

Morphology vs. Syntax in Adjective Class Acquisition

Gemma Boleda GLiCom Pompeu Fabra University Barcelona gemma.boleda@upf.edu	Toni Badia GLiCom Pompeu Fabra University Barcelona toni.badia@upf.edu	Sabine Schulte im Walde Computational Linguistics Saarland University Saarbrücken schulthe@CoLi.Uni-SB.DE
-----------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------

Abstract

This paper discusses the role of morphological and syntactic information in the automatic acquisition of semantic classes for Catalan adjectives, using decision trees as a tool for exploratory data analysis. We show that a simple mapping from the derivational type to the semantic class achieves 70.1% accuracy; syntactic function reaches a slightly higher accuracy of 73.5%. Although the accuracy scores are quite similar with the two resulting classifications, the kinds of mistakes are qualitatively very different. Morphology can be used as a baseline classification, and syntax can be used as a clue when there are mismatches between morphology and semantics.

1 Introduction

This paper fits into a broader effort addressing the automatic acquisition of semantic classes for Catalan adjectives. So far, no established standard of such semantic classes is available in theoretical or empirical linguistic research. Our aim is to reach a classification that is empirically adequate and theoretically sound, and we use computational techniques as a means to explore large amounts of data which would be impossible to explore by hand to help us define and characterise the classification.

In previous research (Boleda et al., 2004), we developed a three-way classification according to generally accepted adjective properties (see Section 2),

and applied a cluster analysis to further examine the classes. While the cluster analysis confirmed our classification to a large extent, it was clear that one of the classes needed further exploration. Also, we used only syntactic features modelled as pairs of POS-bigrams; we explored neither other syntactic features nor the role of morphological evidence for the classification.

In this paper we apply a supervised classification technique, decision trees, for exploratory data analysis. Our aim is to explore the linguistic features and description levels that are relevant for the semantic classification, focusing on morphology and syntax. We check how far we get with morphological information, and whether syntax is helpful to overcome the ceiling reached with morphology.

Decision trees are appropriate for our task, to test and compare sets of features, based on our gold standard. They are also known for their easy interpretation, by reading feature combinations off the tree paths. This property will help us get insight into relevant characteristics of our adjective classes, and in the error analysis.

The paper is structured as follows: Section 2 presents the adjective classification and the gold standard used for the experiments. Sections 3 and 4 explore the morphology-semantics interface and the syntax-semantics interface with respect to the classification proposed, and Section 5 focuses on the differences in the kind of information each level provides for the classification. Sections 6 and 7 are devoted to discussion of related work and conclusions.

2 Classification and gold standard

2.1 Classification proposal

To date, no semantic classification of adjectives is generally accepted in theoretical linguistics. Much research in formal semantics has focused on entailment properties, while other kinds of lexical semantic properties are left uncovered. Standard descriptive grammars propose broader classifications (see Picallo (2002) for Catalan), but these usually do not follow a single classification parameter: they mix morphological, syntactic and semantic criteria and end up with classifications that are not consistent.

We are interested in properties of the lexical semantics of adjectives that have a bearing on their syntactic behaviour. This property makes the semantic distinctions traceable at another linguistic level, which is desirable to ensure falsability of the classification criteria. On more practical terms, it also allows the exploitation of the syntax-semantics interface as is usual in Lexical Acquisition, to automate the acquisition of the relevant classes.

Our proposal is largely inspired by the Ontological Semantics framework (Raskin and Nirenburg, 1995). The assumption of an ontology as a model of the world allows us to distinguish linguistic aspects (e.g. syntactic properties) from the actual content of the lexical entries, formalised as a link to an element of the ontology. We assume an ontology of basic denotations composed of properties (or attributes), objects (or entities), and events. Adjectives participate in each of these possible denotations, and can be *basic*, *object-related* or *event-related*, depending on their lexical meaning.¹ We next characterise each class.

Basic adjectives are the prototypical adjectives, which denote attributes or properties which cannot be decomposed (*bonic* ‘beautiful’, *sòlid* ‘solid’). Event adjectives have an event component in their meaning. For instance, if something is *tangible* (‘tangible’), then it can be touched: *tangible* necessarily evokes a potential event of touching which is embedded in the meaning of the adjective. Other examples are *alterat* (‘altered’) and *ofensiu* (‘offensive’). Similarly, object adjectives have an embed-

ded object component in their meaning: *deformació nasal* (‘nasal deformity’) can be paraphrased as *deformity that affects the nose*, so that *nasal* evokes the object *nose*. Other examples are *peninsular* (‘peninsular’) and *sociolingüístic* (‘sociolinguistic’).

This proposal shares many aspects with discussions in descriptive grammar (2002) and proposals in other lexical resources, such as WordNet (Miller, 1998). In particular, the distinction between prototypical, attribute-denoting adjectives and object-related adjectives is found both in descriptive grammar and in WordNet. As for event-related adjectives, they are not usually found as a class in Romance descriptive grammar, and in WordNet they are distinguished but only if they are participial; other kinds of deverbal adjectives are considered basic (in our terminology). More on the morphology-semantics relationship in Section 3.

Our classification focuses on the semantic content of adjectives, rather than on formal properties such as entailment patterns (contrary to the tradition in formal semantics). The semantic distinctions proposed have an effect on the syntactic distribution of adjectives, as will be shown throughout the paper, and can be exploited in low-level NLP tasks (POS-tagging), and also in more demanding tasks, such as paraphrase detection and generation (e.g. exploiting the relationship *tangible* → *can be touched*, or *deformació nasal* → *deformity affecting the nose*).

2.2 Gold standard

To perform the experiments, we built a set of annotated data based on this classification (*gold standard* from now on). We extracted the lemmata and data for the gold standard from a 16.5 million word Catalan corpus (Rafel, 1994), lemmatised, POS-tagged and shallow parsed with the CatCG tool (Alsina et al., 2002). The shallow parser gives information on the syntactic function of each word (subject, object, etc.), not on phrase structure.

186 lemmata were randomly chosen among all 2564 adjectives occurring more than 25 times in the corpus. 86 of the 186 lemmata were classified by 3 human judges into each of the classes (basic, object, event).² In case of polysemy affecting the class as-

¹Raskin and Nirenburg (1995) account separately for other kinds of adjectives, such as membership adjectives (‘fake’). We will abstract away from these less numerous classes.

²The 3 human judges were PhD students with training in linguistics, one of which had done research on adjectives. As it was defined, the level of training in linguistics needed for the

signment, the judges were instructed to return the class for the most frequent sense as the primary class, and a secondary class for the other sense.

Polysemy typically arises in cases where an adjective has developed a *noncompositional* sense. One of these cases would be the adjective *puntual*, a denominal adjective (derived from *punt*, ‘point’). Its most frequent sense is ‘punctual’ as in ‘I expect Mary to be punctual for this meeting’. This is a basic meaning, noncompositional in the sense that it cannot be predicted from the meaning of the originating noun in combination with the suffix.

The adjective has a compositional sense, namely, ‘related to a point’ (usually, a point in time), as in *això va ser un esdeveniment puntual*, ‘this was a once-occurring event’. This is the meaning we would expect from the derivation *punt* (‘point’) + *al*, and is an object meaning. In this case, the judge should assign the adjective to two classes, primary basic, secondary object. Compositional meanings are thus those corresponding to active morphological processes, and can be predicted from the meaning of the noun and the derivation with the suffix (be it denominal, deverbal or participial).

The judges had an acceptable 0.74 mean κ agreement (Carletta, 1996) for the assignment of the primary class, but a meaningless 0.21 for the secondary class (they did not even agree on which lemmata were polysemous). As a reaction to the low agreement about polysemy, we incorporated polysemy information from a Catalan dictionary (DLC, 1993). This information was incorporated only in addition to the gathered gold standard: In many cases the dictionary only lists the compositional sense. We added it as a second reading if our judges considered the noncompositional one as most frequent.

One of the authors of the paper classified the remaining 100 lemmata according to the same criteria. For our experiment, we use the complete gold standard containing 186 lemmata (87 basic, 46 event, and 53 object adjectives).

3 Morphological evidence

There is an obvious relationship between the derivational type of an adjective (whether it is denominal, deverbal, or not derived) and the semantic class task was quite high.

sification we have put forth: Usually, a denominal adjective has an object embedded in its meaning (corresponding to the object denoted by the noun from which it is derived). Similarly, a deverbal or participial adjective tends to denote a relationship with an event (the event denoted by the originating verb), and a nonderived adjective tends to have a basic meaning. Therefore, the simplest classification strategy is to associate each derivational type with a semantic class: nonderived \rightarrow basic, participial \rightarrow event, deverbal \rightarrow event, and denominal \rightarrow object.

Table 1 reflects the accuracy results of this theoretically defined mapping between morphology and semantics, compared to our gold standard (cases corresponding to the predicted mapping in boldface).³ For instance, the first line of this table shows that 39 of the 42 nonderived adjectives, predicted to be basic by the morphology-semantics mapping, are actually deemed basic by the human judges, while the remaining 3 are classified as object adjectives.

	basic	event	object	Total
nonderived (basic)	39	0	3	42
deverbal (event)	12	11	2	25
participial (event)	12	35	0	47
denominal (object)	24	0	48	72
<i>Total</i>	87	46	53	186
precision	.93	.64	.67	.74
recall	.45	1	.91	.78
f-score ($\alpha = 0.5$)	.69	.82	.79	.76

Table 1: Morphology-semantics mapping: results

Note that the table correctly reflects the general tendencies just outlined: This simple classification achieves 0.76 f-score. However, there are obvious mismatches. Most of these mismatches are concentrated in the first column, namely many of the deverbal, participial and denominal adjectives (predicted to denote event or object meanings) actually have a basic meaning as their most frequent sense. This fact is reflected in the low recall score for basic adjectives (0.45), and in precision being much lower than recall for the other two classes (0.64 vs. 1 for event, 0.67 vs. 0.91 for object adjectives).

³The morphological information was obtained from a manually constructed electronic database of adjectives, kindly provided by Roser Sanromà (2003).

The mismatches usually correspond to polysemy due to noncompositional senses of the adjectives, such as the denominal adjective *puntual* discussed above. Another case is the participial *abatut*, which compositionally means ‘shot-down’, but is most frequently used as a synonym to ‘depressed, down-cast’, and therefore is classified as basic. Similarly, a deverbal adjective such as *radiant* most frequently means ‘happy’, but also has a compositional sense, ‘irradiating’.

Sometimes the compositional meaning is completely lost, as with most deverbal adjectives classified as basic. In some cases the underlying verb no longer exists in Catalan (*horrible*-**horrir*, *compatible*-**compatir*), and they are not perceived as derived.⁴ In other cases, although the verb exists, it is a stative predicate (e.g. *inestable*, ‘unstable’, from *estar* ‘stand/be’; *pudent* ‘stinking’, from *pudir*, ‘stink’), and thus are much more similar to basic adjectives than deverbal adjectives deriving from dynamic predicates, such as *ofensiu* (‘offensive’). Aspectuality of the deriving verb is a factor that has to be examined more carefully in the future.

To summarise, the results for the morphology-semantics mapping indicate that there is a clear relationship between these two levels: Morphology does most of the job right, because each morphological rule has an associated semantic operation. However, this level of information has a clear performance ceiling. In case of noncompositional meanings the morphological class will systematically be misleading, which cannot be overcome unless other kinds of information are let into play.

4 Syntactic evidence

If we adhere to the hypothesis that semantics has a reflection in syntactic distribution (basis for most work in Lexical Acquisition), we can expect that syntax gives us a better clue to semantics than morphology, particularly in cases of noncompositional meanings. We expect that adjectives with a noncom-

⁴The question may arise of whether these adjectives are really deverbal. In the current version of the adjective database, all adjectives bearing a suffix that is active in the Catalan derivational system are classified as derived. The problem is that Catalan shares suffixes with Latin, so that fixed forms from Latin that have been incorporated into Catalan cannot be superficially distinguished from active derived forms.

positional meaning behave in the syntax as basic adjectives, not as event or object adjectives.

Before getting into the experiments using syntactic information, we briefly present the syntax of adjectives in Catalan and the predictions with respect to the syntactic behaviour of each class.

4.1 Adjective syntax in Catalan

The default function of the adjective in Catalan is that of modifying a noun; the default position is the postnominal one (about 66% of adjective tokens in the corpus modify nouns postnominally). Examples are *taula gran* (‘big table’), *arquitecte tècnic* (‘technical architect’), and *element constitutiu* (‘constitutive element’).

However, some adjectives can appear prenominal, mainly when used non-restrictively (so-called “epithets”; 26% of the tokens occur in prenominal position). In English, this epithetic use is not typically distinguished by position, but some adjectives can epithetically modify proper nouns (‘big John’ vs. ‘*technical John’). ‘Big’ in ‘big John’ does not restrict the reference of ‘John’, but highlights a property. In Catalan and other Romance languages, prenominal position is systematically associated to this use, with proper or common nouns.

The other main function of the adjective is that of predicate in a copular sentence (6% of the tokens), such as *aquesta taula és gran* (‘this table is big’). Other predicative contexts, such as adjunct predicates (as in *la vaig veure borratxa*, ‘I saw her drunk’), are much less frequent: approx. 1% of the adjectives in the corpus.

From empirical exploration and literature review, we gathered the following tentative predictions as to the syntactic behaviour of each class in Catalan:

Basic adjectives occur in predicative environments, have scope over other adjectives modifying the same head (most notably, object adjectives), and can have epithetic uses and therefore occur prenominaly.

Event adjectives occur in predicative environments and after object adjectives.

Object adjectives occur in a rigid position, directly after their head noun; they do not allow pred-

icative constructions nor epithetic uses (therefore not prenominal position).

4.2 Setup

We modelled the syntactic behaviour of adjectives using three different representation strategies. The values in the three cases were frequency counts, that is, the percentage of occurrence of each adjective in that syntactic environment. The frequency of the adjectives from the gold standard in the corpus ranges from 27 to 7154 (median: 129.5). All in all, 56,692 out of the approx 600,000 sentences in the corpus were used as data for this experiment. We have not analysed the influence of frequency on the results, but each adjective is represented by a reasonable amount of data, so that the representation of the syntactic evidence in terms of frequency is adequate.

The simplest modelling strategy is unigram representation, taking the POS of the word to the left of the adjective and the POS of the word to the right as separate features. Adjectives have a limited syntactic distribution (much more restricted than e.g. verbs), so that even this simple representation should provide relevant evidence. The second one is bigram representation, with features consisting of the POS of the word to the left of the adjective and the POS of the word to the right as a single feature. This representation results in a much larger number of features (see Table 2), thus potentially leading to data sparseness, but it should be more informative, because left and right context are taken into account at the same time.

The third one is the syntactic function, as given by CatCG. For adjectives, these functions are noun modifier (distinguishing between prenominal and postnominal position), predicate in a copular sentence, and predicative adjunct (more information in Section 4.4). CatCG does not yield completely disambiguated output, and the ambiguous functions were also taken into account, so as not to miss any potentially relevant source of evidence.

To perform the experiment, we used C5.0, a commercial decision tree and rule induction engine developed by Ross Quinlan (Quinlan, 1993). We tried several options, including the default, winnowing, and adaptive boosting. Although the results varied a bit within each representation strategy (boosting tended to perform better, winnowing did not have

a homogeneous behaviour), the general picture remained the same as to the relative performance of each level of representation. Therefore, and for clarity of exposure and exploration reasons, we will only present and discuss results using the default options.

For comparison, we ran the tool on the 3 syntactic representation levels and on morphological information, using derivational type, a finer-grained derivational type, and the suffix.⁵

4.3 Results

The results of the experiment, obtained averaging ten 10-fold cross-validation runs, are depicted in Table 2. In this table, *#f* is the number of features for each representation strategy, *size* the size of the trees (number of leaves), *accuracy* the accuracy rate of the classifiers (in percentage), and *SE* the standard error of each parameter. We currently assume a majority baseline, that of assigning all adjectives to the most numerous class (basic). Given that there are 87 basic adjectives and 186 items in the gold standard (see Table 1), this baseline results in 46.8% accuracy.

	#f	size		accuracy	
		mean	SE	mean	SE
baseline	-	-	-	46.8	-
morphology	3	4.3	0.1	70.1	0.3
unigram	24	19.1	0.2	68.8	0.6
bigram	135	18.8	0.4	67.4	0.8
synt. funct.	14	3.5	0.1	73.8	0.3

Table 2: Decision Tree experiment

Note that all four classifiers are well above the majority baseline (46.8%). The best results are obtained with the lowest number of features (3 for morphology, 14 for syntactic function, vs. 24 and 135 for unigram and bigram), and correspondingly, with the smallest trees (average 4.3 and 3.5 leaves for morphology and function, 19.1 and 18.8 for n-grams). We interpret this result as indicating that the levels of description of morphology and syntactic function are more adequate than the n-gram representation, although this is only a tentative conclusion, because the differences in accuracy are not large. Function abstracts away from particular POS

⁵The finer-grained derivational type states whether the adjective is derived from a noun or verb that still exists in Catalan or not.

syntactic function	basic		event		object	
postnominal modifier	.69	+/- .16	.68	+/- .19	.94	+/- .06
prenominal modifier	.07	+/- .09	.02	+/- .04	.01	+/- .03
predicative adjunct	.09	+/- .08	.19	+/- .16	.02	+/- .03
predicate in a copular sentence	.10	+/- .10	.08	+/- .07	.01	+/- .02

Table 3: Average values for the syntactic functions in each adjective class.

environments, and summarises the most relevant information without the data sparseness problems inherent in n-gram representation.

Also noteworthy is that the accuracy rates for syntax are lower than we would have expected, according to the hypothesis that it better reflects synchronic meaning. For the first two syntactic representations, unigrams and bigrams, results are worse than using the simple morphological mapping explained above (respectively 68.8% and 67.4% accuracy, compared to 70.1% accuracy achieved with morphology).⁶ Only syntactic function improves upon the morphological results, and only slightly (73.8% average accuracy). However, as will be explored in the rest of the Section, the mistakes of the morphological classifier are qualitatively different from those of the syntactic classifiers, which can be used to gain insight into the nature of the problem handled, and to build better classifiers.

4.4 Error analysis

For the analysis of the results, we will focus on the syntactic function features, because it is the best system and allows clearer exploration of the hypotheses stated so far than the n-gram representation.

Table 3 contains the data for the 4 main syntactic functions for adjectives. For each class (all adjectives classified as basic, event or object in the gold standard), it contains the average percentage of occurrence with each syntactic function, along with the standard deviation. A set of 10 remaining syntactic features represented cases not disambiguated by CatCG, which had really low mean values and were rarely used in the DTs.

The values of the 4 syntactic functions confirm to a large extent the predictions made with respect to the syntactic behaviour of each adjective class, but

⁶When using morphological features, DTs used almost only the main derivational type, according to the hypothesis stated in Section 3.

also evidence an additional fact: basic and event adjectives, in the current definition of the classes, have only slight differences in their syntax.

Basic and event adjectives have similar mean values for the default adjective position in Catalan (postnominal modifier; 0.69 and 0.68 mean values), and also for the predicative function in a copular sentence (0.10 and 0.084 mean values). The two-sample t-test confirms that the differences in mean are not significant ($p=0.73$ and $p=0.88$ at the 95% confidence interval).⁷

Basic adjectives occur more frequently as prenominal modifiers (0.07 compared to 0.02), but note the large standard deviation (0.09 and 0.04), which means that there is a large within-class variability. In addition, event adjectives have a larger mean value for the predicative adjunct function (0.19 vs. 0.09), but again, the standard deviation of both classes is very large (0.16 and 0.08). Nevertheless, a t-test returns significant p values (< 0.001 , 95% conf. int.) for the differences in mean of these two features, so that they can be used as a clue to the characterisation of the event class.⁸ The bias of event adjectives towards predicative uses can be attributed to participials – the most frequent kind of adjectives in the event class (35 vs. 11).

Object adjectives do present a distinct syntactic behaviour: They act (as expected) as rigid postnominal modifiers (mean value 0.94), and cannot be used as prenominal modifiers (mean value 0.01) or as predicates (mean values 0.018 and 0.008 for predicative functions). Also note that the standard deviation for each feature is lower in the case of object adjectives than in the case of basic and event adjectives, which indicates a higher homogeneity of the object class. T-tests for the difference in means with

⁷Alternatives “not equal” and “basic smaller than event” respectively.

⁸Alternatives: “basic greater than event” for prenominal modification, “event greater than basic” for predicative adjunct.

respect to the basic and event class return significant p values (< 0.001) except for the difference in prenominal modification values between event and object adjectives ($p=0.26$).⁹

Decision trees built with this feature set use the information consistent with the observations just outlined. In general, they characterise object adjectives as postnominal modifiers (usual threshold: 0.9), basic adjectives as prenominal modifiers (usual threshold: 0.01), and event adjectives as not being prenominal modifiers. In some trees, information about predicativity is also included (event adjectives act as predicative adjuncts; usual threshold: 0.04).

From the discussion of the feature values, it is to be expected that most of the mistakes when using the syntactic function feature set are due to basic-event confusion, and this is indeed the case. For the error analysis, we divided the gold standard into three equal sets, and successively trained on two sets and classified the third. The classification of the gold standard that resulted is reflected in Table 4 (correctly classified items in boldface).

true class →	basic	event	object	Total
basic	56	7	5	68
event	18	35	4	57
object	13	4	44	61
Total	87	46	53	186
precision	.82	.61	.72	.72
recall	.64	.76	.83	.69
f-score	.73	.69	.78	.73

Table 4: Syntax-semantics mapping: results

Table 4 shows that the object class is best characterised (0.78 f-score), followed by the basic (0.73) and event (0.69) classes. Particularly low are precision for event (0.61) and recall for basic (0.64) adjectives. This distribution indicates that many adjectives are classified as event while belonging to other classes (18 to basic, 4 to object), and many basic adjectives are classified into other classes (18 as event, 13 as object).

The basic-event confusion mainly takes place with basic adjectives not used as epithets (in

⁹Alternatives: all means of basic and event greater than those of object, except for postnominal modification, testing against a greater mean for object.

prenominal position; *curull* ‘full’, *dispers* ‘scattered’) and event adjectives used as epithets (*interminable* ‘endless’, *ofensiu* ‘offensive’). Although more analysis is needed, in many of these cases (such as *interminable*) the underlying verb is stative, which makes the adjectives very similar to basic adjectives, as mentioned in Section 3. The judges reported difficulties particularly in distinguishing event from basic adjectives, which matches the results of the experiments. The classification is fuzzy in this point, and we intend to develop clearer criteria to distinguish adjectives with an “active” event in their lexical meaning from basic adjectives.

As for the basic-object confusion, it is due to two factors. The first one is basic being the default class: In the gold standard, if an adjective does not fit into the other 2 classes, it is considered basic, even if it does not denote a prototypical kind of attribute or property. Examples are *radioactiu* (‘radioactive’) and *recíproc* ‘reciprocal’. These tend to be used less in predicative and epithetic functions.

The second one is polysemy. 4 adjectives classified in the gold standard as polysemous between a basic (primary) and an object (secondary) reading are classified by C5.0 as object because they almost only ($> 90\%$ of the time) occur postnominally: *artesanal*, *mecànic*, *moral*, *ornamental* (‘artisanal, mechanical, moral, ornamental’). All of these cases have a compositional meaning paraphrasable by ‘related-to X’, where X is the derived noun, and a noncompositional meaning such as ‘automatic’ for *mecànic*. The syntactic behaviour of the adjective is mixed according to the two classes, so that the values for environments typical of basic adjectives are too low to meet the thresholds.¹⁰

To sum up, event adjectives do not seem to have consistent syntactic characteristics that tell them apart from basic adjectives, while object adjectives have a consistent behaviour distinct from the other two classes. This result backs up previous experimentation with clustering (Boleda et al., 2004), where half of the event adjectives were systematically clustered together with basic adjectives.¹¹ Pol-

¹⁰Note, however, that in 6 other cases with the same polysemy, syntax does tell them apart from typical object adjectives, and are classified as basic (such as the *puntual* case discussed above; see discussion in next Section).

¹¹The ones that were distinguished from basic adjectives

ysemy plays a tricky role, because depending on the uses of the adjective it leads to a continuum in the feature values which sometimes does not allow a clear identification of the most frequent sense.

5 Differences between morphology and syntax

A crucial point to understand the roles of morphology and syntax for our semantic classification is the differences in the kinds of mistakes that each of the information level carries with it. From the discussion up to this point, we would expect that the default morphological classification causes less mistakes with event vs. basic, because the deverbal morphological rules carry the associated “related-to-event” meaning. On the contrary, syntax should handle better the cases where the relationship between morphology and semantics is lost, what we have termed noncompositional meanings.

If we compare the mistakes made by each mapping, both morphology and syntax assign the expected class to 103 lemmata (55.4% of the gold standard), and both coincide in assigning a wrong class for 21 (11.3%). The cases where one mapping achieves the right classification and the other one makes a mistake are reflected in Tables 5 and 6.

true class →	basic	event	object	Total
basic		7	5	12
event	6		4	10
object	4	4		8
Total	10	11	9	

Table 5: Morphology right, syntax wrong

true class →	basic	event	object	Total
basic		2	3	3
event	10			12
object	17			17
Total	27	2	3	

Table 6: Syntax right, morphology wrong

Cases where morphology achieves the right class and syntax does not (Table 6) do not present a very clear pattern, although the basic-event confusion in

was so due to their bearing complements, a parameter orthogonal to the targeted classification.

syntax is indeed reflected as the most numerous in Table 5 (6+7 cases). In absence of a syntactic characterisation of the class, applying the default mapping will yield better results.

As for the cases where syntax classifies correctly and morphology does not (Table 6), they do present a clear pattern: They correspond, as expected, to deverbal (8), participial (2) and denominal (17) adjectives with a meaning that does not correspond to the morphological rule. Among denominals, examples are *elemental* and *horrorós* (‘elementary’ and ‘horrifying’); among deverbals, *reasonable* and *present* (‘reasonable’ and ‘present’); among participials, *innat* and *inesperat* (‘innate’ and ‘unexpected’).

Note that syntax is most helpful in the identification of basic denominal adjectives (17 cases), providing support for the hypothesis that adjectives with a noncompositional meaning behave in the syntax as basic adjectives, which can be exploited in a lexical acquisition setting. In contrast, event and basic classes not having a clearly distinct syntactic distribution, the syntactic features do not help in telling these two classes apart. This problem accounts for the little overall accuracy improvement from morphology (70.1%) to syntax (73.8%): It improves the object vs. basic distinction, but it does not consistently improve the event vs. basic distinction.

5.1 Combining morphological and syntactic features

The next logical step in building a better classifier for adjectives is to use both morphological and syntactic function information. When doing that, a slightly better result is obtained, although no dramatic jump in improvement: 74.7% mean accuracy averaged across ten 10-fold cross-validation runs, with trees of average 8 leaves (mean accuracy being 70.1% with morphology and 73.8% with syntactic function; see Table 2).

In most of the partitions of the data when using this feature set, the first node uses syntactic evidence (high values for postnominal position for object adjectives vs. the rest), and the second level nodes use the derivational type. The remaining morphological features (suffix, fine-grained derivational type; see footnote 4.2) are seldom used.

In all the decision trees, nonderived adjectives are directly assigned to the basic class, and in 80% par-

ticipial adjectives are classified as event. The last rule causes a large number of errors, because 12 out of 47 participles were classified as basic in the gold standard. For the other two derivational types, syntactic evidence is used again in almost all decision trees (99% for deverbal, 80% for denominal adjectives). Deverbal or denominal adjectives that occur preminally are deemed basic, according to expectation. Contrary to expectation, however, deverbal adjectives that occur predicatively are classified as basic. This result confirms the suspicion that frequent predicative use is associated with participial, but not with other kinds of deverbal adjectives, as stated in Section 4.4.

6 Related work

In recent years much research (Merlo and Stevenson, 2001; Schulte im Walde and Brew, 2002; Korhonen et al., 2003) has aimed at exploiting the syntax-semantics interface for classification tasks, mostly based on verbs. In particular, Merlo and Stevenson (2001) present a classification experiment which bears similarities to ours. They use decision trees to classify intransitive English verbs into three semantic classes: unergatives, unaccusatives, and object-drop. As in our experiments, they define three classes, and use only 60 verbs for the experiments. Merlo and Stevenson identify linguistic features referring to verb argument structure (crucially involving thematic relations), and classify the verbs into the three classes with an accuracy of 69.8%. They compare their results with a random baseline of 33%.

There has been much less research in Lexical Acquisition for adjectives. Early efforts include Hatzivassiloglou and McKeown (1993), a cluster analysis directed to the automatic identification of adjectives belonging to the same scale (such as *cold-tempered-hot*). More recently, Bohnet et al. (2002) used bootstrapping to assign German adjectives to “functional” classes (of a more traditional sort, based on a German descriptive grammar). They relied on ordering restrictions and coordination data which can be adapted to Catalan.

As for Romance languages, the only related work we are aware of is Carvalho and Ranchod (2003), who developed a finite-state approach to disam-

biguating homograph adjectives and nouns in Portuguese. They manually classified the adjectival uses of the homographs into six syntactic classes with characteristics used in our classification (predicative uses, position with respect to the head noun, etc.). They used that information to build finite state transducers aimed at determining the POS of the homographs in each context, with a high accuracy (99.3%) and coverage (94%). The research undergone in this paper leads to the automatic acquisition of the classes, defined however at a semantic rather than syntactic level.

7 Conclusion and future work

In this paper, we have presented and discussed the role of two sources of evidence for the automatic classification of adjectives into ontological semantic classes: morphology and syntax. Both levels provide relevant information, as indicated by their respective accuracy results (70.1% for morphology, 73.8% for syntax), both well above a majority baseline (46.8%). Morphology fails in cases of noncompositional meaning, when the relationship to the deriving word has been lost, cases that syntax tends to correctly classify. In contrast, syntax systematically confuses event and basic adjectives due to the lack of a sufficiently distinct syntactic profile of the event class. Therefore, the default morphology-semantics mapping handles these cases better.

Not surprisingly, the best classifier is obtained combining both kinds of information (74.7%), although it is not even 1% better than the syntactic classifier. More research is needed to achieve better ways of combining both levels of description.

We can summarise our results as indicating that morphology can give a reliable initial hypothesis with respect to the semantic class of an adjective, which syntax can refine in cases of noncompositional meaning, particularly for object adjectives. Therefore, morphology can be used as a baseline in future classification experiments.

The experiments presented in this paper also shed light on the characteristics of each class. In particular, we have shown that event adjectives do not have a homogeneous and distinct syntactic profile. One factor to take into account is that the morphological variability within the class (suffixes *-ble*, *iu*, *nt*,

participles) is associated with a high semantic variability. This semantic variability is not found in the object class, where the several suffixes (*al, ic, à, etc.*) all have a similar semantic effect. Another factor which seems to play a role, and which has been identified in the error analysis, is the aspectuality of the deriving verb, particularly whether it is stative or dynamic. In the near future, we intend to use the best classifier to automatically classify more adjectives of our database, so as to allow further exploration of the data and a clearer definition of the class.

A major issue we leave for future research is polysemy detection. Up to now, we have only aimed at single-class classification, and not attempted to capture multiple uses of an adjective. E.g. the approach in Bohnet et al. (2002) could be adapted to Catalan: We can use data on coordination and ordering for polysemy detection, once the class of the most frequent sense is established with the methodology explained in this paper.

Finally, the results presented in this paper seem to point in a fruitful direction for the study of adjective semantics: Adjectives that are flexibly used, those that fully exploit the syntactic possibilities of the language (in Catalan, being used predicatively and as epithets), tend to correspond to adjectives with a basic meaning, that is, tend to be viewed as a compact attribute, as a prototypical adjective. In contrast, derived adjectives which retain much of the semantic link to the noun or verb from which they derive do not behave like prototypical adjectives, are tied to certain positions, and do not exhibit the full range of syntactic possibilities of adjectives as a class. We intend to explore the consequences of this hypothesis in more detail in the future.

Acknowledgements

Many thanks to the people who have manually annotated the data: Àngel Gil, Martí Quixal, Roser Sanromà. Also thanks to Louise McNally, Maite Melero, Martí Quixal, and three anonymous reviewers for revision and criticism of previous versions of the paper. We thank Eric Joanis, Alexander Koller, and Oana Postolache for suggestions that lead to this paper. Special thanks are due to Roser Sanromà for kindly providing us with her manual morphological classification (Sanromà, 2003), and to the Institut d'Estudis Catalans for lending us the research corpus. This work is supported by the Fundación Caja Madrid.

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 the 3rd LREC*, pages 1130-1135.
- 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 Workshop on Linguistic Knowledge Acquisition and Representation at the 3rd LREC Conference*.
- G. Boleda, T. Badia, and E. Batlle. 2004. Acquisition of semantic classes for adjectives from distributional evidence. In *Proceedings of the 20th COLING*, pages 1119-1125.
- J. Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249-254.
- P. Carvalho and E. Ranchhod. 2003. Analysis and disambiguation of nouns and adjectives in Portuguese by FST. In *Proceedings of the Workshop on Finite-State Methods for Natural Language Processing at EACL2003*, pages 105-112.
- DLC. 1993. *Diccionari de la Llengua Catalana*. Enciclopèdia Catalana, Barcelona, third edition.
- V. Hatzivassiloglou and K. R. McKeown. 1993. Towards the automatic identification of adjectival scales: Clustering adjectives according to meaning. In *Proceedings of the 31st ACL*, pages 172-182.
- A. Korhonen, Y. Krymolowski, and Z. Marx. 2003. Clustering polysemic subcategorization frame distributions semantically. In *Proceedings of the 41st ACL*, pages 64-71.
- 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.
- C. Picallo. 2002. L'adjectiu i el sintagma adjectival. In Joan Solà, editor, *Gramàtica del català contemporani*, pages 1643-1688. Empúries, Barcelona.
- R. Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco.
- J. Rafel. 1994. Un corpus general de referència de la llengua catalana. *Caplletra*, 17:219-250.
- V. Raskin and S. Nirenburg. 1998. An applied ontological semantic microtheory of adjective meaning for natural language processing. *Machine Translation*, 13:135-227.
- S. Schulte im Walde and C. Brew. 2002. Inducing German semantic verb classes from purely syntactic subcategorisation information. In *Proceedings of the 40th ACL*, pages 223-230.
- Roser Sanromà. 2003. Aspectes morfològics i sintàctics dels adjectius en català. Master's thesis, Universitat Pompeu Fabra.