Clustering Adjectives for Class Acquisition

Gemma Boleda Torrent
GLICOM
Departament de Traducció i Interpretació
Universitat Pompeu Fabra
gemma.boleda@trad.upf.es

Laura Alonso i Alemany
GRIAL
Departament de Lingüística General
Universitat de Barcelona
lalonso@lingua.fil.ub.es

Abstract

This paper presents an exploratory data analysis in lexical acquisition for adjective classes using clustering techniques. From a theoretical point of view, this approach provides large-scale empirical evidence for a sound classification. From a computational point of view, it helps develop a reliable automatic subclassification method.

Results show that the features used in theoretical work can be successfully modelled in terms of shallow cues. The resulting clusters parallel to a large extent with proposals in the literature, which indicates that automatic acquisition of adjective classes for large-scale lexicons is possible.

1 Introduction

This paper reports on experiments applying clustering techniques to explore the behaviour of Catalan adjectives in running text. The objectives of the exploratory data analysis were twofold: from a theoretical point of view, to get large-scale empirical insight in order to develop a sound classification; and from a practical perspective, to test whether the features mentioned in the literature could be successfully modelled in terms of shallow cues, which would allow an automatic classification.

A sound classification of adjectives should allow one to predict morphological, syntactic and semantic properties of particular items. This predictive power can be exploited in several NLP tasks (see Section 3).

Bootstrapping techniques have been recently applied to German adjective class acquisition (Bohnet et al., 2002). In contrast, we have taken an unsupervised approach in order to test the classes proposed in the literature and investigate their properties (see Section 2). Clustering is suitable for linguistic investigation in classification tasks (Pereira et al., 1993), and has been specially applied in the lexical domain, where no consensus on classification criteria has been reached yet (see Schulte im Walde and Brew (2002) and Merlo and Stevenson (2001) on verb classes and Hatzivassiloglou and McKeown (1993) on adjectival scales). Clustering can be used as an exploratory technique that provides insight into the organization of such domains, by finding classes of homogeneous objects and the features that crucially characterise them.

The initial hypothesis was a three-way classification based on proposals in the literature (Section 2). Features discussed in theoretical work were modelled in terms of shallow cues (Section 4) and a series of clustering experiments were performed over more than 3500 lemmata from a tagged corpus (Section 5). The results obtained support the initial hypothesis, with caveats (Section 6).

2 Catalan Adjective Classes

After a review of the literature on adjective classification (Hamann, 1991; Bouillon and Viegas, 1999; Demonte, 1999; Picallo, 2002), a three-way classification was tested, similar to the ones established in resources such as MikroKosmos (Raskin and Nirenburg, 1995) and WordNet (Miller, 1998) for several Indoeuropean languages. These three classes can be roughly characterised as follows (terms follow Picallo (2002) and Kamp (1975) for nonpredicatives):

Qualitative adjectives (*drunk, serious, rich*)

- **syntax**: they occur as predicates in copular sentences (*my taylor is rich*) and in predicative constructions (*I saw her drunk*). In Catalan, they usually appear after the nominal head but can occasionally appear before.
- **semantics**: they are gradable and comparable: *richer, more serious, very drunk.* Gradability can be observed in Catalan (in addition to adverbial modification) in regular derivational degree and diminutive suffixes, such as *-íssim* (*petit*: *little*, *petitíssim*: *very little*).
- **denotation**: they are said to denote attributes or properties of their referents.

Relational adjectives (*thoracic, neurological*)

- **morphology**: they are usually morphologically complex, either denominal or deverbal: *thoracic*, from *thorax.*
- **syntax**: they only appear as attributes of copular verbs under constrained conditions: **the evolution of the patient is neurological*, but *the problem of the patient is neurological.* In Catalan, they cannot occur before the nominal head.
- **semantics**: they are neither gradable nor comparable.
- **denotation**: they relate the referent of the noun to an external entity (Demonte, 1999).

Nonpredicative adjectives (*mere, alleged*)

- **syntax**: in Catalan, nonpredicative adjectives only appear before the nominal head and cannot occur as predicates in copular sentences.
- **semantics**: they are nongradable and noncomparable.
- **denotation**: they do not denote properties but properties of properties, so they are nonintersective (Hamann, 1991).

In addition to the differences explained so far, there is a major syntactic argument for adopting this classification: adjectives belonging to the same class coordinate when modifying the same noun, but adjectives from different classes do not: **A serious and neurological problem* vs. *a serious neurological problem*, **A mere and rich taylor* vs. *a mere rich taylor.* When co-occuring, relational adjectives are always nearer to the nominal head than qualitative ones, and qualitative nearer than nonpredicative ones.

3 Adjective classification: motivation and challenges

Identifying adjective subclasses is useful for several tasks in NLP, at different levels of linguistic description. For example, in Catalan, if an adjective is qualitative, it can bear a regular grade morpheme (*íssim*), so this information can help broaden the morphological coverage of computational lexicons. As for syntax, coordination and adjective order can be exploited to disambiguate lexical items which are ambiguous between noun and adjective, one of the most pervasive ambiguities in POS-tagging for many Romance languages. Finally, class information could help predict and detect potential shifts in meaning.

However, it is very difficult to establish a sharp line between the classes proposed. Adjective classification, being a matter involving lexical semantics, is hampered by polysemy. The main problem in this case is class-associated polysemy: many adjectives exhibit mixed behaviour between qualitative and nonpredicative, on the one hand, or relational and qualitative, on the other.

As an example of the qualitative-nonpredicative polysemy, take an adjective such as English *poor.* It has at least two meanings associated to different classes: *poor man* means 'not rich' (qualitative reading) and 'pitiable' (nonpredicative reading). The difference can be observed when translated into Catalan: the usual translation of *poor* is *pobre*, but when used as qualitative it will follow the noun (*home pobre*) and when used as a nonpredicative

feature	mean	std dev	shallow cue
gradable	0.04	0.08	degree or diminutive morpheme, modification by degree adverb (like *molt* 'very')
comparable	0.03	0.07	modification by comparison adverb (such as *menys* 'less')
attributive	0.06	0.10	syntactic tag *Atr* ('predicate of a copular verb')
predicative	0.03	0.06	syntactic tag *Pred* ('predicative adjunct of subject or object of a noncopular verb')
non-intersective	0.04	0.08	syntactic tag *AN*> ('modifier of a noun to the right')
first adjective	0.03	0.05	first one of two or more adjectives modifying the same noun
last adjective	0.03	0.04	last one of two or more adjectives modifying the same noun

Table 1: Theoretically motivated features used in modelling adjectives for clustering, with their mean and standard deviation values (in percentages) .

it will precede it (*pobre home*). As for relational-qualitative polysemy, Catalan *econòmic* has two meanings which can be translated as *economic* (relational) and *cheap* (qualitative), so that *econòmic* can modify a noun like *pantalons* ('trousers').

Both kinds of polysemy are regular (Apresjan, 1974); moreover, we believe that all relational adjectives could be potentially used as qualitative, and all qualitative adjectives could eventually develop a nonpredicative use. Why posit different classes, then?

There are at least two answers to that question. One the one hand, not all adjectives actually have readings corresponding to both classes. For instance, it is difficult to find an appropriate context where *thoracic* can be used as qualitative.

On the other hand, as discussed in Section 2, each class presents a different set of linguistic properties. Therefore, even those adjectives which are ambiguous exhibit the properties of one single class in each particular context. For instance, as nonpredicatives cannot be used in copular sentences, *poor* is unambiguously qualitative in such contexts: *the man is poor* is not synonymous with *the man is pitiable*. Conversely, *econòmic* is gradable when used as qualitative: *pantalons molt econòmics* ('very cheap trousers'). It is therefore useful to draw a distinction between the three classes, even if a particular adjective does not necessarily belong to one single class.

4 Modelling the Data

4.1 Corpus and tools

The corpus used was a fragment of CTILC (*Corpus Textual Informatitzat de la Llengua Catalana*), collected by the Institute for Catalan Studies (IEC). The fragment contains 8.5 million words of Catalan written texts from 1970 onwards, belonging to a formal register.

The corpus had previously been automatically tagged and hand-corrected, with information on lemma, part-of-speech and other morphological information (following the EAGLES standard). It was additionally parsed with CATCG, a shallow parser for Catalan developed at GLICOM (Alsina et al., 2002). This shallow parser provides each word with a tag indicating its syntactic funcion (main verb, subject, etc.).

In order to minimise problems due to data sparseness, only adjectives occuring more than 10 times in the corpus were taken into account, totalling 3522 adjective lemmata.

4.2 Shallow cues

The features used to model the adjectives were features detectable in the annotated text itself with no external knowledge sources, because we wanted to model the linguistic behaviour of adjectives using no previous lexical knowledge. Therefore, information on derivational morphology was not used at this stage (but it was for analysis; see Section 5.3).

The values for each feature were set as true percentages, that is, the number of times a feature is detected for a given adjective, divided by the number of occurrences of that adjective in the corpus.

The lemmata were modelled using features of two kinds. On the one hand, shallow textual correlates were defined for some of the parameters discussed in Section 2. Table 1 lists these *theoretically motivated* features, together with their mean and standard deviation values, as well as the shallow cues that were defined as textual correlates of the features.

On the other hand, and in order to test the strengths and weaknesses of the starting hypothesis, a relatively unbiased description of adjetives was also provided: each lemma was described by *distributional* features, the POS of a five-word window (two at each side), as seen in Table 2.

feature	mean	std dev
word -1 Noun	0.52	0.25
word -2 Det	0.39	0.20
word +1 Prep	0.21	0.15
word +1 PT	0.42	0.15
word +2 Det	0.24	0.13
word -1 Adv	0.10	0.11
word -1 Verb	0.08	0.11
word -1 Det	0.06	0.10
word +1 Noun	0.06	0.10
word -2 Prep	0.13	0.09

Table 2: Distributional features used for describing adjectives for clustering. Of the 36 features, only the 10 with highest standard deviation are shown here.

It can be argued that some theoretically motivated features, like *comparability* or *gradability*, should not be modelled with percentages, because they do not convey relative properties, but rather absolute ones. However, by assigning these features percentages we hope to distinguish adjectives belonging to more than one class from unambiguous ones. It is reasonable to hypothesise that ambiguous adjectives will have different values in these features than unambiguous ones, which might yield differences in clustering results.

5 Experiments

5.1 Clustering Parameters

A number of clustering experiments were carried out using CLUTO (Karypis, 2002). CLUTO is a stand-alone clustering tool that provides information for analysing the characteristics of the obtained clusters, by means of a list of descriptive and discriminating features of each cluster.

The clustering parameters of CLUTO were held constant throughout the whole set of experiments, so as to focus on the effects of variations in the features used, that is, in the description of the objects. The clustering parameters used were:

- clustering algorithm: partitioning, based on the clustering criterion *H2*, a combination of cluster-internal similarity and inter-cluster dissimilarity (Zhao and Karypis, 2002)
- distance measure: cosine of the vectors
- number of clusters: 2 to 7 for each dataset, to find the most informative granularity level

5.2 Attribute Sets

The 3522 adjectives were clustered with three different subsets of the features described in Section 4.2: **distributional** features, textual correlates of **theoretically motivated** features, and **both** distributional and theoretically motivated.

After a series of experiments, the following distributional attributes were found to perturb classification and were left out of adjective description:

- *word -2 is determiner* and *word -1 is noun* present a very high correlation coefficient (0.8), which strengthens their discriminating power. Since they characterise the default adjectival context in Catalan, solutions using these features group more than half of the adjectives in one cluster. If not used, less discriminating features yield finer distinctions.
- *followed by preposition* and *word +2 is determiner* are very discriminating features, but they distinguish adjectives by their subcategorisation behaviour, which is not related to the targeted classification.
- *followed by punctuation* is also very discriminating, but the classes characterised by this feature are meaningless.

5.3 Analysis Procedure

Of the six clustering solutions obtained for the three approaches, ranging from 2 to 7 clusters, 5-cluster solutions were studied in detail, because they presented the best joint values for cluster quality (see Figure 1).

Due to the high number of objects that were clustered, the qualitative analysis of the obtained solutions was problematic. Since a comprehensive hand-made judgement of the solutions was almost impossible, two alternative strategies were adopted: a comparison of the distribution of adjective classes in clusters with a classification built by human judges and with a morphology-based classification; and a cluster-internal analysis via examination of their characterising features and the objects that prototypically represent them.

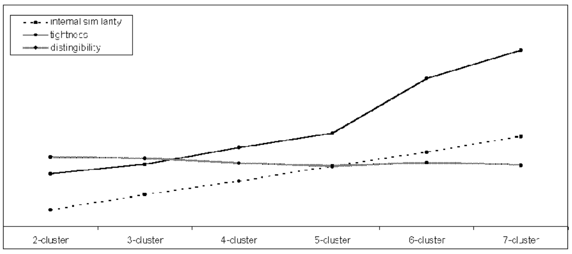

Figure 1: Quality of 2- to 7-cluster solutions with theoretically motivated features, measured by average cluster internal similarity, cluster tightness (internal similarity divided by standard deviation) and distinguishibility (internal similarity divided by external similarity).

In order to analyse the distribution of adjective classes in clusters, two a-priori classifications were built. First, 4 human judges classified a subset of 102 adjectives. 100 adjectives were randomly chosen, and two nonpredicatives (*mer* 'mere' and *presumpte* 'alleged'), were included so that this class, not found by random selection, was also represented. Judges assigned each of the adjectives to one of the hypothesised classes (Section 2): nonpredicative, qualitative or relational, or to two additional classes in order to account for polysemy: ambiguous between nonpredicative and qualitative or ambiguous between qualitative and relational.

The kappa measure for agreement between judges ranged from 0.52 to 0.64, with a confidence interval at 95% of +/-0.14. According to Landis and Koch, (1977) these figures indicate moderate (< 0.61) to substantial (> 0.61) agreement, but Carletta (1996) considers that values below 0.67 are too low to be significant for linguistic tasks. The agreement among judges is thus relatively poor, which is a clear indicator of the difficulty of the task. In spite of that, a consensus classification of the subset was established as a comparison ground for clustering solutions.

However, this evaluation only considered 2% of the objects in the dataset. To avoid low representativity, a large-scale classification was performed using derivational morphology. Some derivational suffixes yield either qualitative or relational adjectives: for instance, the denominal suffix *ós*, roughly denoting 'bearing N' and thus a kind of property, forms qualitative adjectives such as *vergonyós* ('shy', from *vergonya*, 'shyness').

Adjectives formed by derivational processes were classified according to the suffix they bear, using a list of 54 adjective forming suffixes detected by pattern matching. Morphologically simple adjectives and very ambiguous suffixes, as far as adjective classes are concerned, were not classified. No information could be gathered for nonpredicative adjectives, because there are no nonpredicative-forming suffixes.

With this procedure, 2132 adjectives (59% of the whole set) were classified. In spite of the fact that pattern matching is very error prone, and that morphological processes are not fully regular, this classification shows general tendencies for the distribution of relational and qualitative adjectives across clusters, as shown by the kappa coefficient with the manually annotated set (0.45, close to the lowest agreement between human judges, 0.52).

Figure 2 shows the distribution of adjectives across clusters in the approaches with distributional and with theoretically motivated attributes.

As a second analysis strategy, the characteristics of the clusters as such were explored, based on the list of features that CLUTO provides as most descriptive for the resulting clusters (see Section 5.1). Moreover, cluster centroids were also examined in order to obtain an example-based, human-friendly idea of the content of the clusters. An adjective was considered a centroid for a cluster when its values for descriptive features were very close to the mean value of that feature within the cluster. A summary is presented in Table 3.

6 Results and discussion

As can be seen in Figure 2, results in both approaches coincide to a large extent, with a kappa coefficient of 0.45. Kappa agreement was calculated between equivalent clusters for every pair of solutions, considering that cluster equivalence corresponded to sharing a majority of objects.

High statistical correlation coefficients were found between some theoretically motivated features and some distributional ones, as can be seen in Table 4.

This correlation is motivated by the fact that the theoretically motivated features are represented in terms of distributional properties. Remarkably, though, from the 31 distributional features used,

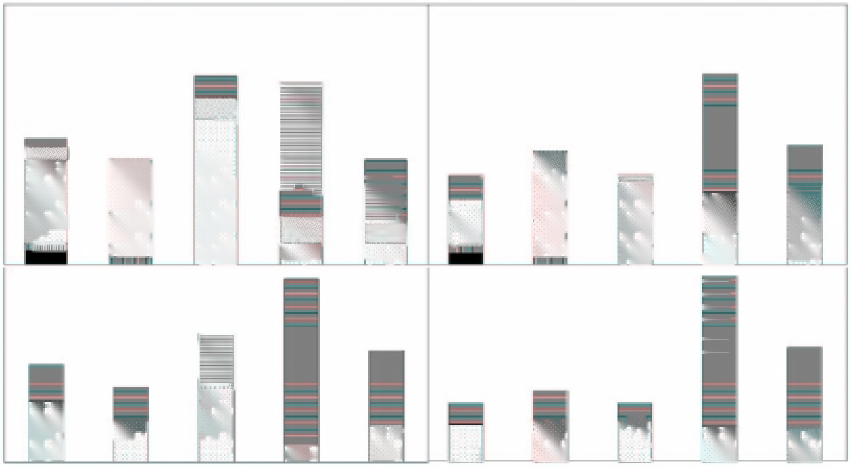

Figure 2: 5-cluster solutions with theoretically motivated (left) and distributional (right) attributes compared to classifications obtained by human judges (upper row) and using derivational morphology (lower row). Columns are clusters, patterns are classes: black is *nonpredicative*, light spots is *qualitative*, horizontal lines is *relational*, vertical lines is *ambiguous between qualitative and nonpredicative*, and tight spots is *ambiguous between qualitative and relational*. The order of clusters follows Table 3.

theoretical	distributional	corr. coeff.
non-intersective	word -1 det.	0.72
non-intersective	word +1 noun	0.88
gradable	word -1 adv.	0.67
comparable	word -1 adv	0.75
attributive	word -1 verb	0.71
predicative	word -1 verb	0.49

Table 4: Correlation coefficients between some theoretically motivated features and some distributional ones.

those ellicited as most discriminating by clustering coincide to a large extent with those that are most highly correlated with the theoretically motivated features. This provides empirical support for the theoretically motivated features chosen for the experiments.

Table 3 shows the most discriminating features for each of the clusters in the three approaches. In both approaches there are two clusters containing mainly relational adjectives and three containing mainly qualitatives. Also, the two nonpredicative adjectives are grouped together, but in a cluster also containing many qualitatives. The features used are not discriminating enough to differentiate nonpredicative adjectives from qualitative ones.

As for polysemous adjectives, they are scattered through all the clusters also in both approaches. A possible explanation for that is polysemous adjectives do not present a homogeneous behaviour: ambiguous adjectives such as *irònic* 'ironic' are used more as qualitatives than as relationals, and are clustered accordingly together with qualitatives. The reverse is true for adjectives such as *alemany* 'German'.

Three sets of clusters containing qualitative adjectives are consistently distinguished in the three solutions, which in Table 3 bear the labels **gradable**, **attributive** and **non-intersective**. Focusing on theoretical features, the first cluster is characterized by gradability and comparativity, the second one by attributivity and predicativity, and the third one by non-intersectivity. This variation seems to indicate that qualitative adjectives do not have an homogeneous behaviour.

However, it has to be taken into account that the three clusters are described in the solutions provided by CLUTO by the same descriptive features, only with different discriminating strength in the different clusters. It can be concluded, then, that qualitatives are not distinguished categorically, but gradually.

As for relationals, in the solution using theoretically motivated features they are characterised positively by positional features. In the distributional approach, however, they are defined almost negatively, as lacking all of the features that describe the other clusters. The lack of any strong, distinctive feature for these groups of adjectives can be explained by the fact that relationals are

size	most characterising features	cluster centroids		cluster label
		theoretically motivated		
624	non-intersective, first adjective	inestimable, feixuc	*inestimable, heavy*	**non-intersective**
598	gradable, comparable	agradable, recent	*pleasant, recent*	**gradable**
789	attributive, predicative	aleatori, igual	*random, equal*	**attributive**
832	first adjective, attributive	cultural, decoratiu	*cultural, decorative*	**first adj.**
678	last adjective, attributive	residual, rus	*residual, Russian*	**last adj.**
		distributional		
476	word -1 det., word +1 verb	ancià, misteriós	*ancient, mysterious*	**non-intersective**
648	word -1 adverb	brillant, dolorós	*bright, painful*	**gradable**
502	word -1 verb	raonable, cautelós	*reasonable, careful*	**attributive**
1130	word 2 verb	basc, educatiu	*Basque, educative*	**first adj.**
765	word -2 adj.	artístic, emotiu	*artistic, emotive*	-
		theoretically motivated and distributional		
505	word +1 noun, -1 det., non-intersective	misteriós, alt	*mysterious, tall*	**non-intersective**
592	word -1 adverb, comparable, gradable	dolorós, assenyat	*painful, sensible*	**gradable**
574	word -1 verb, attributive	igual, increïble	*equal, incredible*	**attributive**
1097	word -2 verb	legal, europeu	*legal, European*	**first. adj**
753	word -2 adj.	corporal, religiós	*corporal, religious*	-

Table 3: Most discriminating features in clustering solutions and some of the adjectives closest to the centroid of each cluster. The label in the last column summarizes the main characteristic of the clusters across solutions.

distributionally unmarked in Catalan (recall that the attributes representing the default context were not used; see Section 5.2).

In the solution using theoretically motivated features, cluster **first adj.** is characterised by occurring in the first position when combined with other adjectives. As suggested by the discussion in Section 2, relationals occur closer to the noun than qualitatives, so this result is consistent with the initial hypothesis.

Surprisingly, though, cluster **last adj.** also contains mainly relational adjectives, even though it has as main characteristic the fact that the adjectives occur in the *last* position when combined with other adjectives. This cluster contains many nationality-denoting adjectives, which have been classified as relational or ambiguous between relational and qualitative in the manually annotated corpus. When co-occurring with other relational adjectives, they appear further from the nominal head: *English empiricist philosopher*. This solution, thus, indicates that a finer-grained classification might be needed for relational adjectives.

Since theoretically motivated and distributional features provide comparable solutions, the combination of the two should strengthen the tendencies already noted when used separately. Indeed, when adjectives are described by the union of these two sets of features, clusters are more neatly defined, as can be seen in Figure 3. The distribution of clusters is totally comparable to the one sketched above. As for the features describing the clusters, theoretically motivated features combine with the corresponding distributional ones as expected from the discussion so far (see Table 3).

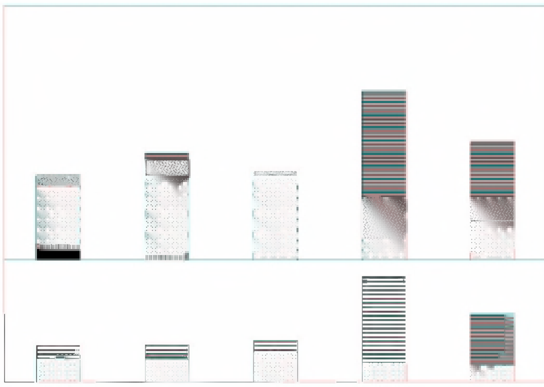

Figure 3: 5-cluster solution with distributional and theoretically motivated features, compared to classifications obtained by human judges (upper row) and using derivational morphology (lower row), following Figure 2.

7 Conclusions and Further Work

Clustering has proven to be a successful method for linguistic investigation in a relatively unexplored area such as adjectival classification. Results show that the features used in theoretical work can be successfully modelled in terms of shallow cues, so that automatic acquisition of adjective classes is possible for large-scale lexicons.

Results using distributional features parallel to a large extent with results using theoretically motivated features, which provides empirical evidence that the properties mentioned in Section 2 are relevant for adjective classification. Moreover, clustering results largely support the initial hypothesis, as qualitative adjectives are distinguished from relational ones and nonpredicative adjectives are grouped together.

Nevertheless, this approach is not useful for detecting and acquiring data on class-associated polysemy, probably due to heterogeneity in the behaviour of polysemous adjectives. Several alternatives will be explored in the near future: soft clustering can be used to test whether polysemous adjectives fall into several clusters; also, a bootstrapping approach has been envisaged that exploits information on coordination and adjective order.

Acknowledgments

Part of this material was presented at the Workshop on Quantitative Investigations in Theoretical Linguistics (Osnabrück, 3-5 October 2002). The authors wish to thank the reviewers and audience of the workshop for their helpful comments. They also thank the reviewers of the EACL03 student session, as well as Toni Badia, Louise McNally, Gretel DeCuyper and Martí Quixal for their detailed criticism of the paper. Special thanks are due to Àngel Gil, Martí Quixal and Clara Soler for manual annotation of the data.

This research has been conducted thanks to grants 2001FI-00582 from the government of Catalonia (Gemma Boleda) and PB98-1226 from the X-TRACT project, of the Spanish Ministry of Science and Technology (Laura Alonso). It has also been partially funded by projects HERMES (TIC2000-0335-C03-02) and PETRA (TIC2000-1735-C02-02), by CLiC (Centre de Llenguatge i Computació).

References

À. Alsina, T. Badia, G. Boleda, S. Bott, À. Gil, M. Quixal, and O. Valentín. 2002. CATCG: a general purpose parsing tool applied. In *Proceedings of Third International Conference on Language Resources and Evaluation.*

I. D. Apresjan. 1974. Regular polysemy. *Linguistics*, 142:5–32.

B. Bohnet, S. Klatt, and L. Wanner. 2002. An approach to automatic annotation of functional information to adjectives with an application to German. In *Proceedings of the 3rd LREC Conference, Workshop: Linguistic Knowledge Acquisition and Representation.*

P. Bouillon and E. Viegas. 1999. The Description of Adjectives for Natural Language Processing: Theoretical and applied perspectives. In *Proceedings of Description des Adjectifs pour les Traitements Informatiques. Traitement Automatique des Langues Naturelles*, Cargèse, Corsica.

J. Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

V. Demonte, 1999. *El adjetivo: clases y usos. La posición del adjetivo en el sintagma nominal*, pages 129–215. Espasa-Calpe, Madrid.

C. Hamann, 1991. *Adjectivsemantik / Adjectival Semantics*, pages 657–673. De Gruyter, Berlin/NY.

V. Hatzivassiloglou and K. R. McKeown. 1993. Towards the automatic identification of adjectival scales: Clustering adjectives according to meaning. In *Proceedings of the 31st Annual Meeting of the ACL*, pages 172–182.

S. Schulte im Walde and C. Brew. 2002. Inducing German semantic verb classes from purely syntactic subcategorisation information. In *Proceedings of the 40th Annual Meeting of the ACL*, pages 223–230.

J. A. W. Kamp. 1975. Two theories about adjectives. In E. L. Keenan, editor, *Formal Semantics of Natural Language*, pages 123–155. Cambridge University Press, Cambridge.

G. Karypis. CLUTO 2002.
http://www-users.cs.umn.edu/ karypis/cluto/index.html

J. R. Landis and G. C. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*.

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

K. J. Miller. 1998. Modifiers in WordNet. In Christiane Fellbaum, editor, *WordNet: an electronic lexical database*, pages 47–67. MIT, London.

F. C. N. Pereira, N. Tishby, and L. Lee. 1993. Distributional clustering of English words. In *Meeting of the Association for Computational Linguistics*, pages 183–190.

C. Picallo. 2002. L'adjectiu i el sintagma adjectival. In Joan Solà, editor, *Gramàtica del català contemporani*, pages 1643–1688. Empúries, Barcelona.

V. Raskin and S. Nirenburg. 1995. Lexical semantics of adjectives: A microtheory of adjectival meaning. Technical report, New Mexico State University.

Y. Zhao and G. Karypis. 2002. Evaluation of hierarchical clustering algorithms for document datasets. Technical Report 02-022, Department of Computer Science & Engineering, University of Minnesota.