Integration of probabilistic and symbolic methods for semantic categorization

Basili R.(*), M.T. Pazienza (*), P. Velardi (§)

(*) Dipartimento Ingegneria Elettronica, Università di Roma,Tor Vergata
{rbas,pazienza}@tovvx1.ccd.utovrm.it

(§) Istituto di Informatica, Università di Ancona
vela@anvax2.cineca.it

Abstract. Several research groups have been engaged with the problem of automatically acquiring selectional patterns for syntactic and semantic disambiguation from training corpora. More recently, few papers proposed probabilistic word association models to generalize co-occurrence patterns, in order to improve the coverage of the acquired knowledge. Though lately the devised probabilistic models became rather sophisticated, the evaluation of the acquired word clusters is rather disomogeneous and controversial. In this paper the experience made at the NLP laboratory of the University of Roma, Tor Vergata on the acquisition of selectional restrictions with different degrees of expressivity is described. The nature and the coverage of the acquired symbolic knowledge is studied in the light of several experiments of corpus-driven probability-based methods. The integration of the induced information with available on-line thesaurus (e.g. WordNet) is also analyzed.

1. From lexical to conceptual patterns.

Several research groups have been engaged with the problem of automatically acquiring selectional patterns for syntactic and semantic disambiguation from training corpora [Basili et al. 1993b], [Sekine et al, 1992], [Hindle et al. 1993], [Grishman and Sterling,1992], [Yarowsky,1992]. Though all the aforementioned papers report good performances on relatively small test sets (generally of the order of hundred ambiguous sentences), the major problem is coverage. In fact, even when very large corpora are available, reliable co-occurrence patterns are obtained only for a small fragment of the corpus. The problem gets even worst when acquiring word triples (e.g. subject-verb-object, noun-preposition-noun, etc.) rather than pairs.

The growing literature on lexical acquisition systems based on statistical methods is dramatically posing the problem of the required interaction between quantitative and qualitative information about the language (Alshawi,1994). This is particularly true in the case of acquisition of selectional patterns in the above mentioned papers. *Coverage can be effectively improved only when an explanation of single instances is available*, i.e. *when the correct generalization has been obtained*. It is quite clear that any generalization step requires a set of information not simply reducible to collocations, even of syntactic nature.

It is of course true that some methods for generalizing co-occurrence patterns [Brown et al. 1992] [Grishman and Sterling 1994], [Pereira et al, 1993], [Dagan et al. 1994] have been recently proposed. These methods move from word-based (i.e. lexical) to class-based patterns. Meaningful classes are hopefully related to (lexical or already semantic) *concepts*. Some algorithms are referred to as "smoothing techniques", as they augment coverage just by guessing the probability of unseen patterns, by means of class-driven approximation.

An interesting evaluation of the effectiveness of distributional approaches to data sparseness is presented in [Grishman and Sterling, 1994]. In this paper it is argued that smoothing techniques introduce a certain degree of additional error, because co-occurrences may be erroneously conflated in a cluster, and some of the co-occurrences being generalized are themselves incorrect. In general, the effect is a higher recall at the price of a lower precision.

Even if (though not fully demonstrated) distributional clustering techniques are effective at improving the coverage of co-occurrence analyses, a more important limitation of distributional approaches, is that probabilistic clustering algorithms lack of a true explanatory power as quantitative information on collocations is not worth expressive of many linguistic phenomena. The degree of similarity of the clusters is purely numerical (thus implying that their linguistic plausibility can only be tested by inspection). What is acquired is a "blind" criterion for choosing among alternatives, *without any explanation*.

The induced information is thus useful for certain tasks such as syntactic disambiguation, but does not provide a linguistic knowledge amenable for semantic interpretation.

Furthermore, these methods rely on very shallow information as for example syntactic similarity, i.e. co-occurrence in syntactic patterns, and this may give rise to wrong generalizations or meaningless classes.

However, since some group of corageous researchers took the responsibility of designing and making available a full on-line word taxonomy, symbolic classification of words seems to gain new consensus. In particular, many papers [Hearst and Schuetze, 1993] [Beckwith et al, 1991] report on the use of WordNet. Of course, when using WordNet, as well as Roget's and other on-line thesaura, one finds all the

disadvantages of a human entered classification, except for the need of heavy human work.

Table 1 lists the plus and minus of probabilistic vrs. symbolic (human-entered) classification.

Probabilistic clustering methods	Human classification (e.g. WordNet)
(+) fully automatic (+) coherent classification principle (+) domain-appropriate clusters (-) can't evaluate linguistic plausibility of clusters (no symbolic description) (-) estimate of clustering parameters is complex (-) limited coverage (even with large corpora) (-) verbs are a problem (-) applicable to limited tasks	(+) categories have a (symbolic) description (+) coverage is very high (except for proper nouns and technical words) (+) applicable to any task where taxonomic knowledge is required (-) unnecessary ambiguity (given a sub domain) (-) non domain-appropriate clusters (-) verbs are a problem (-) questionable choices. The classification principia are not always coherently applied

Table 1. Plus and minus of probabilistic vrs. manual word classification methods

It is seen that the advantages/disadvantages are almost specular. A specific discussion is necessary for verbs that is the most problematic word class for inductive as well as for fully supervised (i.e. hand-coded) categorization.
There is evidence of a much greater polysemy of verbs with respect to nouns. Furthermore, there exists a very complex relationship between syntactic and semantic information conveyed by verbs. When using pure probabilistic models, data on verbs are too sparse to create reliable categories. We will support this claim with some experimental evidence in section 3. When relying on human intuition, we realize that the classification principia for verbs are much less understood, when not contrasting.
On the other side, verbs are perhaps the most important lexical category in a language. An efficient use of knowledge about verbs is essential for improving the overall efficiency of any NLP system.

2. Combining Statistics with Symbolic Representation in ARIOSTO

In this paper we will sketch the architecture of a possible acquisition system that integrates the capability of a purely inductive method with the explanatory power of available hand-coded lexical knowledge. Results of the acquisition algorithms applied at different levels of lexical representation will be discussed in section 3.

The underlying representational framework is that of ARIOSTO (Basili et al.,1994a), a corpus-driven acquisition system. ARIOSTO allows to derive the generalized argument structure of verb(s) from the incoming sentences of a given corpus. With generalized argument structure we intend the full description of the selectional constraints with which a given verb appear in a sentence, expressed by a system of types-relations that are domain specific. The acquisition of such relations is not a central issue of this paper. Technical details are in (Basili et al.,1994a). The involved relations are valid selectional constraints between words belonging to high-level semantic classes (e.g. Animate or Human_Entity). For example, the sentence

(1) *...moving to Boston by bus ...*

is a typical manifestation of the following selectional constraints of the verb *to move*:

(2) [Act] -(Destination)->[Location]
[Act]-(Instrument)->[Instrumentality]

The acquisition phase that moves from *syntactic triples* like:

(3) G_V_P_N(move,to,Boston)(v pp)
G_V_P_N(move,by,bus) (v pp)

to *semantic relations* in (2), also referred as concept-relation-concept (CRC) triples, is labeled Triple Generalization in Fig. 1. The initial bias of this acquisition step is a set of high-level categories for nouns, like Instrumentality or Location. Actually high level WordNet lexical concepts, i.e. *synsets* (Miller et al., 1993), are used as class labels. A semi-automatic method is used to extract noun classes of interest for a given domain. This phase is labeled Concept Selection in Fig. 1. A simple weighting algorithm is adopted that counts the number of words in the corpus that are subsumed by a label. Higher values are related to most important labels for the domain. A similar algorithm to detect the *best* WordNet classes for an incoming corpus of texts is also reported in (Hearst and Schuetze,1993). A final (and not so expensive) manual inspection suffices to complete the global coverage of the corpus lemmas. High level classes, like INSTRUMENTALITY, ORGANIZATION, NATURAL OBJECT OR COGNITIVE PROCESS have been selected from WordNet by processing a collection of abstracts and technical reports on Remote Sensing topics in English. We will refer to this corpus as the Remote Sensing Domain corpus (RSD).

Whenever semantic class labels are available the interpretation of syntactic triples like (3) into generalized triples (2) is made possible (i.e. step Triple Generalization in Fig. 1). Generalized triples are meaningful relations as (2) that can be used to rewrite/disambiguate the incoming sentences. The adopted role system expresses the canonical relations of a common sense

understanding of the underlying domain (e.g. the (Agentship) of an ORGANIZATION). It provides a domain model able to guide the next phase, i.e. the surface interpretation of single verbal phrases of the corpus. This further step is represented in Fig. 1 by the Shallow Understanding box. The generalized thematic structures of our example (1) is in this phase rewritten into:

(4) (move (destination LOCATION)
(instrument INSTRUMENTALITY)).

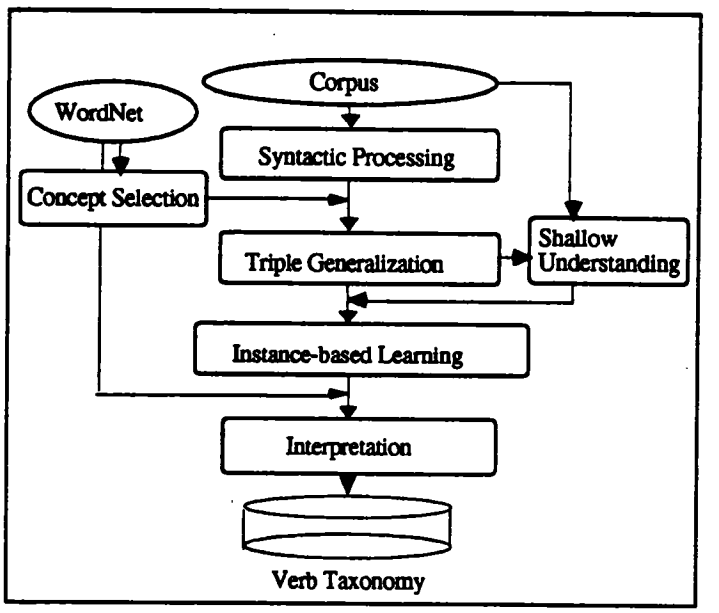

Fig. 1: The architecture of the integrated acquisition system based on ARIOSTO

Note that (4) does not express the verb argument structure but only the typical semantic constraints that rule the use of *move* in sentence (1). Focusing on single sentences is helpful to detect polisemic uses that are quite common for verbs. Distributional methods miss the differences in distinct uses related to independent senses of verbs.

The thematic structure (4) is a feature vector that expresses attributes as roles and values as classes. Learning methods can fruitfully exploit the similarity between feature descriptions of instances. Similar instances are clustered according to the shared conceptual rather than merely syntactic description. An incremental conceptual clustering algorithm, called CIAULA (Basili et al.,1993,1995), a modified version of COBWEB (Fisher,1989) has been used to build verb classes. This step is labeled Instance-based Learning in Fig. 1. Typical examples of classes automatically derived from the Remote Sensing Corpus are in fig. 2. I_AGENT labels the relation "Inanimate Agent".

The clusters are represented by an *extension* of the corresponding verbal class, i.e. the verb set, as well as an *intention*, i.e. the typical expected shared semantic pattern named prototype. The representation includes also weighted verb memberships to the class, according to their *average* behavior in the whole learning set. The first score measures the typicality of the class prototype for the verb. The second measures the verb attitude to appear as a member of the class with respect to all its occurrences in the corpus (i.e. the learning set). Meaningful classes arise spontaneously from the data. Some parameters of the algorithm allow to control the generalization power of a class as well as the average typicality of its members (Basili et al.,1995).

```
Class:1870
PROTOTYPE (i.e., Predicted Roles):
   --   (AFFECTED)  -- [ABSTRACTION]

Verbs (and their 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
PROTOTYPE (i.e., Predicted Roles):
   --  (I_AGENT)  -- [ABSTRACTION]
   --    (MANNER)  -- [PROPERTY,

      COGNITIVE_PROCESS]

Verbs (and their degree of membership):
  illustrate    (0.500 - 1.000)
  calculate     (0.600 - 0.333)
  deal          (1.000 - 0.333)
```

Figure 2. Two CIAULA clusters obtained from RSD

The most informative classes can be easily detected in the resulting taxonomic structure (implied by the overall learning strategy). These classes are selected as useful concepts, i.e. events/status of interest for the underlying domain. In the next section the results for RSD are reported and discussed.

The evaluation problem is twofold. On one side it is important *to numerically evaluate the predictive power* of the derived clusters. In CIAULA, the numeric descriptors of the clusters represent *per se* an evaluation parameter, since they measure the semantic closeness of the cluster members and the generalization of the cluster. On the other side, it is very important *to evaluate as well the linguistic coverage* and *cognitive plausibility* of the resulting taxonomy from an ontological perspective. This problem is particularly complex. In the literature "cognitive" evaluation has been generally approached by manual inspection of the data, that may or may not sound appropriate according to the personal intuition of the human judge. However, any text driven clustering method is inherently missing knowledge about the ontology underlying the language domain. But this has a counterpart for humans: the frustrating difficulty that a human judge has in associating a unifying concept to a flat word cluster depends on the fact that he has too much ontological evidence! Specific choices in a given domain may be (and of course they are) very difficult..

An external source of psychologically motivated deep lexical knowledge could provide a suitable ontological base to the overall learning activity as well as consistently guide the cognitive evaluation. In our study we used WordNet (Miller et al.,1993), as it is a robust, psychologically well founded lexical model. In this stage we rely on the semantic network of verbs in WordNet. An heavy use of the hyponimy/hyperonimy relation can realize the necessary supervision of the previous taxonomic learning. Note that with supervision we do not mean *validation* as a number of acquired new classes may suggest meaningful semantic concepts that are missing in a general purpose network like WordNet.

We defined an algorithm to assign to each verb cluster, generated by CIAULA, the "best" WordNet semantic supertype. The "best" tag is the most specific supertype for the larger number of cluster members.

This stage of acquisition is labeled as Interpretation in Fig. 1, as it allows the system to assign a well defined lexical notion (or sense in WordNet terms) to the acquired clusters. The labeling activity configures as a search into a space of concepts that are possibly valid for a group of verbs. Maximizing the number of subtypes that a label has in the class, the system is able to focus on the more plausible interpretation of a given class. The corresponding synset is used as an entry point in WordNet to collect: further synonymies of the class members as well as additional information like the verb suggested argument structure. Note that a class prototype combined with the corresponding synset argument structure can easily produce a specific Syntax-Semantic mapping. Relations other than hyponimy/hyperonimy may also be of interest. Our experiments did not take them into account as they are not yet fully implemented in WordNet.

3. Experimental set-ups.

In the previous section an overall architecture to acquire a domain specific classification of lemmas in a given sub language has been proposed. Many claims have been only sketched. Although this kind of research has several open issues, a wide experimentation on two different corpora gave very interesting results in the area of coarse grained semantic patterns acquisition as well as in the induction of a verb taxonomy. Two sublanguages have been selected: a corpus of legal texts in Italian on the V.A.T. law and the mentioned RSD corpus in English. The size of the two corpora is of about 400.000 words. Table 2 shows data on the number of syntactic triples (as in (3)) and generalized triples (or CRC triples, as in (2)) that have been obtained from the syntactic preprocessing of the two corpora.

	Legal Corpus	Remote Sensing Corpus
Corpus size	428,380	315,419
Semantic tags	13	19
Syntactic triples	166,386	43,333
CRC triples	198	329

Table 2. Words, Semantic Classes and Triples in two sublanguages.

A first analysis has been performed in order to evaluate the feasibility and complexity of the verb taxonomy acquisition task. Verbs appear in the corpus with a variegated and poorly overlapping argument structure. Table 3 shows data on the average number of verb selectional constraints of the type in (2) that have been derived from the two different domains. We can see that independently from their frequency in the corpus verbs show several semantic patterns. As each CRC pattern derives from an average of 2.3 different syntactic structures, the typical argument structure is shown idiosyncratic even for low frequency verbs.

The task of observing regularities is prohibitive in such a sparse sea of information, even though semantic patterns are probably the best quality information that we can observe in raw texts. Furthermore, more granular classes of verbs are of great interest not only from a lexical point of view, that is for tasks like morphologic, syntactic or word sense disambiguation, but mainly from a semantic point of view: the acquisition of meaningful verb classes allows the induction of semantic primitives related to the underlying domain. For example, a legal expert system using a natural language interface would profitably make use of a separate category for verbs expressing `movement_of_money`.

Verb Frequency Ranges	Legal Domain		Remote Sensing Domain	
	Average # relations per verb	# of different detected relations	Average # relations per verb	# of different detected relations
x≤10	3.76	81	6.4	76
10<x≤100	9.8	85	17.9	80
x>100	27	85	35.2	80

Table 3. Average number of selectional restrictions per verb in two sublanguages

In order to study the feasibility of acquisition of verb semantic classes we applied CIAULA to both the legal and RSD domains. For the first a test set of about 962 (real) sentences of the corpus have been used. The 962 sentences include a set of 184 different verbs. Each verb has thus an average frequency of 5.01 within the learning set. The acquired classification includes 59 basic level classes extracted out of the global hierarchical structure. An example of basic level class is shown in Fig. 3.

```
Class: 2414
Intention:
        [2414:*]-(Affected)->[Goods:*]
        [2414:*]-(Affected]->[Amount_of_Money:*]
Extension:      assegnare (*assign)
        alienare  (*alienate)
        acquisire (*acquire,*buy)
        comprare (*buy)
        erogare (*erogate)
```

- Fig. 3: An example of basic level verb class derived by CIAULA in the legal domain -

In RSD we performed an equivalent experiment on a set of 560 source verbal phrases of 167 different verbs, with an average frequency of 3.35 occurrences per verb. The resulting classification includes 34 basic level classes grouping 60% of the analyzed verbs. Remaining verbs are markers of sparse relations that was not possible to cluster in meaningful way. The lower frequency of the English verbs in the test set is a potential explanation for this lower coverage. However, a deeper insight in the results as well as an evaluation guided by the Interpretation phase is useful to stress essential positive aspects.

In Fig. 4 two basic level classes are shown together with the corresponding WordNet labels generated during the Interpretation phase. The suggested interpretation is very accurate. Labels are mildly general, but peculiar aspects of verb meanings are further clarified by the prototypical semantic pattern. The source text (i.e. the RSD corpus) suggests specific senses of verbs, that are wide spread captured from the data. For example the verb *derive* looses in the Class 12 its physical meaning of movement from a place to another that is quite common. It is instead related to the abstract passage from a source problem/phenomena (e.g. a mental activity) to a target MENTAL OBJET or COGNITIVE PROCESS by the agentship of an ABSTRACTION. Similarly it is captured the abstract sense of *calculate* in Class 25. Its meaning of evaluation/cogitation (as for example in '... *automated and used to calculate or adjust the calibration* ...') is kept distinct from the meaning of 'production of a result' (as in '... *the method calculates both the amplitude image and the radar signal* ...') that is explicitly captured in class 12.

A similar example for the Italian corpus has been obtained by a flat mapping of Italian verbs into English translations. WordNet here is used for Italian language as high level concepts may be shared by the two languages. As many translations exist for each Italian verb (about 3 per verb), the process is less precise. The algorithm is forced to converge in any case: as a consequence we derive slightly higher level tags.
As an example the interpretation of the class in fig. 3 is 'get, acquire, enter upon, come upon, get possession of' that is the WordNet synset mostly close to the (conceptual) notion of movement_of_money.

```
Class: 12
WN_label:     'decide, make up one's mind,
              decide upon, determine'

Intention: [12:*] -(Agent)-> [*:ABSTRACTION]
Extension:  relate      compute      solve
        compare         measure      evaluate
        determine       study        derive
        include         plan         identify
        describe        calculate
WN Argument Structure:
        Somebody ---s something
```

```
Class: 25
WN_label:     'think, cogitate, cerebrate'

Intention: [25:*] -(Affected)-> [*:COGNITIVE PROCESS]
Extension:  calculate   base         plot
        approximate     survey       analyse
        plan            record       select
        locate          estimate     describe
        deal            review       include
        derive          measure      compare
        determine       view
WN Argument Structure: Somebody ---s
```

Fig. 4: Verb clusters and their interpretation in RSD

Different acquired classes may be assigned with the same label. In these cases, the different semantic patterns assigned to the class explicitly suggests the different semantic aspects that make the related verbs different. An example for the class of verbs labeled 'decide, make up one's mind, decide upon, determine' is partially shown in fig. 5.

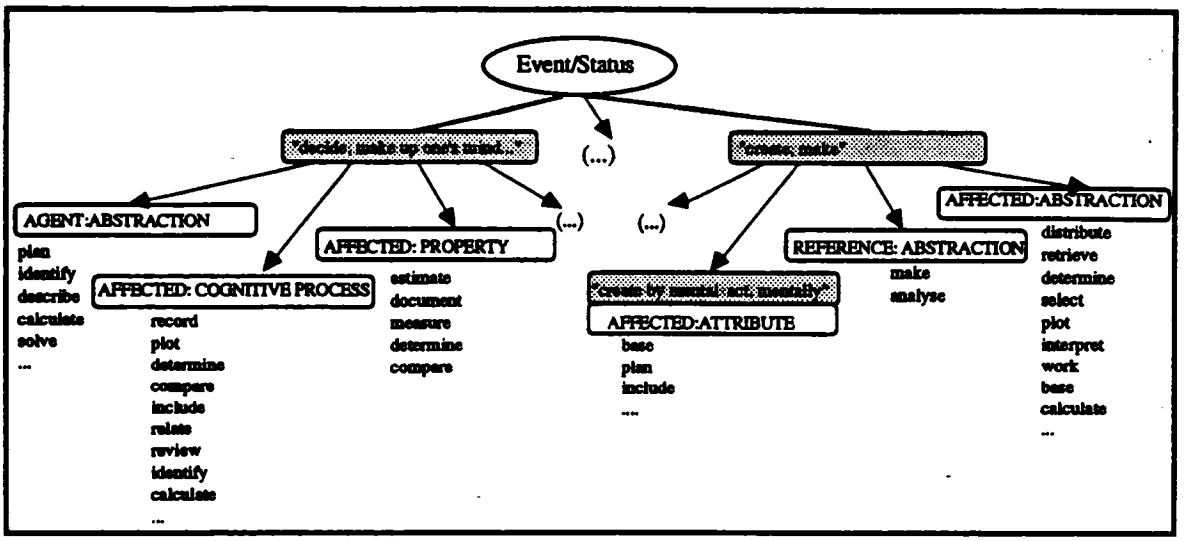

Fig. 5: A portion of the acquired verb taxonomy.

The resulting taxonomy is thus built under a default node labeled 'Event/Status'. Note that basic level classes sharing the same labels are merged. This information is a consequence of the interpretation process. *This inference relies on the hypothesis that higher levels of the WordNet taxonomy generalize across different sublanguages.* An interesting phenomenon arises in this case. The hyponimy relations that links different labels in WordNet are retained in the final taxonomy, and they give rise to multiple levels. One such relation links the nodes 'create,make' and 'create by mental act, create mentally' in fig. 5.

It is then possible to affirm that the combination of a context-driven clustering method with a cognitive-driven, general purpose classification provides a systematic and noise free interpretation of the data. Furthermore, using as source a data-driven classification preserves a desirable domain specificity, thus implying a suitable data compression. The resulting taxonomy includes classes that are described by their:

- extension, i.e. the set of members;
- intention, the semantic prototype;
- syntactic features as the predicted argument structure
- additional semantic relations, e.g. troponimy or opposition relations.

4. Conclusion.

In this paper the architecture of an integrated lexical acquisition system based on corpus driven techniques as well as on hand coded knowledge has been described. The derivation of generalized patterns in terms of concept-relation-concept (CRC) triples and the generation of a taxonomy of verb classes is shown feasible. The methods and algorithms described are portable through different linguistic domains. Only minimal changes to limited amount of semantic information required to trigger the acquisition process are needed. An Italian as well an English corpus have been processed according to the same principles (relying on an independent morphology and syntax). The derived knowledge has a number of interesting features that makes it suitable for linguistic as well non-linguistic tasks. It is very specific to the knowledge domain related to the corpus (e.g. the Remote Sensing discipline). The acquired verb classes are absent from the WordNet knowledge base, as they include generally many more verbs according to the more specific contexts (i.e. thematic structures) observed in the corpus. Sometimes the acquired classification completely disagrees with WordNet including verbs that are not part (or hyponimies) of a common synset. Further knowledge is the specific (semantic) thematic structure predicted for a class that is absent in WordNet.

In the acquired taxonomy there is less information as well as a lower degree of ambiguity. In our test set, we measured 1.04 sense per verb after the Interpretation phase, against 4.5 senses in WordNet. Of course, we loose the suitable synonymy relation linking the verbs in any WordNet synset, but we have larger sets with a different meaning: a common underlying concept that is easily derivable from the prototype, i.e. the shared thematic structure. Even if the coverage of the resulting taxonomy is not full, the first results are very encouraging. A possible solution to the coverage problem may be to augment the quality of the source information by using, possibly incomplete, outputs of a broad-coverage parser in order to rely on richer thematic structures. An open problem is the effective expressiveness of the resulting classes: the interaction between WordNet synset argument structures and the thematic structures assigned, as a prototypical feature, to a class should provide more precise (and deep) lexical rules. This issue requires obviously further insight.

Acknowledgment. The authors would like to thank Carlo Consoli for his great effort in the

implementation of the WordNet labeling algorithm.

References

(Alshawi,1994) H. Alshawi, Qualitative and Quantitative Designs for Speech Translation, The Balancing Act: Combining Symbolic and Statistical Approaches to Language, ACL Workshop, Las Cruces, New Mexico, USA, July 1994.

(Basili et al.,1993a) Basili, R., M.T. Pazienza, P. Velardi, Acquisition of selectional patterns, Journal of Machine Translation, 8:175-201,1993.

(Basili et al.,1993b) Basili, R., M.T. Pazienza, P. Velardi, Semi-automatic extraction of linguistic information for syntactic disambiguation, Applied Artificial Intelligence, vol. 4, 1993.

(Basili et al.,1993) Basili, R., M.T. Pazienza, P. Velardi, Hierarchical clustering of verbs, ACL-SIGLEX Workshop on Lexical Acquisition, Columbus Ohio, June, 1993.

(Basili et al.,1995) Basili, R., M.T. Pazienza, P. Velardi, A context driven conceptual clustering method for verb classification, in B. Boguraev, J. Pustejovsky Eds., MIT Press, 1995.

(Beckwith et al.,1993) R. Beckwith, C. Fellbaum, D. Gross, G. Miller, WordNet: A Lexical Database Organized on Psycholinguistic Principles, in Lexical Acquisition: Exploting On-Line Resources to Build a Lexicon, U. Zernik Ed., Lawrence-Erlbaum Ass., 1991.

(Brown et al., 1992) Brown P., Della Pietra V.J., deSouza P.V., Lai J.C., Mercer R., Class-based n-gram Models of Natural Language, Computational Linguistics, 18, n. 4, 1992.

(Dagan et al.,1994) Dagan I., Pereira F., Lee L., Similarity-based Estimation of Word Co-occurrences Probabilities, Proc. of ACL, Las Cruces, New Mexico, USA, 1994.

(Grishman and Sterling,1992) R. Grishman, J. Sterling, Acquisition of selectional patterns, Proc. of COLING '92, Nantes, July 1992.

(Grishman and Sterling,1994) R. Grishman, J. Sterling, Generalizing Automatically Generated Selectional Patterns, Proc. of COLING '94, Kyoto, August 1994.

(Hearst and Schuetze,1993) M. Hearst and H. Schuetze, Customizing a Lexicon to Better Suite a Computational Task, ACL SIGLEX, Workshop on Lexical Acquisition from Text, Columbus, Ohio, USA, 1993.

(Hindle 1990) D. Hindle, Noun Classification form predicate-argument structure, Proc. of ACL 1990

(Hindle and Rooths,1993) D. Hindle, M. Rooths, Structural ambiguity and lexical relations, Computattional Linguistics, March 1993.

(Pereira et al.,1993) Pereira F., N. Tishby, L. Lee, Distributional Clustering of English Verbs, Proc. of ACL, Columbus, Ohio, USA, 1994.

(Sekine et al., 1992) S. Sekine, Carrol J.J., Anadianou S., Tsujii J., Automatic Learning of Semantic Collocation, Proc. of Third ANLP, Trento, Italy, 1992.

(Yarowsky, 1992) D. Yarowsky, Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora, Proc. of COLING '92, Nantes, July 1992.

(Zernik,1989) U. Zernik, "Lexical acquisition: Learning from corpus..", in Proc. of 11th IJCAI, Detroit, Michigan,1989.

(Webster,1989) Webster M., M. Marcus, Automatic Acquisition of Lexical Semantics of Verbsfrom Sentence Frames, Proceedings of ACL, Vancouver, 1989.