

# An empirical approach to Lexical Tuning

Roberto Basili (\*), Roberta Catizone (◇), Maria Teresa Paziienza (\*),  
Mark Stevenson (◇), Paola Velardi (#), Michele Vindigni (\*), Yorick Wilks (◇)

(\*) University of Roma, Tor Vergata, Italy  
(◇) University of Sheffield, United Kingdom  
(#) University of Roma, La Sapienza, Italy

**Abstract.** NLP systems crucially depend on the knowledge structures devoted to describing and representing word senses. Although automatic Word Sense Disambiguation (WSD) is now an established task within empirically-based computational approaches to NLP, the suitability of the available set (and granularity) of senses is still a problem. Application domains exhibit specific behaviors that cannot be fully predicted in advance. Suitable adaptation mechanisms have to be made available to NLP systems to tune existing large scale sense repositories to the practical needs of the target application, such as information extraction or machine translation. In this paper we describe a model of "lexical tuning"—the systematic adaptation of a lexicon to a corpus—that specializes the set of verb senses required for an NLP application, and builds inductively the corresponding lexical descriptions for those senses.

## 1 Word Sense Disambiguation and Lexical Tuning

It is a commonplace observation (and the basis of much research e.g. [Riloff and Lehnert1993]) that lexicons must be tuned or adapted to new domain corpora. This aspect, now often called Lexical Tuning, can take a number of forms, including:

- (a) adding a new sense to the lexical entry for a word
- (b) adding an entry for a word not already in the lexicon
- (c) adding a subcategorization or preference pattern etc. to any existing sense entry

The system we describe is an original architecture for the overall task of corpus-based lexical tuning.

This task is of general theoretical interest, but one that it is difficult to test directly, as a distinct NLP task, largely because of the difficulty of incorporating the phenomenon into the standard markup-model- and-test paradigm of current empirical linguistics. A central issue in any application of empirical methods to computational linguistics is the evaluation procedure used, which is normally taken to consist in some form of experiment using premarked-up text divided into training and (unseen) test portions. Apart from the well-known problem of the difference between sense-sets in different lexicons, there are problems concerned with subjects having difficulty assigning a word occurrence to one and only one

sense during this markup phase. Kilgarriff [Kilgarriff1993] has described such problems, though his figures suggest the difficulties are probably not as serious as he claims [Wilks1997]. However, we have to ask what it means to evaluate the process of Lexical Tuning: this seems to require annotating in advance a new sense in a corpus that does not occur in the reference lexicon, when developing gold standard data for testing basic WSD. The clear answer is that, on the description given above, the sense extension (task (a) above: tuning to a new sense) CANNOT be pre-tagged and so no success rate for WSD can possibly exceed 100% MINUS the percentage of extended sense occurrences.

One issue about lexical tuning that is not often discussed is: what the percentage of senses needing tuning IS in normal text?

One anecdotal fact sometimes used is that, in any randomly chosen newspaper paragraph, each sentence will be likely to have an extended sense of at least one word, usually a verb, in the sense of a use that breaks conventional preferences and which might be considered extended or metaphorical use, and quite likely not in a standard lexicon. This is a claim that can be easily tested by anyone with a newspaper and a standard dictionary.

The assumption under test in our project is that lexical tuning will assist the adaptation of a NLP task (e.g. Information Extraction) to a new domain and can therefore best be tested indirectly by its augmentation of the target system performances. There is already substantial evidence that some form of word sense disambiguation (WSD) assists any NL task, when applied as a separate module, and lexical tuning can be seen as a more advanced form of WSD. A second assumption - not addressed in literature - is that a tuned lexicon can significantly help the task of automatic pattern acquisition for template filling in an IE system. Currently, this task is largely performed by hand, with the help of more or less sophisticated interfaces [R. Yangarber1997]. The key idea adopted here is that an established initial lexicon can be tuned or adapted for verb senses in a given application domain. First verb occurrences in a corpus are distributed over a classifier that clusters their subcategorization patterns. This distribution allows a judgment of when a new pattern in the corpus, and not in the initial lexicon, should be assigned to an existing sense of the target dictionary, or established as a

new sense to be added to it.

## 2 A general architecture for lexical tuning

The proposed Lexical Tuning system first processes a corpus with a tagger [Brill1992] and shallow parser<sup>1</sup>. The structures so derived (essentially the subcategorization patterns (hereafter subcat) of individual verb occurrences in the corpus) are distributed over a lattice structure, called a Galois Lattice (hereafter *RGL*), by an inductive method described in [Basili et al.1997], and briefly summarized in the next section. This is a device by which each occurring set of syntactic properties, for a given verb, is assigned to one node on the lattice, in an appropriate position for partial orderings with respect to all other subcat distributions. It is thus a sorting frame, with set inclusion relations, for the contexts of each appearance of the verb in the corpus.

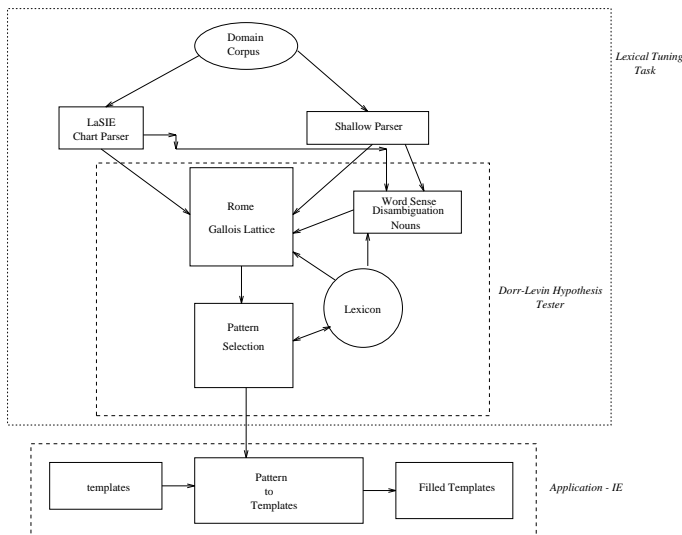


Figure 1. The functional architecture for Lexical Tuning

## 3 Learning subcategorization patterns from corpora

Methods of clustering are widely adopted within the machine learning community for example-driven learning tasks [J.H. Gennari1989]. They are generally based on incremental search within concept (or class) spaces. The target problem here is the derivation of the subcategorization frame for each verb observed in the source training text. In this case a separate search space is used for each verb. In particular:

1. All the verb phrases associated with a given verb are collected from the corpus.

<sup>1</sup> Two available shallow parsers [Gaizauskas et al.1995] and [Basili et al.1992] are used for comparison and calibration

2. Clusters of similar verb behaviors are organized into a hierarchical structure (e.g. a lattice)
3. Finally, a set of valid subcategorization rules corresponding to some derived classes are selected: clusters that suggest separate verbal senses and their corresponding grammatical constraints.

Clustering techniques based on conceptual lattices derive classes as conjunctive concepts according to a boolean feature-value language. Each derived concept is a couple  $(S, F)$  where  $S$  is a subset of instances, sometimes called an *extension*, and  $F$  is the set of features of the cluster, correspondingly called an *intension*.

The following representation is adopted to express the context information: grammatical relations observed in a sentence are attributes of the corresponding instance, and prepositions are used to indicate associated phrases to a verb in a standard manner (except for verb subject and direct object).

When the *RGL* is built it is necessary to select the nodes corresponding to true verb subcategorization frames. This inference depends on the whole lattice structure and on the intension  $F$  of the underlying node. Two measures are defined in [Basili et al.1997]: *selectivity* of a node and *linguistic preference*. The first relates to the amount of information at the node with respect to the whole lattice structure. The second expresses the utility of an intension  $F$  as a valid subcategorization frame. A threshold is imposed which is a combination of the two measures: nodes assigned to values over the threshold are retained as valid subcategorization patterns. As an example the patterns valid for the verb *hire* from a set of 200 source contexts<sup>2</sup> are the following:

```
((subj X) hire )
((subj X) hire (obj Y))
((subj X) hire (for Y))
((subj X) hire (obj Y) (for Z))
```

The extracted patterns of the verb *hire* are plainly not independent. In this case, the structure of the lattice includes the corresponding subsumption relations (i.e.  $((\text{subj } X) \text{ hire } (\text{obj } Y) (\text{for } Z))$  implies all the other patterns). Each pattern is characterized by its intension, the empirical score of the node and the set of its members, i.e. the contexts sharing that pattern. Technical details can be found in [Basili et al.1997].

## 4 Relating Patterns to senses

The patterns acquired by the process outlined in the previous section lack some relevant information:

- source corpus data can be noisy, so that wrong and incomplete patterns may have been acquired;
- semantic information pertinent to the arguments in the patterns is missing, with the result that several verb senses may have been collapsed into the same pattern;
- verbs characterized by low frequency in the corpus are likely to be described by incomplete information, which is to say, missing patterns or partial patterns.

In this section, we describe an automatic method to:

<sup>2</sup> The parsed sentences in the Wall Street Journal are here used as the source.

- add semantic information to the extracted grammatical patterns
- assign these augmented patterns to the dictionary senses
- match source instances against the extended patterns, and
- assign to source sentences those senses pertinent to the valid patterns.

The result of this whole phase is a set of dictionary senses augmented by related semantically annotated patterns from the corpus together with a corresponding reordering of senses on the base of their frequency in the training texts. Note that for some of the patterns the assignment to sense(s) in the dictionary may be problematic. These patterns are candidates for potential new senses that will be initially described as semantically annotated subcategorization frames.

#### 4.1 Using patterns to improve WSD

Selectional restrictions have been used in many word sense disambiguation system [Wilks1975] [McCarthy1997] [Resnik1997]. However, [McCarthy1997] has described a 'vicious circle' in this enterprise: selectional restrictions help WSD, but WSD is needed to acquire high quality selectional restrictions. One way to avoid this is use the restrictions in some lexical resource as a first approximation and augment, or tune, them with the patterns found in a corpus. For example, the Longman Dictionary of Contemporary English [Procter1978] contains subcategorization patterns and selectional restrictions indicating preferred semantic classes for verb and modifier arguments. The integration between the source dictionary  $d$  and the acquired patterns (set  $P$ ) is possible given appropriate syntactic information available in  $d$ . In the Longman Dictionary of Contemporary English (LDOCE) then available syntactic information includes:

- subcategorization patterns
- selectional restrictions in terms of the semantic classification of arguments

Table 1. Semantic classes of arguments for verb *hire*

hire	<i>H</i>	<i>Org</i>	<i>T</i>	<i>M</i>	<i>U</i>	<i>J</i>
Subject	33%	20%	10%	2.9%	2.9%	-
Object	50%	0.9%	2.9%	8%	10%	4.8%

Each verb sense thus includes a structure where the number of arguments is specified, and an expected semantic type is expressed. For example, the verb *hire* has two senses with the following descriptions:

```
hire_2_1 ((subj X:Human) hire (obj J:Movable))
    %to get the use of (something) for
    %a special occasion on payment
hire_2_2 ((subj X:Human) hire (obj H:Human))
    %to employ (someone) for a time
    %for a payment
```

The *RGL* lattice currently clusters verbs according to their syntactic subcategorization. However, each node in the Galois lattice for a given verb  $v$  represents a set of contexts of  $v$  whose descriptions show all the attributes (i.e. detected arguments)

of the node. So the node for the ((subj X) hire (obj J)) patterns refers to source examples like

```
'... creditors hired experts ...'
'... Apple hired president ...'
'... office hire reporters ...'
'... network hired Hill ...'
```

At first glance the distribution of semantic classes in the examples is not random. Typical subjects and objects are *Humans*. Data in Table 1 has been obtained by applying the sense tagger to each context covered by the related node of the *hire* lattice: the distribution of different LDOCE classes over the subject and object relations is shown<sup>3</sup>.

The semantic classes of arguments are generally ambiguous in such dictionaries. For example, *expert* is vaguely classified as an "Not concrete or animal" in LDOCE. Although some noise is introduced by the semantic tagger, the sense distribution of each argument converges to a reasonably selective preference via the simple relative frequency. For instance, among the source senses of LDOCE, only the sense *hire\_2\_2* is retained because it fits the most frequent semantic types for the two arguments. Corpus sentences appear to be reducible to that sense (see the *expert* example). This coercion imposed on underspecified semantic types is alone an important aspect of this method. Manual analysis of the examples shows that no sentence in the 200 training instances from the WSJ uses the other sense of *hire*.

As we previously remarked, the device to modify the lexicon is still under development. However, an intuitive way to evaluate the efficacy of the architecture is to use semantically annotated patterns to improve verb sense disambiguation, effectively evaluating the quality of the selectional restrictions derived. We use the following algorithm to carry out this evaluation.

```
For each verb  $v$  in the training set  $T$ 
  Build the set  $patt(v)$  of its
  subcategorization patterns
Each  $p \in patt(v)$  is a triple  $(args, pref, ex)$  with
   $args$  = list of arguments
   $pref$  = empirical preference score
   $ex$  = set of examples
For each  $p \in patt(v)$ 
  For each  $e \in ex$ 
    For each  $a \in args$ 
      Tag  $a$  in  $e$  by the semantic tagger
  For each  $a \in args$ 
    Build its distribution  $\pi$  of semantic classes
  Build the set  $sempatt(v, p)$  of semantically
  annotated patterns (by thresholding over  $\pi$ )
  Map each  $sp \in sempatt(v, p)$  in one LDOCE sense  $ldcs$ 
  Attach  $sp$  to the set of acquired domain patterns of  $ldcs$ 
  If no such  $ldcs$  exists
    guess  $sp$  as an early new verb sense for  $v$ 
```

Semantic tagger here refers to our word sense disambiguation engine, applied to head nouns in the verb phrases, and which gives at present the best overall WSD rate for open-class words in general text [Stevenson et al.1998]. The use of the WSD semantic tagger is a key element here in the ability of the system to rate new patterns not in the original dictionary but sufficiently close to be retained by the RGL within

<sup>3</sup>  $H$ ,  $Org$ ,  $T$ ,  $M$ ,  $U$  and  $J$  denote the class "Human", "Organization" (this named entity class has been created, "Abstract"(ion), "Male and Human", "Collective, Animal or Human" and "Movable", respectively

the existing sense set. We can see in the examples above that there is a range of examples with subject marked ORG.

At the end of the analysis of the verb *hire*, the following data are acquired. *hire\_2\_2* is extended via the patterns

```
((subj X:Org) hire (obj Human))
((subj X:Org) hire (obj Human) (for Z:Any))}
```

As a set of potential new senses the following patterns are retained:

```
{ ((subj X:Human) hire ) }
```

## 5 Early experimental evidence

The algorithm of Section 4.1 has been applied to a set of verbs characterized by [Levin1993] as “*Change of Possession* verbs”. 10 verbs have been selected and all their source contexts processed. These 10 verbs are *accept, collect, seize, keep, hire, save, obtain, select, win, earn*. The *RGL* module and the semantic tagger have been applied to the sentences and the output patterns gave rise to **2899** semantically annotated patterns (an average of about **289.9** patterns per verb). The assignment of senses to the augmented pattern was possible for an about 2500 of the 10 verbs totalling 52 senses (an average of 5 senses per verb): a semantic tag has been considered valid for an argument if its relative frequency over the source contexts was over the 18%. (We are currently experimenting with ways of optimising the value of this threshold.) An average of around 40 corpus examples were found for each subcategorization frame. LDOCE senses have been assigned to those patterns for which a consistent mapping between expected argument types in LDOCE and the observed semantic class derived from the examples could be found. In addition, we took into account the named-entity class of ORGANIZATION which is significant in the financial domain. The issue of integrating named-entities into the semantic classification scheme has not been fully resolved, but our aim is to add the significant named-entities to our lexicon, thus tailoring the classification scheme to our domain. Although this may diminish the general nature of our approach, named entities need handling specially in most NLP applications.

Some constraints were relaxed, as for example merging counts for classes like “*Collective, Animal or Human*” and “*Human*”. A sense is considered *valid* if at least one semantic pattern consistent with it has been found. 4-10% of patterns were proposed as potential new senses. Conversely, 50-60% of LDOCE senses were rejected, as no corresponding corpus pattern could be found.

There were two verbs that appeared to have new senses. They were ‘seize’ and ‘save’. The example sentences containing these 2 verbs have verb argument preferences that are not in the sense in the lexicon. Below is a list of the senses of *seize* with a subject (\$E), and object preferences(\$F) and the dictionary definition (\$I). In our financial domain the meaning of *seize* is different than those in the lexicon below and hence this fact is reflected in the verb argument pairs. The closest sense is *seize\_0\_4*, but even that does not adequately reflect the use of ‘seize’ in the financial domain. From the corpus examples the most frequently occurring argument classes were a Human subject and an Abstract object

```
seize_0_2
$E O (Animal or Human)
$F Z (Abstract)
$I to take hold of eagerly , quickly , or forcefully ; GRAB ;
```

```
seize_0_3
$E T (Abstract)
$F H (Human)
$I to attack or take control of (someone’s body or mind) ; OVE
```

```
seize_0_4
$E X (not concrete or Animal)
$F H (Human)
$I [fml or law] to give ownership (of property) to
```

```
seize_0_1a
$E Z (Unmarked)
$F Z (Unmarked)
$I to take possession of by official order
```

```
seize_0_1b
$E Z (Unmarked)
$F Z (Unmarked)
$I to take possession of by force
```

The system shows an interesting behavior for the verb *accept*. LDOCE senses are the following:

```
accept_0_1 ((subj X:Human) accept (obj Z:Any))
%to accept as a gift
accept_0_2 ((subj X:Human) accept (obj T:Abstract))
%to believe, admit, agree
accept_0_3 ((subj X:Human) accept (obj T:Abstract))
%to take responsibility for
```

Note that the Levin’s classes for this verb are *Change of Possession.Obtain* (13.5.2) (*accept\_0\_1* and *accept\_0\_2*), and *Predicative Complements.Characterize* (29.2) (*accept\_0\_3*). The senses present in the source sentences are *accept\_0\_2* and *accept\_0\_3*, with a higher frequency for the first. These are exactly captured by the patterns extracted from the source sentences:

```
((subj X:Human) accept (obj T:Abstract))
```

Note that the score assigned to the two patterns is the same (as the same frequency is used for the arguments of the two frames), so that sorting of senses has no impact in this case. However, if the method could make use, as we plan, of other grammatical features (e.g. predicative complements) a more precise separation between the two senses could be made by the proposed method.

One last point is to say that we also found that 90 analyzed contain a ‘Human subject’ and an ‘Abstract’ object. This provides supporting evidence justifying Levin’s clustering of these verbs into a single class based on syntactic behavior. More analysis will be necessary in order to make a general claim about the regularity of verb-argument structure for the remaining Levin classes, but the initial results are encouraging.

## 6 Discussion and future work

The selectional preferences derived could be used to improve the sense tagging results. Since the tuned lexicon has been

adapted to more accurately reflect the semantics of the corpus it is likely that our sense tagger will achieve better results when tagging against this lexicon compared to a more general, untuned, lexicon. We could, then, sense tag the text with the untuned lexicon, use the information in the tags to tune the lexicon, then repeat the tagging process, with better results. The re-tagged text could then be used to tune the lexicon further. This process could be repeated with the sense tagger providing more accurately tagged text for the lexical tuning mechanism which would, in turn, provide a better tuned lexicon to use for semantic tagging. These experiments show that by taking currently available semantic tagging and corpus analysis algorithms, it is possible to discover novel senses, not in dictionaries.

## 7 References

- [Basili et al.1992] R. Basili, M.T. Pazienza, and P. Velardi. 1992. A shallow syntactic analyzer to extract word associations from corpora. *Literary and Linguistic Computing*, 7(2).
- [Basili et al.1997] R. Basili, M. Vindigni, and M.T. Pazienza. 1997. Corpus-driven unsupervised learning of verb subcategorization frames. In *Proc. of Conference of the Italian Association for Artificial Intelligence, AI\*IA 97, Rome*.
- [Brill1992] E. Brill. 1992. A simple rule-based part-of-speech tagger. In *Proceeding of the Third Conference on Applied Natural Language Processing*, Trento, Italy.
- [Gaizauskas et al.1995] R. Gaizauskas, T. Wakao, K Humphreys, H. Cunningham, and Y. Wilks. 1995. Description of the LaSIE system as used for MUC-6. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*. Morgan Kaufmann.
- [J.H. Gennari1989] P. Langley D.H. Fisher J.H. Gennari. 1989. Models of incremental concept formation. *Artificial Intelligence*, 40.
- [Kilgarriff1993] A. Kilgarriff. 1993. Dictionary word sense distinctions: An enquiry into their nature. *Computers and the Humanities*, 26:356–387.
- [Levin1993] B. Levin. 1993. *English Verb Classes and Alternations*. Chicago, IL.
- [McCarthy1997] D. McCarthy. 1997. Word sense disambiguation for the acquisition of selectional restrictions. In *Proceedings of ACL/EACL Workshop on Automatic Information Extraction and building of lexical semantic resources*, pages 52–60, Madrid, Spain, July.
- [Procter1978] P. Procter. 1978. *Longman Dictionary of Contemporary English*. Longman Group.
- [R. Yangarber1997] R. Grishman. R. Yangarber. 1997. Customization of information extraction systems. In *Proceedings of Workshop on Lexically-Driven Information Extraction*, Dept. of Computer Science, University of Roma "La Sapienza".
- [Resnik1997] P. Resnik. 1997. Selectional preferences and word sense disambiguation. In *Proceedings of the SIGLEX Workshop "Tagging Text with Lexical Semantics: What, why and how?"*, pages 52–57, Washington, D.C., April.
- [Riloff and Lehnert1993] E. Riloff and W. Lehnert. 1993. Automated dictionary construction for information extraction from text. In *Proceedings of Ninth IEEE Conference on Artificial Intelligence for Applications*, pages 93–99.
- [Stevenson et al.1998] M. Stevenson, H. Cunningham, and Y. Wilks. 1998. Sense tagging and language engineering. In *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI-98)*, pages 185–189, Brighton, UK.
- [Wilks1975] Y. Wilks. 1975. A preferential pattern-seeking semantics for natural language inference. *Artificial Intelligence*, 6:53–74.
- [Wilks1997] Y. Wilks. 1997. Senses and texts. *Computers and the Humanities*.