Enhancing FreeLing Rule-Based Dependency Grammars with Subcategorization Frames

Marina Lloberes
U. de Barcelona
Barcelona, Spain
mllobesa8@alumnes.ub.edu

Irene Castellón
U. de Barcelona
Barcelona, Spain
icastellon@ub.edu

Lluís Padró
U. Politècnica de Catalunya
Barcelona, Spain
padro@cs.upc.edu

Abstract

Despite the recent advances in parsing, significant efforts are needed to improve the current parsers performance, such as the enhancement of the argument/adjunct recognition. There is evidence that verb subcategorization frames can contribute to parser accuracy, but a number of issues remain open. The main aim of this paper is to show how subcategorization frames acquired from a syntactically annotated corpus and organized into fine-grained classes can improve the performance of two rule-based dependency grammars.

1 Introduction

Statistical parsers and rule-based parsers have advanced over recent years. However, significant efforts are required to increase the performance of current parsers (Klein and Manning, 2003; Nivre et al., 2006; Ballesteros and Nivre, 2012; Marimon et al., 2014).

One of the linguistic phenomena which parsers often fail to handle correctly is the argument/adjunct distinction (Carroll et al., 1998). For this reason, the main goal of this paper is to test empirically the accuracy of rule-based dependency grammars working exclusively with syntactic rules or adding subcategorization frames to the rules.

A number of studies shows that subcategorization frames can contribute to improve parser performance (Carroll et al., 1998; Zeman, 2002; Mirroshandel et al., 2013). Particularly, these studies are mainly concerned with the integration of subcategorization information into statistical parsers.

The list of studies about rule-based parsers integrating subcategorization information is also extensive (Lin, 1998; Alsina et al., 2002; Bick, 2006; Calvo and Gelbukh, 2011). However, they do not explicitly relate the improvements in parser performance to the addition of subcategorization.

This paper analyses in detail how subcategorization frames acquired from an annotated corpus and distributed among fine-grained classes increase accuracy in rule-based dependency grammars.

The framework used is that of the FreeLing Dependency Grammars (FDGs) for Spanish and Catalan, using enriched lexical-syntactic information about the argument structure of the verb. FreeLing (Padró and Stanilovsky, 2012) is an open-source library of multilingual Natural Language Processing (NLP) tools that provide linguistic analysis for written texts. The FDGs are the core of the FreeLing dependency parser, the Txala Parser (Atserias et al., 2005).

The remainder of this paper is organized as follows. Section 2 contains an overview of previous work related to this research. Section 3 presents the rule-based dependency parser used and the Spanish and Catalan grammars. Section 4 describes the strategy followed initially to integrate subcategorization into the grammars and how this information has been redesigned. Section 5 focuses on the evaluation and the analysis of several experiments testing versions of the grammars including or discarding subcategorization frames. Finally, the main conclusions and the further research goals arisen from the results of the experiments are exposed in Section 6.

2 Related Work

There has been an extensive research on parser development, and most approaches can be classified as *statistical* or *rule-based*. In the former, a statistical model learnt from annotated or unannotated texts is applied to build the syntactic tree (Klein and Manning, 2003; Collins and Koo, 2005; Nivre et al., 2006; Ballesteros and Nivre, 2012), whereas the latter uses hand-built grammars to guide the

parser in the construction of the tree (Sleator and Temperley, 1991; Järvinen and Tapanainen, 1998; Lin, 1998).

Concerning the languages this study is based on, some research on Spanish has been performed from the perspective of Constraint Grammar (Bick, 2006), Unification Grammar (Ferrández and Moreno, 2000), Head-Driven Phrase Structure Grammar (Marimon et al., 2014), and Dependency Grammar for statistical parsing, both supervised (Carreras et al., 2006) and semi-supervised (Calvo and Gelbukh, 2011). For Catalan, a rule-based parser based on Constraint Grammar (Alsina et al., 2002) and a statistical dependency parser (Carreras, 2007) are available.

Despite the huge achievements in the area of parsing, argument/adjunct recognition is still a linguistic problem in which parsers still show low accuracy and in which there is still no generalized consensus in Theoretical Linguistics (Tesnière, 1959; Chomsky, 1965). This phenomenon refers to the subcategorization notion, which corresponds to the definition of the type and the number of arguments of a syntactic head.

The acquisition of subcategorization frames from corpora is one of the strategies for integrating information about the argument structure into a parser. Depending on the level of language analysis of the annotated corpus, two main strategies are used in automatic acquisition.

If the acquisition is performed over a *morphosyntactically annotated text*, the subcategorization frames are inferred by applying statistical techniques on morphosyntactically annotated data (Brent, 1993; Manning, 1993; Korhonen et al., 2003).

Alternatively, acquisition can be performed with *syntactically annotated texts* (Sarkar and Zeman, 2000; O'Donovan et al., 2005; Aparicio et al., 2008). Subcategorization acquisition can be performed straightforwardly because the information about the argument structure is available in the corpus. Therefore, this approach generally focuses on the methods for subcategorization frames classification.

The final classification in a lexicon of frames is a computational resource for several NLP tools. In the framework which this research focuses on, the integration of the acquired subcategorization is orientated to the contribution towards building the syntactic tree when the parser has incomplete in-

formation to make a decision (Carroll et al., 1998).

Depending on the characteristics of the parser, subcategorization assists in this task in a different way. Subcategorization information can be used to assign a probability to every possible syntactic tree and to rank them in parsers that perform the whole set of possible syntactic analysis of a particular sentence (Carroll et al., 1998; Zeman, 2002; Mirroshandel et al., 2013).

In contrast, subcategorization may help to restrict the application of certain rules. Then, when the parser detects the subcategorization frame in the input sentence, it labels the syntactic tree according to the frame discarding any other possible analysis (Lin, 1998; Calvo and Gelbukh, 2011).

3 Dependency Parsing in FreeLing

The rule-based dependency grammars presented in this article are the core of the Txala Parser (Atserias et al., 2005), the NLP module in charge of Dependency Parsing in the FreeLing library (Padró and Stanilovsky, 2012).[1]

FreeLing is an open-source project that has been developed for more than ten years. It is a complete NLP pipeline built on a chain of modules that provide a general and robust linguistic analysis. Among the available tools, FreeLing offers sentence recognition, tokenization, named entity recognition, tagging, chunking, dependency parsing, word sense disambiguation, and coreference resolution.

3.1 Txala Parser

The Txala Parser is one of the dependency parsing modules available in FreeLing. It is a rule-based, non-projective and multilingual dependency parser that provides robust syntactic analysis in three steps.

Txala receives the partial syntactic trees produced by the chunker (Civit, 2003) as input. Firstly, the head-child relations are identified using a set of heuristic rules that iteratively decide whether two adjacent trees must be merged, and in which way, until there is only one tree left. Secondly, it is converted into syntactic dependencies according to Mel'čuk (1988). Finally, each dependency arch of the tree is labelled with a syntactic function tag.

---

[1]http://nlp.cs.upc.edu/freeling/

| | **Rules** | | |
Language	Total	Linking	Labelling
English	2961	2239	722
Spanish	4042	3310	732
Catalan	2879	2099	780
Galician	178	87	91
Asturian	4438	3842	596

Table 1: Sizes of the FDGs

3.2 FreeLing Dependency Grammars

The current version of FreeLing includes rule-based dependency grammars for English, Spanish, Catalan, Galician and Asturian (see Table 1 for a brief overview of their sizes). In this paper, the Spanish and Catalan dependency grammars are described.

The FDGs follow the linguistic basis of syntactic dependencies (Tesnière, 1959; Mel'čuk, 1988). However, we propose a different analysis for prepositional phrases (preposition-headed), subordinate clauses (conjunction-headed) and co-ordinating structures (conjunction-headed).

A FDG is structured as a set of manually defined rules which link two adjacent syntactic partial trees (*linking rules*) and assign a syntactic function to every link of the tree (*labelling rules*), according to certain conditions and priority. They are applied based on this priority: at every step, two adjacent partial trees will be attached or will be labelled with a syntactic function tag if their rule is the highest ranked for which all the conditions are met.

Linking rules can contain four kind of conditions, regarding morphological (part-of-speech tag), lexical (word form, lemma), syntactic (syntactic context, syntactic features of lemmas) and semantic features (semantic properties predefined by the user).

For instance, the rule shown in Figure 1 has priority `911`, and states that a sub-tree marked as a subordinate clause (`subord`) whose head is a relative pronoun (`PR`) attached as a child to the noun phrase (`sn`) to its left (`top_left`) when these two consecutive sub-trees are not located to the right of a verb phrase (`!grup-verb_$$`).

Concerning the *labelling rules*, the set of conditions that the parent or the child of the dependency must meet may refer to morphological (part-of-speech tag), lexical (word form, lemma), syntactic (lower/upper sub-tree nodes, syntactic features of lemmas) and semantic properties (EuroWord-Net Top Concept Ontology -TCO- features, Word-

```
911 !grup-verb_$$ - (sn,subord{^PR})
       top_left RELABEL -
```

Figure 1: Linking rule for relative clauses

```
grup-verb    dobj
             d.label=grup-sp
             p.class=trans
             d.side=right
             d.lemma=a|al
             d:sn.tonto=Human
             d:sn.tonto!=Building|Place
```

Figure 2: Labelling rule for human direct objects

Net Semantic File, WordNet Synonyms and Hypernyms and other semantic features predefined by the user).

In the rule illustrated in Figure 2, the direct object label (`dobj`) is assigned to the link between a verbal head (`grup-verb`) and a prepositional phrase (`grup-sp`) child when the head belongs to the transitive verbs class (`trans`) and the child is post-verbal (`right`), the preposition is `a` (or the contraction `al`), and the nominal head inside the prepositional phrase has the TCO feature `Human` but not (`!=`) the features `Building` or `Place` (to prevent organizations from being identified as a direct object).

4 CompLex-VS lexicon for Parsing

Following the hypothesis that subcategorization frames improve the parsing performance (Carroll et al., 1998), the first version of FDGs included verbal and nominal frames in order to improve argument/adjunct recognition and prepositional attachment (Lloberes et al., 2010). In this paper, only the verbal lexicon is presented because it is the resource used for the argument/adjunct recognition task in the grammars.

4.1 Initial CompLex-VS lexicon in FDGs

The initial Computational Lexicon of Verb Subcategorization (CompLex-VS) was automatically extracted from the subcategorization frames of the SenSem Corpus (Fernández and Vàzquez, 2014), which contains 30231 syntactically and semantically annotated sentences per language, and of the Volem Multilingual Lexicon (Fernández et al., 2002), which has 1700 syntactically and semantically annotated verbal lemmas per language. The patterns extracted from both resources are orga-

nized according to the linguistic-motivated classification proposed by Alonso et al. (2007).

The final lexicon applied to the FDGs has 11 subcategorization classes containing a total of 1314 Spanish verbal lemmas and 847 Catalan verbal lemmas with a different subcategorization frame.

A first experimental evaluation of the Spanish Grammar with the initial subcategorization lexicon (Lloberes et al., 2010) showed that incorporating subcategorization information is promising.

4.2 Redesign of the CompLex-VS lexicon

According to the evaluation results of the grammars with the initial CompLex-VS included, the lexicon has been redesigned, proposing a set of more fine-grained subcategorization frame classes in order to represent verb subcategorization in the dependency rules in a controlled and detailed way.

New syntactic-semantic patterns have been extracted automatically from the SenSem Corpus according to the idea that every verbal lemma with a different subcategorization frame expresses a different meaning. Therefore, a new lexicon entry is created every time an annotated verbal lemma with a different frame is detected.

The CompLex-VS contains 3102 syntactic patterns in the Spanish lexicon and 2630 patterns in the Catalan lexicon (see Section 4.3 for detailed numbers). They are organized into 15 subcategorization frames as well as into 4 subcategorization classes. The lexicon is distributed in XML format under the Creative Commons Attribution-ShareAlike 3.0 Unported License.[2]

Certain patterns have been discarded because they are non-prototypical in the corpus (e.g. clitic left dislocations), they alter the sentence order (e.g. relative clauses), or they involve controversial argument classes (e.g. prepositional phrases seen as arguments or adjuncts depending on the context).

As Figure 3 shows, the extracted patterns (`<verb>`) have been classified into `<frame>` classes according to the whole set of argument structures occurring in the corpus (`subj` for intransitive verbs, `subj,dobj` for transitive verbs, etc.). Simultaneously, frames have been organized in `<subcategorization>` classes (monoargumental, biargumental, triargumental and quatri-argumental).

```
<subcategorization
    class="monoargumental"
    ref="1" freq="0.188480">
    <frame class="subj" ref="1"
        freq="0.188480">
        <verb lemma="pensar"
            id="2531"
            ref="1:1" fs="subj"
            cat="np" rs="exp"
            head="null"
            construction="active"
            se="no" freq="0.000070"/>
    </frame>
</subcategorization>
<subcategorization
    class="biargumental"
    ref="2" freq="0.733349">
    <frame class="subj,dobj" ref="2"
        freq="0.617452">
        <verb lemma="agradecer"
            id="454" ref="2:2" fs="subj,dobj"
            cat="np,complsc" rs="ag_exp,t"
            head="null,null"
            construction="active"
            se="no" freq="0.000140"/>
    </frame>
</subcategorization>
```

Figure 3: Example of the CompLex-VS

Every lexicon entry contains the syntactic function of every argument (`fs`), the grammatical category of the head of the argument (`cat`) and the thematic role (`rs`). The type of `construction` (e.g. active, passive, impersonal, etc.) has been inferred from the predicate and aspect annotations available in the SenSem Corpus.

Two non-annotated lexical items of the sentence have also been inserted into the subcategorization frame because the information that they provide is crucial for the argument structure configuration (e.g. the particle '`se`' and the lexical value of the prepositional phrase `head`).

In addition, meta-linguistic information has been added to every entry: a unique `id` and the relative frequency of the pattern in the corpus (`freq`). A threshold frequency has been established at $7 \cdot 10^{-5}$ (Spanish) and at $8.5 \cdot 10^{-5}$ (Catalan). Patterns below this threshold have been considered marginal in the corpus and they have been discarded.

Every pattern contains a link to the frame and subcategorization class that they belong to (`ref`). For example, if an entry has the reference `1:1`, it means that the pattern corresponds to a monoargumental verb whose unique argument is a subject.

4.3 Integration of CompLex-VS in the FDGs

From the CompLex-VS, two derived lexicons per language containing the verbal lemmas for every recorded pattern have been created to be integrated into the FDGs. The CompLex-SynF lexicon con-

Frames	Spanish	Catalan
subj	203	386
subj,att	3	7
subj,dobj	440	230
subj,iobj	37	61
subj,pobj	126	93
subj,pred	45	31
subj,attr,iobj	2	1
subj,dobj,iobj	113	72
subj,dobj,pobj	42	34
subj,dobj,pred	21	18
subj,pobj,iobj	2	1
subj,pobj,pobj	14	9
subj,pobj,pred	1	0
subj,pred,iobj	4	5
subj,dobj,pobj,iobj	1	0

Table 2: CompLex-SynF lexicon in numbers

tains the subcategorization patterns generalized by the syntactic function (Table 2). The CompLex-SynF+Cat lexicon collects the syntactic patterns combining syntactic function and grammatical category (adjective/noun/prepositional phrase, infinitive/interrogative/completive clause).

The addition of grammatical categories makes it possible to restrict the grammar rules. For example, a class of verbs containing the verb *quedarse* ('to get') whose argument is a predicative and a prepositional phrase allows the rules to identify that the prepositional phrase of the sentence *Se ha quedado de piedra* ('[He/She] got shocked') is a predicative argument. Furthermore, it allows for discarding the prepositional phrase of the sentence *Aparece de madrugada* ('[He/She] shows up at late night') being a predicative argument, although *aparecer* belongs to the class of predicative verbs but conveying a noun phrase as argument.

While in the CompLex-SynF lexicon the information is more compacted (1054 syntactic patterns classified in 15 frames), in the CompLex-SynF+Cat lexicon the classes are more granular (1356 syntactic patterns organized in 77 frames).

Only subcategorization patterns corresponding to lexicon entries referring to the active voice have been integrated in the FDGs, since they involve non-marked word order. Both lexicons also exclude information about the thematic role, although they take into account the value of the head (if the frame contains a prepositional argument) and the pronominal verbs (lexical entries that accept 'se' particle whose value neither is reflexive nor reciprocal).

Two versions of the Spanish dependency grammar and two versions of the Catalan dependency

Grammar	Spanish	Catalan
Bare	450	508
Baseline	732	780
SynF	872	917
SynF+Cat	869	917

Table 3: Labelling rules in the evaluated grammars

grammar have been created. One version contains the CompLex-SynF lexicon and the other one the CompLex-Synf+Cat.

The old CompLex-VS lexicon classes have been replaced with the new ones. Specifically, this information has been inserted in the part of the labelling rules about the syntactic properties of the parent node (observe `p.class` in Figure 2).

Finally, new rules have been added for frames of CompLex-SynF and CompLex-SynF+Cat that are not present in the old CompLex-VS lexicon. Furthermore, some rules have been disabled for frames of the old CompLex-VS lexicon that do not exist in the CompLex-SynF and CompLex-SynF+Cat lexicons (see Table 3 for the detailed size of the grammars).

5 Evaluation

An evaluation task has been carried out to test empirically how the FDGs performance changes when subcategorization information is added or subtracted. Several versions of the grammars have been tested using a controlled annotated linguistic data set.

This evaluation specifically focuses on analysing the results of the experiments qualitatively. This kind of analysis makes it possible to track the decisions that the parser has made, so that it is possible to provide an explanation about the accuracy of the FDGs running with different linguistic information.

5.1 Experiments

Four versions of both Spanish and Catalan grammars are tested in order to assess the differences of the performance depending on the linguistic information added.

- Bare FDG. A version of the FDGs running without subcategorization frames.

- Baseline FDG. A version of the FDGs running with the old CompLex-VS lexicon.

- SynF FDG. A version of the FDGs running with the CompLex-SynF lexicon.

Tag	SenSem Spanish	ParTes Spanish	SenSem Catalan	ParTes Catalan
subj	42.23	34.03	43.03	28.08
dobj	35.77	29.86	34.64	34.25
pobj	16.73	13.89	16.56	17.12
iobj	4.64	6.25	4.70	2.05
pred	0.49	2.08	0.51	0.68
attr	0.14	13.89	0.56	17.81

Table 4: Comparison of the labelling tags distribution in SenSem and ParTes (%)

Tag	Description
adjt	Adjunct
attr	Attribute
dobj	Direct Object
iobj	Indirect Object
pobj	Prepositional Object
pred	Predicative
subj	Subject

Table 5: Tagset of syntactic functions related to the subcategorization

- SynF+Cat FDG. A version of the FDGs running with the CompLex-Synf+Cat lexicon.

Since this research is focused on the implementation of subcategorization information for argument/adjunct recognition, only the *labelling rules* are discussed in this paper (Table 3). However, metrics related to *linking rules* are also mentioned to provide a general description of the FDGs.

5.2 Evaluation data

To perform a qualitative evaluation, the ParTes test suite has been used (Lloberes et al., 2014). This resource is a multilingual hierarchical test suite of a representative and controlled set of syntactic phenomena which has been developed for evaluating the parsing performance as regards syntactic structure and word order.

It contains 161 syntactic phenomena in Spanish (99 referring to structure and 62 to word order) and 147 syntactic phenomena in Catalan (101 corresponding to structure phenomena and 46 to word order).

The current version of ParTes is distributed with an annotated data set in the CoNLL format. Although this data set is not initially developed for evaluating the argument/adjunct recognition, the number of arguments and adjuncts contained in ParTes is proportional to the number of arguments and adjuncts of the SenSem Corpus (Table 4). Therefore, the ParTes data set is a reduced sample of the linguistic phenomena that occur in a larger corpus, which makes ParTes an appropriate resource for this task.

5.3 Evaluation metrics

The metrics have been computed using the CoNLL-X Shared Task 2007 script (Nivre et al., 2007). The output of the FDGs (*system output*) has been compared to the ParTes annotated data set (*gold standard*).

The metrics used to evaluate the performance of the several FDGs versions are the following ones:

Accuracy[3]

$$\text{LAS} = \frac{\text{correct attachments and labellings}}{\text{total tokens}}$$

$$\text{UAS} = \frac{\text{correct attachments}}{\text{total tokens}}$$

$$\text{LAS2} = \frac{\text{correct labellings}}{\text{total tokens}}$$

Precision

$$\text{P} = \frac{\text{system correct tokens}}{\text{system tokens}}$$

Recall

$$\text{R} = \frac{\text{system correct tokens}}{\text{gold tokens}}$$

Both quantitative and qualitative analysis detailed in Section 5.4 pay special attention to the metric LAS2, which informs about the number of heads with the correct syntactic function tag.

Precision and recall metrics of the labelling rules provide information about how the addition of verbal subcategorization information contributes to the grammar performance. For this reason, in the qualitative analysis, only labelling syntactic function tags directly related to verbal subcategorization are considered (Table 5).

5.4 Accuracy results

The global results of the FDGs evaluation (LAS) show that the whole set of evaluated grammars score over 80% accuracy in Spanish (Table 6) and around 80% in Catalan (Table 7).

In the four Spanish grammar versions (Table 6), the correct head (UAS) has been identified in 90.01% of the cases. On the other hand, the tendency changes in syntactic function labelling (LAS2). The *Baseline* establishes that 85.54% of tokens have the correct syntactic function tag.

---

[3]LAS: Labeled Attachment Score; UAS: Unlabeled Attachment Score; LAS2: Label Accuracy

Grammar	LAS	UAS	LAS2
Bare	81.37	90.01	82.86
Baseline	83.76	90.01	85.54
SynF	84.50	90.01	86.29
SynF+Cat	84.50	90.01	86.29

Table 6: Accuracy scores (%) in Spanish

Grammar	LAS	UAS	LAS2
Bare	78.99	86.84	81.91
Baseline	79.52	86.84	82.85
SynF	81.78	86.84	85.24
SynF+Cat	81.78	86.84	85.24

Table 7: Accuracy scores (%) in Catalan

However, *Bare* drops 2.68 scores and *SynF* and *SynF+Cat* improve 0.75 scores with respect to the baseline.

A parallel behaviour is observed in Catalan, although the scores are slightly lower than in Spanish (Table 7). The four Catalan grammars score 86.84% in attachment (UAS). The *Baseline* scores 82.85% in syntactic function assignment (LAS2). Once again FDGs perform worse without subcategorization information (0.94 points less in *Bare* grammar) and better with subcategorization information (2.39 points more in *SynF* and *SynF+Cat*).

From a general point of view, accuracy metrics show a medium-high accuracy performance of all versions of FDGs in both languages. Specifically, these first results highlight that subcategorization information helps with the syntactic function labelling. However, qualitative results will reveal how subcategorization influences the grammar performance (Sections 5.5 and 5.6).

5.5 Precision results

As observed in the quantitative analysis (Section 5.4), in both languages most of the syntactic function assignments drop in precision when subcategorization classes are blocked in the grammar (Tables 8 and 9), whereas syntactic function labelling tends to improve when subcategorization is available.

For example, the precision of the prepositional object (*pobj*) in both languages drops drastically when subcategorization is disabled (*Bare*). On the contrary, the precision improves significantly when the rules include subcategorization information (*Baseline*). Furthermore, the introduction of more fine-grained frames helps the grammars reach a precision of 94.74% in Spanish and 94.12% in Catalan (*SynF* and *SynF+Cat*). Fig-
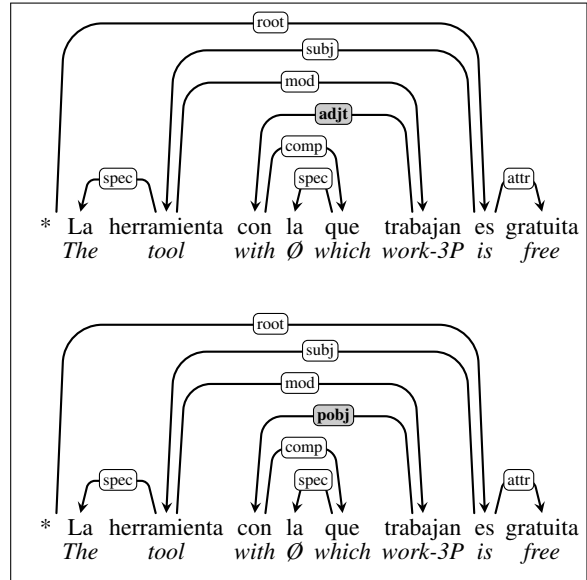

Figure 4: Example of *bare* FDGs wrongly labelling a *pobj* as *adjt* (above) and of *SynF* FDGs correctly labelling it (below)

Tag	Bare	Baseline	SynF	SynF+Cat
adjt	59.26	70.27	61.54	61.54
attr	84.21	71.43	71.43	71.43
dobj	78.26	85.71	87.80	87.80
iobj	100.00	100.00	100.00	100.00
pobj	42.50	77.27	94.74	94.74
pred	12.50	0.00	33.33	33.33
subj	90.24	90.91	91.11	91.11

Table 8: Labelling precision scores (%) in Spanish

ure 4 shows this dichotomy.

Despite these improvements, some items differ from the general tendency.

In Spanish, the improvement of the copulative verbs (*attr*) is due to lexical information in the *Bare* FDG, while they keep stable in *SynF* and *SynF+Cat*. Precision remains the same in the indirect object (*iobj*) because morphological information is enough to detect dative clitics in singular.

The performance of predicative (*pred*) in all the grammars is related to the lack or addition of subcategorization. The *Baseline* FDG subcategorization classes do not include the same set of verbs as in the evaluation data. For this reason, a generic rule for capturing predicatives (*Bare* FDG) covers the lack of verbs in a few cases. The improvement of the coverage with new verbs (*SynF* and *SynF+Cat*) shows an increment of the precision.

Adjunct (*adjt*) recognition drops for mislabellings with predicative because of the ambiguity between the participle clause expressing time and a true predicative complement.

Tag	Bare	Baseline	SynF	SynF+Cat
adjt	60.71	61.76	62.50	62.50
attr	95.65	82.14	95.83	95.83
dobj	75.00	83.33	84.78	82.98
iobj	100.00	100.00	100.00	100.00
pobj	61.29	66.67	94.12	94.12
pred	50.00	0.00	100.00	100.00
subj	72.50	71.43	73.81	73.81

Table 9: Labelling precision scores (%) in Catalan

Tag	Bare	Baseline	SynF	SynF+Cat
adjt	57.14	92.86	85.71	85.71
attr	80.00	100.00	100.00	100.00
dobj	83.72	83.72	83.72	83.72
iobj	33.33	33.33	44.44	44.44
pobj	85.00	85.00	90.00	90.00
pred	33.33	0.00	33.33	33.33
subj	75.51	81.63	83.67	83.67

Table 10: Labelling recall scores (%) in Spanish

FDGs in Catalan show a parallel behaviour to that in Spanish, but they follow the general tendency in more cases. *SynF* and *SynF+Cat* increase the precision in all the cases, except for the direct object (*dobj*) in *SynF+Cat*. Once more the prepositional object (*pobj*) performance raises when subcategorization frames are available.

Although a drop in all the cases in the *Bare* FDG is expected, the attribute (*attr*) and the predicative (*pred*) increase the precision because of the same reasons as the Spanish grammars.

The results of *SynF* and *SynF+Cat* are almost identical. The analysis of their outputs shows that more fine-grained subcategorization classes including grammatical categories do not have a contribution to the precision improvement.

5.6 Recall results

The addition of subcategorization information in the FDGs also contributes to the improvement, almost in all the cases, in Spanish as well as in Catalan (Tables 10 and 11). The use of FDGs without subcategorization involves a decrease in the recall most of times.

In Spanish, the *Baseline* grammar contains very generic rules to capture adjuncts and more fine-grained subcategorization classes restrict these rules. For this reason, the recall slightly drops in *SynF* and *SynF+Cat*. As observed in the precision metric (Section 5.5), small populated classes related to predicative arguments make recall drop in the baseline. Consequently, generic rules for predicative labelling in the *Bare* grammar and better populated predicative classes in *SynF* and *SynF+Cat* allows a recovery in recall.

FDGs in Catalan show a similar tendency. In the *Bare* grammar, prepositional objects and predicatives are better captured than in the baseline because the lack of subcategorization information allows rules to apply in a more irrestrictive way. On the other hand, the addition of subcategorization information does not seem to help with capturing

Tag	Bare	Baseline	SynF	SynF+Cat
adjt	50.00	61.76	73.53	73.53
attr	84.62	88.46	88.46	88.46
dobj	72.00	80.00	78.00	78.00
iobj	33.33	33.33	33.33	33.33
pobj	76.00	56.00	64.00	64.00
pred	100.00	0.00	100.00	100.00
subj	70.73	73.17	75.61	75.61

Table 11: Labelling recall scores (%) in Catalan

more direct objects. Lower results are due to some verbs missing.

Once again there are no significant differences between *SynF* and *SynF+Cat*, which reinforces the idea that grammatical categories do not provide new information for capturing new argument and adjuncts.

5.7 Analysis of the results

The whole set of experiments demonstrate that subcategorization improves significantly the performance of the rule-based FDGs.

However, some arguments, such as the prepositional object and the predicative, are difficult to capture without subcategorization information. Meanwhile, there are others, such as the attribute, that do not need to be handled with subcategorization classes.

Proper subcategorization information also contributes to capture more arguments and adjuncts. The recall scores are stable among the grammars that use subcategorization information. Secondly, most of these scores are medium-high precision.

Overall, the results show that the new CompLex-VS is a suitable resource to improve the performance of rule-based dependency grammars.

The classification of frames proposed is coherent with the methodology. Furhtermore, it is an essential resource for the grammars tested since it ensures medium-high precision results (compared to medium precision results in the FDGs using the old CompLex-VS). It is important to consider the kind of information to define the subcategorization

classes because it can be redundant, such as the combination of syntactic function and grammatical category.

The CompLex-VS lexicon still needs the inclusion of new verbs, since some arguments for verbs missing in the lexicon are not captured properly.

6 Conclusions

This paper presented two rule-based dependency grammars in Spanish and Catalan for the FreeLing NLP library.

Besides the grammars, a new subcategorization lexicon, CompLex-VS, has been designed using frames acquired from the SenSem Corpus. The new frames have been integrated in the argument/adjunct recognition rules of the FDGs.

A set of experiments has been carried out to test how the subcategorization information improves the performance of these grammars.

The results show that subcategorization frames ensure a high accuracy performance. In most cases, the old CompLex-VS frames and the new CompLex-VS frames show an improvement.

However, the increment is more evident in some arguments –such as the prepositional object and the predicative– than others, like the complement in attributive verbs. These results indicate that some arguments necessarily need subcategorization information to be disambiguated, while others can be disambiguated just with syntactic information.

Furthermore, the new frames of CompLex-VS provide better results than the initial ones. Therefore, more fine-grained frames (CompLex-SynF) contribute to raise the accuracy. Despite this evidence, fine-grained classes do not necessarily mean improvement of the parser performance. The most fine-grained lexicon (CompLex-SynF+Cat), which combines syntactic function and grammatical category information, neither improves nor worsens the results of the FDGs.

These conclusions are built on a small set of test data. Although it is a controlled and representative evaluation data set, these results need to be contrasted with a larger evaluation data set.

It would be interesting to evaluate how the parsing performance improves while subcategorization information is added incrementally.

References

L. Alonso, I. Castellón, and N. Tincheva. 2007. Obtaining coarse-grained classes of subcategorization patterns for Spanish. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*.

À. Alsina, T. Badia, G. Boleda, S. Bott, À. Gil, M. Quixal, and O. Valentn. 2002. CATCG: Un sistema de análisis morfosintáctico para el catalán. *Procesamiento del Lenguaje Natural*, 29.

J. Aparicio, M. Taulé, and M.A. Martí. 2008. AnCora-Verb: A Lexical Resource for the Semantic Annotation of Corpora. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*.

J. Atserias, E. Comelles, and A. Mayor. 2005. TXALA un analizador libre de dependencias para el castellano. *Procesamiento del Lenguaje Natural*, 35.

M. Ballesteros and J. Nivre. 2012. MaltOptimizer: A System for MaltParser Optimization. In *Proceedings of the Eight International Conference on Language Resources and Evaluation*.

E. Bick. 2006. A Constraint Grammar-Based Parser for Spanish. In *Proceedings of TIL 2006 - 4th Workshop on Information and Human Language Technology*.

M.R. Brent. 1993. From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax. *Computational Linguistics*, 19(2).

H. Calvo and A. Gelbukh. 2011. *DILUCT: Análisis Sintáctico Semisupervisado Para El Español*. Editorial Academica Espanola.

X. Carreras, M. Surdeanu, and L. Màrquez. 2006. Projective Dependency Parsing with Perceptron. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*.

X. Carreras. 2007. Experiments with a Higher-Order Projective Dependency Parser. In *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007*.

J. Carroll, G. Minnen, and T. Briscoe. 1998. Can Subcategorisation Probabilities Help a Statistical Parser? In *Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora*.

N. Chomsky. 1965. *Aspects of the Theory of Syntax*. MIT Press.

M. Civit. 2003. Criterios de etiquetación y desambiguación morfosintáctica de corpus en español. In *Colección de Monografías de la Sociedad Española para el Procesamiento del Lenguaje Natural: 8*. Sociedad Española para el Procesamiento del Lenguaje Natural.

M. Collins and T. Koo. 2005. Discriminative Reranking for Natural Language Parsing. *Computational Linguistics*, 31(1).

A. Fernández and G. Vàzquez. 2014. The SenSem Corpus: an annotated corpus for Spanish and Catalan with information about aspectuality, modality, polarity and factuality. *Corpus Linguistics and Linguistic Theory*, 10(2).

A. Fernández, G. Vazquez, P. Saint-Dizier, F. Benamara, and M. Kamel. 2002. The VOLEM Project: A Framework for the Construction of Advanced Multilingual Lexicons. In *Proceedings of the Language Engineering Conference*.

A. Ferrández and L. Moreno. 2000. Slot Unification Grammar and Anaphora Resolution. In N. Nicolov and R. Mitkov, editors, *Recent Advances in Natural Language Processing II. Selected papers from RANLP 1997*. John Benjamins Publishing Co.

T. Järvinen and P. Tapanainen. 1998. Towards an implementable dependency grammar. In *Proceedings of Workshop on Processing of Dependence-Based Grammars, CoLing-ACL'98*.

D. Klein and C.D. Manning. 2003. Accurate Unlexicalized Parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*.

A. Korhonen, Y. Krymolowski, and Z. Marx. 2003. Clustering Polysemic Subcategorization Frame Distributions Semantically. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*.

D. Lin. 1998. Dependency-Based Evaluation of MINIPAR. In *Workshop on the Evaluation of Parsing Systems, First International Conference on Language Resources and Evaluation*.

M. Lloberes, I. Castellón, and L. Padró. 2010. Spanish FreeLing Dependency Grammar. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation*.

M. Lloberes, I. Castellón, L. Padró, and E. Gonzàlez. 2014. ParTes. Test Suite for Parsing Evaluation. *Procesamiento del Lenguaje Natural*, 53.

C.D. Manning. 1993. Automatic Acquisition of a Large Subcategorization Dictionary from Corpora. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*.

M. Marimon, N. Bel, and L. Padró. 2014. Automatic Selection of HPSG-parsed Sentences for Treebank Construction. *Computational Linguistics*, 40(3).

I.A. Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. State U. Press of NY.

S.A. Mirroshandel, A. Nasr, and B. Sagot. 2013. Enforcing Subcategorization Constraints in a Parser Using Sub-parses Recombining. In *NAACL 2013 - Conference of the North American Chapter of the Association for Computational Linguistics*.

J. Nivre, J. Hall, J. Nilsson, G. Eryiğit, and S. Marinov. 2006. Labeled Pseudo-projective Dependency Parsing with Support Vector Machines. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*.

J. Nivre, J. Hall, S. Kübler, R. McDonald, J. Nilsson, S. Riedel, and D. Yuret. 2007. The CoNLL 2007 Shared Task on Dependency Parsing. In *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007*.

R. O'Donovan, M. Burke, A. Cahill, J. Van Genabith, and A. Way. 2005. Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II and Penn-III Treebanks. *Computational Linguistics*, 31(3).

L. Padró and E. Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Eight International Conference on Language Resources and Evaluation*.

A. Sarkar and D. Zeman. 2000. Automatic Extraction of Subcategorization Frames for Czech. In *Proceedings of the 18th Conference on Computational Linguistics - Volume 2*.

D. Sleator and D. Temperley. 1991. Parsing English with a Link Grammar. In *Third International Workshop on Parsing Technologies*.

L. Tesnière. 1959. *Eléments de syntaxe structurale*. Klincksieck.

D. Zeman. 2002. Can Subcategorization Help a Statistical Dependency Parser? In *19th International Conference on Computational Linguistics*.