the basic level categories derived from a small learning set, named DPR633, that belongs to the legal corpus. CIAULA receives in input 293 examples of 30 verbs. The reason for showing DPR633 rather than an excerpt of the results derived from the full corpus is that there was no objective way to select among the over 300 basic level classes. In Appendix 2, the relatively low values of μ_1 and μ_2 are due to the exiguity of the example set, rather than to errors in parsing, as remarked in the previous section. Of course, the basic-level classes extracted from the larger corpora exhibit a more striking similarity among their members, indicated by highest values of global and local membership. An example of cluster extracted from the whole legal corpus was shown in Figure 1.

The example shown in Appendix 2 is however "good enough" to highlight some interesting property of our clustering method. Each cluster has a semantic description, and the degree of local and global membership of verbs give an objective measure of the similarity among cluster members. It is interesting to observe that the algorithm classifies in distinct clusters different verb usages. For example, the cluster 4 and the cluster 6 classify two different usages of the verb *indicare*, e.g. *indicare un'ammontare* (to indicate an amount) and *indicare un motivo* (to specify a motivation), where "ammontare" is a type of AMOUNT(AM) and "motivo" is a type of ABSTRACT_ENTITY (AE).

The two clusters 13 and 14 capture the physical and abstract use of *eseguire*, e.g. *eseguire un'opera* (to build a building(=REAL_ESTATE) vrs. *eseguire un pagamento* (to make a payment(=AMOUNT,ACT)).

The clusters 3 and 6 classify two uses of the verb *tenere*, i.e. *tenere un registro* (to keep a record(=DOCUMENT) vrs. *tenere un discorso* (to hold a speech(=ABSTRACT_ENTITY)). Many other (often domain-dependent) examples are reflected in the derived classification.

To sum up, we believe that CIAULA has several advantages over other clustering algorithms presented in literature.

- (1) The derived clusters have a semantic description, i.e. the predicted *thematic roles* of its members.
- (2) The clustering algorithm incrementally assigns instances to classes, evaluating its choices on the basis of a formal criterium, *the global utility*.
- (3) The defined measures of typicality and generalization power make it possible to select the basic-level classes of a hierarchy, i.e. those that are repository of most lexical information about their members. These classes demonstrated substantially stable with respect to the order of presentation of option, i.e. the predicted thematic roles of its members.
- (4) It is possible to discriminate different usages of verbs, since verb instances are considered individually.

The hierarchy, as obtained by CIAULA, is not usable *tout court* by a NLP system, however class descriptions and basic-level categories appear to be greatly useful at addressing the intuition of the linguist.

References.

- R. Basili, M.T. Pazienza, P. Velardi, (1992a) *Computational Lexicons: the neat examples and the odd exemplars*, Proc. of 3rd. Conf. on Applied NLP, 1992.
- R. Basili, M.T. Pazienza, P. Velardi, (1992b) "A shallow Syntax to extract word associations from corpora", in *Literary and Linguistic Computing*, vol. 2, 1992
- R. Basili, M.T. Pazienza, P. Velardi, (in press) *Semi-automatic extraction of linguistic information for syntactic disambiguation*, Applied Artificial Intelligence.
- R. Byrd, N. Calzolari, M. Chodorow, I. Klavans, M. Neff, O. Rizk, (1987), "Large lexicons for

- Natural Language Processing: Utilizing the Grammar Coding System of LDOCE*", In *Computational Linguistics*, December 1987
- Dubois, D., Prade, H. (1988), "Possibility Theory: an Approach to Computerized Processing of Uncertainty", Plenum Press, New York, 1988.
- Fisher, D., (1987), *Knowledge acquisition via incremental conceptual clustering*, Machine Learning, 2, 1987.
- J. Gennari, P. Langley, D. Fisher, (1989), *Model of incremental Concept Formation*, in *Artificial Intelligence* n.1-3, 1989.
- Jacobs, P., (1991), "Integrating language and meaning in structured inheritance networks", in "Principles of Semantic Networks", J. Sowa Ed., Morgan Kaufmann, 1991.
- R. Jackendoff, (1983), "Semantics and cognition", MIT Press, 1983.
- D. Hindle, (1990), *Noun classification from predicate argument structures*, in Proc. of ACL , 1990
- F. Keil, (1989), *Concepts, kinds and cognitive development*, The MIT press, 1989.
- G. Lakoff, (1987), *Woman, fire and dangerous things*, University of Chicago Press, 1987.
- S. Nirenburg, V. Raskin, (1987), *The subworld concept lexicon and the lexicon management system*, In *Computational Linguistics*, n. 13, December 1987
- F.Pereira, H. Tishby, (1992), "Distributional similarity, Phase Transition and Hierarchical Clustering"; in Proc. of AAAI Fall Symposium Series, Probabilistic Approaches to Natural Language, Cambridge, October, 1992.
- S.Pinker, (1989), *Learnability and Cognition - The Acquisition of Argument Structure*, MIT Press, 1989.
- Luigi Pirandello, (1922), *Novelle per un anno*, Editore R. Bemporad, Mondadori, 1922.
- Pustejovsky, J., (1991), "The Generative Lexicon", *Computational Linguistics*, vol. 17, n. 4, 1991.
- E. Rosch, (1978), *Principle of categorization*, in *Cognition and Categorization*, Erlbaum 1978.

Appendix 1: The Algorithm for Conceptual Clustering of Verb Semantic Instances

```

Input:  $T_{op}$  root node of the current taxonomy
          I, Unclassified verb semantic instance
          v(I), verb head of the instance I
Output: An exhaustive conceptual classification of the incoming instances.
Variables:  $\mathcal{C}, \mathcal{P}, \mathcal{S}, \mathcal{N}, \mathcal{M}$  classes of the taxonomy
              x, p, s, n, m, q measures of global utility of a classification

CIAULA( $T_{op}, I, v$ )
  IF  $T_{op}$  is a terminal
  THEN
    IF  $T_{op}$  is the singleton {v(I)}
    THEN
      INCORPORATE( $T_{op}, I$ )

    ELSE NEW_TERMINAL( $T_{op}, I$ )
      INCORPORATE( $T_{op}, I$ )

  ELSE
    INCORPORATE( $T_{op}, I$ )
    FOR EACH subtype  $\mathcal{C}$  of  $T_{op}$ 
      G_UTILITY $x, I, \mathcal{C}, T_{op}, v$ )

    Let: p the best score x for classifying I in the class  $\mathcal{P}$ 
          s the second best score x for classifying I in the class  $\mathcal{S}$ 
          n the score x for classifying I in a new node  $\mathcal{N}$ , subtype of  $T_{op}$ 
          m the score x of classifying I in node  $\mathcal{M}$ , merge between  $\mathcal{P}$  and  $\mathcal{S}$ 
          q the score x in classifying I in a classification obtained removing  $\mathcal{P}$  from the
            current level and picking up the set of its son to the previous  $\mathcal{P}$  level

    IF p is the highest score
    THEN
      CIAULA( $\mathcal{P}, I, v$ )
    ELSE IF n is the highest score
    THEN initialize  $\mathcal{N}$  with values shown by I
    ELSE IF m is the highest score
    THEN
      MERGE( $\mathcal{M}, \mathcal{P}, \mathcal{S}, T_{op}, I$ )
      CIAULA( $\mathcal{M}, I, v$ )
    ELSE IF q is the highest score
    THEN
      SPLIT( $\mathcal{P}, T_{op}$ )
      CIAULA( $T_{op}, I, v$ )

  END

```

Appendix 2: Basic level classes derived from the DPR633 Corpus

Class: 1 Card: 3 Omega=0.67 Tau: 1.00
PROTOTYPE (i.e., Predicted Roles):
-- (FIG_DEST) -- [A]
-- (RECIPIENT) -- [HE]
Verbs (local - global degree membership):
dichiarare (0.50 - 0.20)
applicare (1.00 - 0.07)

Class: 2 Card: 10 Omega=0.50 Tau: 1.00
PROTOTYPE (i.e., Predicted Roles):
-- (OBJ) -- [AM]
-- (RECIPIENT) -- [HE]
Verbs (local - global degree membership):
richiedere (0.33 - 0.25)
eseguire (0.33 - 0.06)
applicare (1.00 - 0.11)
versare (0.66 - 0.13)
pagare (1.00 - 0.30)

Class: 3 Card: 14 Omega=0.50 Tau: 1.00
PROTOTYPE (i.e., Predicted Roles):
-- (OBJ) -- [D]
-- (REFERENCE) -- [D]
Verbs (local - global degree membership):
tenere (0.25 - 0.11)
applicare (0.25 - 0.03)
allegare (0.50 - 0.33)
prevedere (1.00 - 0.44)
emettere (0.25 - 0.07)
indicare (0.50 - 0.03)
eseguire (0.75 - 0.20)

Class: 4 Card: 9 Omega=0.33 Tau: 1.00
PROTOTYPE (i.e., Predicted Roles):
-- (OBJ) -- [AM]
-- (REFERENCE) -- [D]
Verbs (local - global degree membership):
applicare (1.00 - 0.19)
indicare (0.60 - 0.05)
prevedere (0.20 - 0.11)

Class: 5 Card: 3 Omega=0.67 Tau: 0.89
PROTOTYPE (i.e., Predicted Roles):
-- (OBJ) -- [A, AM]
-- (REFERENCE) -- [D]
Verbs (local - global degree membership):
ammettere (1.00 - 0.66)
modificare (0.50 - 0.50)

Class: 6 Card: 8 Omega=0.62 Tau: 1.00
PROTOTYPE (i.e., Predicted Roles):
-- (OBJ) -- [AE]
-- (REFERENCE) -- [D]
Verbs (local - global degree membership):
operare (0.25 - 0.09)
effettuare (0.25 - 0.01)
richiedere (0.25 - 0.25)
tenere (0.25 - 0.11)
indicare (1.00 - 0.07)

Class: 7 Card: 3 Omega=0.67 Tau: 0.78
PROTOTYPE (i.e., Predicted Roles):
-- (OBJ) -- [RE]
-- (FIG_LOC) -- [HE]
-- (REFERENCE) -- [D]
Verbs (local - global degree membership):
comprendere (0.50 - 0.05)
indicare (1.00 - 0.03)

Class: 8 Card: 18 Omega=0.28 Tau: 1.00
PROTOTYPE (i.e., Predicted Roles):
-- (SUBJ) -- [HE]
-- (MANNER) -- [D]
Verbs (local - global degree membership):
prevedere (0.12 - 0.11)
disporre (0.12 - 0.33)
approvare (1.00 - 0.88)
stabilire (0.87 - 0.38)
dichiarare (0.12 - 0.20)

Class: 9 Card: 3 Omega=0.67 Tau: 0.78
PROTOTYPE (i.e., Predicted Roles):
-- (SUBJ) -- [HE]
-- (MANNER) -- [A]
-- (LOCATION) -- []
Verbs (local - global degree membership):
rendere (0.50 - 0.20)
operare (1.00 - 0.18)

Class: 10 Card: 3 Omega=0.67 Tau: 0.78
PROTOTYPE (i.e., Predicted Roles):
-- (OBJ) -- [AM]
-- (FIG_DEST) -- [A]
-- (REFERENCE) -- [D]
Verbs (local - global degree membership):
versare (0.50 - 0.06)
operare (1.00 - 0.18)

Class: 11 Card: 6 Omega=0.67 Tau: 1.00
PROTOTYPE (i.e., Predicted Roles):
-- (OBJ) -- [D]
-- (FIG_DEST) -- [A]
Verbs (local - global degree membership):
presentare (0.33 - 0.06)
tenere (0.33 - 0.11)
emettere (1.00 - 0.23)
indicare (0.33 - 0.01)

Class: 12 Card: 6 Omega=0.50 Tau: 0.78
PROTOTYPE (i.e., Predicted Roles):
-- (OBJ) -- [A, AM]
-- (MANNER) -- [A]
Verbs (local - global degree membership):
versare (0.25 - 0.06)
eseguire (0.25 - 0.06)
effettuare (1.00 - 0.07)

Class: 13 Card: 3 Omega=0.67 Tau: 1.00

PROTOTYPE (i.e., Predicted Roles):

-- (OBJ) -- [RE]

-- (MANNER) -- [A]

Verbs (local - global degree membership):

eseguire (0.50 - 0.06)

comprendere (1.00 - 0.10)

Class: 14 Card: 11 Omega=0.55 Tau: 1.00

PROTOTYPE (i.e., Predicted Roles):

-- (OBJ) -- [A, AM]

Verbs (local - global degree membership):

produrre (0.16 - 0.11)

considerare (0.16 - 0.16)

applicare (0.16 - 0.03)

eseguire (0.16 - 0.06)

effettuare (1.00 - 0.11)

indicare (0.16 - 0.01)