

Verb SCF extraction for Spanish with dependency parsing

Extracción de patrones de subcategorización de verbos en castellano con análisis de dependencias

Muntsa Padró
Universidade Federal do Rio
Grande do Sul
Av. Bento Gonçalves, 9500
Porto Alegre -Brasil
muntsa.padro@inf.ufrgs.br

Núria Bel
Universitat Pompeu Fabra
Roc Boronat 138
08018 Barcelona - Spain
nuria.bel@upf.edu

Aina Garí
Universitat de Barcelona
Gran Via de les Corts
Catalanes, 585
08007 Barcelona - Spain

Resumen: En este artículo presentamos los resultados de nuestros experimentos en producción automática de léxicos con información de patrones de subcategorización verbal para castellano. La investigación se llevó a cabo en el marco del proyecto PANACEA de adquisición automática de información léxica que redujera al máximo la intervención humana. En nuestros experimentos, se utilizó una cadena de diferentes herramientas que incluía ‘crawling’ de textos de un dominio particular, normalización y limpieza de los textos, segmentación, identificación de unidades, etiquetado categorial y análisis de dependencias antes de, finalmente, la extracción de los patrones de subcategorización. Los resultados obtenidos muestran una gran dependencia de la calidad de los analizadores de dependencias aunque, no obstante, están en línea con los resultados obtenidos en experimentos similares para otras lenguas.

Palabras clave: Adquisición automática de patrones de subcategorización, análisis de dependencias, adquisición léxica.

Abstract: In this paper we present the results of our experiments in automatic production of verb subcategorization frame lexica for Spanish. The work was carried out in the framework of the PANACEA project aiming at the automatic acquisition of lexical information reducing at maximum human intervention. In our experiments, a chain of different tools was used: domain focused web crawling, automatic cleaning, segmentation and tokenization, PoS tagging, dependency parsing and finally SCFs extraction. The obtained results show a high dependency on the quality of the results of the intervening components, in particular of the dependency parsing, which is the focus of this paper. Nevertheless, the results achieved are in line with the state-of-the-art for other languages in similar experiments.

Keywords: Automatic subcategorization frame acquisition, dependency parsing, lexical acquisition.

1 Introduction

Knowledge of Subcategorization Frames (SCF) implies the ability to distinguish, given a predicate in raw text and its co-occurring phrases, which of those phrases are arguments (obligatory or optional) and which adjuncts. Access to SCF knowledge is useful for parsing as well as for other NLP tasks such as Information Extraction (Surdeanu et al., 2003) and Machine Translation (Hajič et al., 2002). SCF induction is also important for other

computational linguistic tasks such as automatic verb classification, selectional preference acquisition, and psycholinguistic experiments (Lapata et al., 2001; Schulte im Walde and Brew, 2002; McCarthy and Carroll, 2003; Sun et al., 2008a, 2008b).

In this paper we present the results of our experiments in automatic production of verb subcategorization frame lexica for Spanish with special focus on the use of statistical dependency parsing. The work was carried out in the framework of the PANACEA project (7FP-ICT-248064) aiming at the automatic

acquisition of lexical information reducing at maximum human intervention. Therefore, fully automation of the process and human work reduction were the main criteria for choosing methods and assessing the results. In our experiments, a chain of different tools was used: domain focused web crawling, automatic cleaning, segmentation and tokenization, PoS tagging, dependency parsing and finally SCFs induction¹. In the context of this project, we focused on maximizing precision. In order to contribute to the production of resources for working systems, we understood that it could be an asset to produce automatically a SCFs lexicon where good entries can be clearly separated from dubious ones. Then, human revision could concentrate on the dubious ones while still saving time and effort if those identified as reliable were actually good.

The obtained results show a high dependency on the quality of the results of the intervening components. Nevertheless, the results achieved are in line with the state-of-the-art for other languages in similar experiments.

2 Related work

The possibility to induce SCFs from raw corpus data has been investigated mostly for verbs and it is based on a first hypothesis generation step followed by a filtering step that tries to separate actual complement combination patterns from occasional combinations (see Korhonen, 2010, for a survey on different techniques for different languages). Current researched systems rely on the information supplied by an intermediate parser that identifies constituents and their grammatical function. Thus a first step collects sequences of constituents and their frequency, and a second step tries to select those combinations that are consistently found. Evaluation is made in terms of precision, i.e. only actual SCFs for a particular verb type, that is, a lemma, must be assigned, and coverage, i.e. all the possible SCFs for a particular verb type must be assigned. The main problem of current techniques has to do with maximizing both precision and coverage for each particular verb because (i) SCFs distribution is Zipfian and usual frequency filters fail to select infrequent patterns, and (ii) because the correlation between the conditional distribution of SCFs given a particular verb type and the

unconditional distribution independent of specific verb types is very small.

Verb SCF acquisition for Spanish has already been addressed. Chrupala (2003) presented a system to learn subcategorization frames for 10 frequent verbs² of two classes, verbs of change and verbs of path, from a 370,000 word corpus by adopting the existing scheme of classification of Spanish SCFs from the SENSEM verb database developed in the VOLEM project (Fernández et al. 2002). The experiment searched chunked corpora and detected potential SCFs for 10 Spanish verbs. Semantic information of nouns, in particular the ‘human’ feature, is added in the chunking step in order to handle phenomena such as direct objects marked with the preposition ‘a’. The SCFs hypothesis generation is based on matching the actual co-occurrences against a number of previously defined syntactic patterns associated with specific SCFs in the form of templates. A number of rules generate different variants of a number of initial, canonical templates.c For instance, a rule generates cliticized variants of full NP SCFs.

As for the evaluation, 20 sentences for each one of 10 verbs were randomly selected and system results were compared with a manually corrected version of the SCFs selected. The Chrupala (2003) system achieved a precision of 0.56 in token SCF detection. The results were also evaluated for types: for each verb the number of detected SCFs was collected and compared with the manual reference with a filtering phase based on a relative frequency cut-off. Best results published were, at a cut-off of 0.07, 0.77 precision and 0.70 recall.

Esteve Ferrer (2004) carried out the SCF extraction experiment on a corpus of 50 million words, also PoS tagged and chunked. The task was to assign acquired SCFs to verb types after the two explained phases of hypothesis generation and posterior filtering. A predefined list of 11 possible SCFs, each made of plausible combinations of a maximum of two constituents, were considered. The predefined SCFs considered different prepositions that were grouped manually with semantic criteria. Hypothesis selection was performed with a Maximum Likelihood Estimate (MLE, Korhonen and Krymolowski 2002). Evaluation was carried out comparing with a manually

¹ All the PANACEA materials and tools are available at www.panacea-lr.eu

² Bajar, convertir, dejar, desatar, deshacer, llenar, preocupar, reducir, sorprender, decir.

constructed gold standard for a sample of 41 randomly chosen verbs that included frequent but also infrequent verbs. These experiments gave the following results at a frequency cut-off of 0.05: 0.71 precision, 0.61 recall.

The main novelties of our work with respect to these previous experiments for SCF extraction for Spanish verbs are: the use of a dependency-parsed corpus and no need to have a list of predefined templates or SCFs to match. A further innovative aspect investigated in our project has to do with the amount of expert language dependent knowledge involved in the used methods. Until recently, state of the art SCF acquisition systems used handcrafted rules to generate hypothesis (Chrupala, 2003) or to match natural language parser output to a set of pre-defined SCFs (Briscoe and Carroll, 1997; Korhonen, 2002; Preiss et al., 2007, Esteve Ferrer, 2004). More recent works, however, propose to use an inductive approach, in which the inventory of SCFs is also induced directly from parsed corpus data (O'Donovan et al., 2004; Cesley and Salmon-Alt, 2006; Ienco et al., 2008; Lenci et al., 2008; Kawahara and Kurohashi, 2011). In Messiant (2008) inductive system, that we used as explained in section 3, candidate frames are identified by grammatical relation (GR) co-occurrences. The only given information is the label of GRs that are to be considered. Statistical filtering or empirically-tuned thresholds are again used to select frames for the final lexicon. This inductive approach has achieved respectable accuracy for different languages (0.60-0.70 F1-measure against a dictionary), do not involve predefined expert knowledge and is more portable than earlier methods. They are also highly scalable to large volumes of data, since the identification and selection of frames for producing the lexicon generally takes minimal time and resources.

The application of the inductive method is dependent, however, on the availability of a parser. The IULA Spanish Treebank (Marimon et al. 2012) allowed us to train two different statistical parsers: Malt (Nivre and Hall, 2005; Nivre et al., 2007) and one of the parsers in Mate-tools (Bohnet, 2010). These parsers were used to obtain the syntactic information to test the Messiant (2008) inductive method.

In the next sections we present the results of using the inductive approach for Spanish verbs. The only comparable exercise for Spanish is Altamirano and Alonso (2010). They developed a SCF extraction system based on SCF

induction and frequency based selection. The system was tested using the SENSEM corpus (Castellón et al., 2006). The corpus contained 100 sentences for each of the 250 most frequent Spanish verbs. For the SCF induction experiment, sentences were manually annotated with GR information, what makes the experiment similar to ours. Note, however, that in our experiments automatic parsing introduced errors that affected the induction results, as we will discuss in section 5.1. The Altamirano and Alonso (2010) experiment evaluation was carried out by manually inspecting the results for the 20 most frequent verb senses. Results obtained were: 0.79 precision and 0.70 recall.

3 Methodology

For the SCF induction, we used a Messiant (2008) based SCF extractor as implemented in the *tcp_subcat_inductive* web service developed by University of Cambridge³. The input to the web service is the output of a parser either in RASP parser format or in the CoNLL format. The user can decide which GR labels are candidates to be arguments of a verb, and hence part of subcategorization frames, and which are not and should not be considered. Note that the user does not define specific combination patterns, as in earlier SCF acquisition approaches: if the user specifies DOBJ and XCOMP as GR labels of interest, but not MODIFIER, then the SCF inventory will consist of all observed combinations of DOBJ and XCOMP, while MODIFIER will never appear in any SCF.

The system outputs the observed frequency of combinations of the addressed GRs for each verb as potential SCFs, what allows filtering them by their frequency. An adequate filtering threshold is tuned heuristically, as we will see later. The concrete information we extracted in our experiments was:

- Subject and verb complements: Direct Object (DOBJ), Indirect Object (IO), predicative and object-predicative complements and prepositional phrase complements (PP): bounded preposition, direction and location PPs.
- For subject and complements, we also considered whether the complement is

³Available at <http://registry.elda.org/services/304>

realised by a noun phrase, or a clause phrase.

- For PP complements with bounded preposition we also extract the particular preposition.

All this information is extracted by the *tcp_subcat_inductive* tool using the adequate parametrization.

The experiments were carried out on two domain specific corpora: Environment (46.2M tokens) and Labour Legislation (53.9M tokens). The corpora were automatically crawled and cleaned (Bel et al. 2012). From these corpora, all sentences containing the target lemmas were extracted and parsed. For each corpus, 30 target verbs were selected and a gold-standard was manually annotated for evaluation purposes (up to 200 sentences were annotated for each verb). The gold-standard in the form of a lexicon of possible frames associated to each verb type was derived from the actual occurrences of target verb types. Because of the restriction of having a minimum number of occurrences, the final list of verb types differs for the two corpora.

The evaluation was made in terms of number of acquired SCFs that were indeed SCFs in the gold standard per verb type (precision) and SCFs acquired with respect to the number of SCFs for every verb type in the gold standard (recall).

4 Experiments

As already mentioned, the goal of the experiments was to assess the induction method for Spanish using an automatic chain of processing tools, in particular dependency parsing. In our experiments, we tried two different dependency parsers and we experimented with different filtering thresholds as well as with other filtering approaches (ensemble and pattern-based filtering) in order to raise precision and to guarantee a clear-cut between reliable SCF assignments and dubious ones that would still need human revision. In what follows we present the configuration of the different experiments performed.

4.1 Different parsers

The SCF extractor can be applied to any parsed corpus. Thus, we used two different parsers to produce the input of the SCF extractor. The parsers were: (i) Malt parser (Nivre and Hall, 2005; Nivre et al, 2007) optimized with

MaltOptimizer (Ballesteros and Nivre, 2012). (ii) Mate graph-based re-scoring (completion model) parser (Bohnet and Kuhn, 2012; Bohnet and Nivre, 2012). Both parsers were trained with the IULA Treebank⁴ (Marimon et al. 2012). The parsers in turn were applied to PoS tagged text obtained with FreeLing v3 tools (Padró and Stanilovsky, 2012). Both parses had a high performance in terms of Labelled Attachment Score (LAS), being Mate the parser with higher LAS (94,7% vs 93,2% for Malt, with a test set from the IULA Treebank). However the exact match score, i.e. every complement in the parsed sentence is correctly analysed, was around 50%. Note that SCF extraction identifies frequent GR combinations in whole sentences, and therefore is very much affected when the parser repeatedly delivers combinations of correct but also wrong GR. In order to sort out this problem we tried two different strategies: combining the results of two parsers and filtering known bad combinations.

4.2 Ensemble Strategy

Given that we had data from two different parsers, we tried to raise SCF extraction precision by selecting as good ones only those SCF-verb assignments that resulted from considering the data from the two parsers in the extraction phase. The hypothesis behind was that if a particular SCF is not output by two systems, using each a different parser, it is unlikely that the GR combination is correct.

4.3 Filter Strategy

In order to assess the frequency filtering with respect to precision, we cleaned parser results by applying hand-made filters for erasing known parser frequent errors, for instance SCFs with more than one subject or direct object, with both a *by_agent* and a direct object, and so on. This strategy was only applied to Malt parser results to assess the benefits of this strategy that, note, requires expert human knowledge.

5 Results

In this section we present the obtained results for the Labour Legislation (LAB) and the Environment (ENV) corpora.

⁴ http://www.iula.upf.edu/recurs01_tbk_uk.htm

Table 1 and 2 show the summary of the obtained results with the experimental settings presented in the previous section for both corpora. Since we are especially interested in developing systems with high precision, in that table we present the best results maximizing F1 and maximizing precision. The table shows Precision, Recall and F1 values averaged for all verbs and the likelihood-based cut off to get these results. We also present results with the additional filters devised for known errors delivered by the Malt parser to assess their impact (Malt+F).

| method | thresh | P | R | F1 |
|-------------------------|--------|---------------|--------|---------------|
| Choosing best F1 | | | | |
| Malt | 0.04 | 0.6923 | 0.5094 | 0.5870 |
| Malt + F | 0.04 | 0.8571 | 0.5094 | 0.6391 |
| Mate | 0.04 | 0.6848 | 0.5943 | 0.6364 |
| Ensemble | 0.04 | 0.7195 | 0.5566 | 0.6277 |
| Choosing best P | | | | |
| Malt | 0.1 | 0.8723 | 0.3868 | 0.5359 |
| Malt + F | 0.09 | 0.9167 | 0.4151 | 0.5714 |
| Mate | 0.1 | 0.8333 | 0.4245 | 0.5625 |
| Ensemble | 0.09 | 0.8800 | 0.4151 | 0.5641 |

Table 1: Best results over LAB corpus

| method | thresh | P | R | F1 |
|-------------------------|--------|---------------|--------|---------------|
| Choosing best F1 | | | | |
| Malt | 0.05 | 0.8421 | 0.5053 | 0.6316 |
| Malt + F | 0.04 | 0.9074 | 0.5158 | 0.6577 |
| Mate | 0.06 | 0.8947 | 0.5368 | 0.6711 |
| Ensemble | 0.05 | 0.8548 | 0.5579 | 0.6752 |
| Choosing best P | | | | |
| Malt | 0.1 | 0.9545 | 0.4421 | 0.6043 |
| Malt + F | 0.07 | 0.9778 | 0.4632 | 0.6286 |
| Mate | 0.1 | 0.9375 | 0.4737 | 0.6294 |
| Ensemble | 0.08 | 0.9787 | 0.4842 | 0.6479 |

Table 2: Best results over ENV corpus

Additionally, Figure 1 shows details for all the results using Malt over the LAB corpus. In this figure we can see how increasing the filtering threshold leads to a better precision but to a loss in recall. Note that the frequency cut-off maximizing precision separates those assignments that are almost certain and would not need further revision. Table 3 shows the best results in terms of F1 obtained (Mate parser and ENV corpus) related to the number

of extracted SCFs. Note that our gold-standard has 32 possible SCFs.

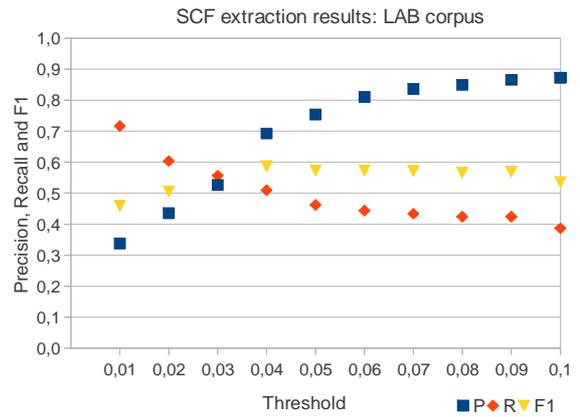


Figure 1: SCF results with Malt and LAB corpus

| Thres. | # SCF | P |
|--------|-------|--------|
| 0.01 | 24 | 0.5462 |
| 0.02 | 22 | 0.6162 |
| 0.03 | 18 | 0.6867 |
| 0.04 | 16 | 0.7794 |
| 0.05 | 14 | 0.8548 |
| 0.06 | 11 | 0.9245 |
| 0.07 | 10 | 0.9583 |
| 0.08 | 10 | 0.9787 |
| 0.09 | 10 | 0.9778 |
| 0.1 | 10 | 0.9778 |

Table 3: Number of SCF extracted with Mate and ENV corpus, and precision results

5.1 Discussion

Figure 2 shows graphically the results of every experimental set up. Malt and Mate parser results influenced the performance of the extractor. The results show that all strategies can achieve good precision scores, but with a dramatic cost in recall, as it was expected: infrequent SCFs are left out.

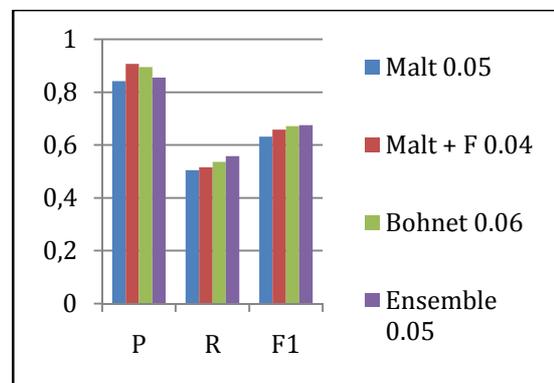


Figure 2: Comparison of Results for best F1 dataset in terms of Precision, Recall and F1 for ENV corpus.

The low recall for all systems is also due to the poor performance of both parsers for identifying some particular dependencies. The clearest example is the difficulties both parsers have in identifying indirect objects (IO). For example, Mate parser annotates IOs with 68% Precision and 52% Recall. Malt parser obtains poorer results. Those figures make the parser output hardly reliable for this low frequent complement (Padró et al., 2013).

The parser limitation to correctly detect IOs is also the reason for the differences between the two corpora results. The evaluation gave quite surprisingly different results for the two corpora, ENV corpus delivering overall better results. Manual inspection demonstrated that the difference came from the particular verbs chosen for the gold standard. In the ENV gold standard only 9 SCFs include IO, while in the LAB gold standard there were 18 SCFs. The parsers rarely deliver parses with this type of complements because they systematically assign a wrong label, therefore the SCF extractor never proposes candidates with IO although they are present at the gold standard. Therefore, in the LAB test set the results show a lower recall because accidentally there were more IOs to be found.

It is clear that using Mate parser to annotate the corpus, the system obtains just a slightly lower precision than the combined Malt and hand-made filter strategy (Malt+F), but better recall and F1 even although with a higher frequency cut-off. This means that using Mate parser we got competitive results without the need of developing hand-made rules, thus, resulting in a more general approach.

The ensemble strategy delivered poorer results in terms of precision gain, but it had better recall scores. This is partially due to the fact that better precision results are obtained with a lower threshold and therefore, more candidates are taken into account. We also found interesting that also for the same threshold, in some cases, an improvement on recall is observed. This is due to the fact that when combining SCFs extracted with both parsers the frequencies associated to each SCF change, making possible that some SCFs that were filtered with a given threshold for one system, are not filtered after the ensemble. The case of "presentar" ('to present') in ENV corpus when using a threshold of 0.03 is an example. In the gold standard this verb has three assigned SCF: transitive verb with noun phrase (both

direct object and subject are NP), intransitive verb (subject as NP) and ditransitive verb (subject and DOBJ are NP, and a further IO). Ditransitive SCF has a frequency in the gold standard of 0.02, and it is not learned by any of the systems (as said, IOs are badly tagged by the parsers, so it is very hard to learn SCFs that contain them). Transitive and intransitive are acquired by both Malt and Mate parsers, but with Malt parser data the extractor assigns to the intransitive frame a frequency of 0.02 and is thus not selected with a threshold of 0.03. On the other hand, with Mate parser data the system do extract the intransitive SCF, but with a frequency of 0.11 (closer to the gold standard frequency, which is 0.16). Thus, after the ensemble, the intransitive frame receives a frequency of 0.07 and is thus selected with 0.03 threshold. Note that we performed the ensemble before the filtering and apply the thresholds to the obtained results, in order to be able to capture these changes in frequency.

Comparing our strategy with previous experiments, we see that, although it is impossible to compare scores, the most noticeable fact is that the use of a dependency parser leads to competitive results in terms of precision, although with a poorer recall. However, this strategy requires less previous specific knowledge and manual work.

Furthermore, the number of SCFs in our gold-standards is bigger than those of previous work (32 different SCFs vs. 23 and 11 of Chrupala, 2003 and Esteve, 2004 respectively). This is also a further factor for assessing the low recall scores we got. In fact we are learning a similar number of SCFs than previous work, as shown in Table 3, but since our gold standard is more fine-grained, the resulting recall is lower.

6 Conclusions and future work

In this work we tested a SCF acquisition method for Spanish verbs. The used system extracts the SCFs automatically from dependency-parsed corpora building the SCF inventory at the same time as the lexicon.

We have seen that parser errors severely affect the SCF extraction results. Even though we are using state of the art parsers with very competitive performance, the systematic errors they produce for some infrequent complements make it impossible the identification of SCFs that contain such complements and thus causes

low recall. Nevertheless, we have obtained a system with good precision, though the recall needs still to be improved.

In order to improve the results, some future work will be necessary to improve parser output. One possible line is to filter out unreliably parsed sentences before running the SCF acquisition system, for example, performing the ensemble of the two parsers at sentence level instead of applying it to the output of the SCF extractor. Nevertheless, we do not expect that to solve the problem of undetected complements, such as IOs. Specific improvements in parsing will, therefore, be needed. In that line, we would encourage the parsing community to start considering other ways of evaluating parser accuracy apart from LAS. We have seen in our work that parsers with very high LAS may fail to label very important but infrequent complements that are needed for subsequent tasks.

7 Acknowledgments

This work was funded by the European Project PANACEA (FP7-ICT-2010- 248064). We specially thank Laura Rimell, Prokopis Prokopidis and Vassilis Papavasiliou, PANACEA partners, for their support. Our gratitude also to Miguel Ballesteros, Héctor Martínez and Bernd Bohnet for their kind support for MALT-optimizer and MATE tools.

8 References

- Altamirano, R. and Alonso, L. (2010) IRASubcat, a highly customizable, language independent tool for the acquisition of verbal subcategorization information from corpus. Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas.
- Ballesteros, M. and Nivre, J. (2012). MaltOptimizer: A System for MaltParser Optimization. In Proceedings of the International Conference on Language Resources and Evaluation, LREC'12.
- Bel, N., Papavasiliou, V., Prokopidis, P., Toral, A., Arranz, V. (2012). "Mining and Exploiting Domain-Specific Corpora in the PANACEA Platform", in The 5th Workshop on Building and Using Comparable Corpora. LREC'12.
- Bohnet, B. (2010). Top Accuracy and Fast Dependency Parsing is not a Contradiction. Proceedings of 23rd International Conference on Computational Linguistics (COLING 2010).
- Bohnet, B. and Kuhn, J (2012). The best of both worlds: a graph-based completion model for transition-based parsers. In Proceedings of the EACL 2012.
- Bohnet, B. and Nivre, J. (2012). A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In Proceedings of the EMNLP-CoNLL 2012.
- Briscoe, E. J. and Carroll, J. (1997). Automatic extraction of subcategorization from corpora. In Proceedings of the 5th ACL Conference on Applied Natural Language Processing,
- Castellón, I., A. Fernández, G. Vázquez, L. Alonso, J.A. Capilla (2006). The Sensem Corpus: a Corpus Annotated at the Syntactic and Semantic Level. In Proceedings of the International Conference on Language Resources and Evaluation, LREC'06.
- Chesley, P. and Salmon-Alt, S. (2006) Automatic extraction of subcategorization frames for French. In Proceedings of the International Conference on Language Resources and Evaluation, LREC'06.
- Chrupala, G. (2003). Acquiring Verb Subcategorization from Spanish Corpora. DEA Thesis, University of Barcelona.
- Esteve-Ferrer, E. (2004). Towards a semantic classification of Spanish verbs based on subcategorisation information. In Proceedings of the ACL 2004 Workshop on Student Research.
- Fernández, A., P. Saint-Dizier, G. Vázquez, F. Benamara, M. Kamel (2002). "The VOLEM Project: a Framework for the Construction of Advanced Multilingual Lexicons", in Proceedings of the Language Engineering Conference.
- Hajič, J; M. Čmejrek; B. Dorr; Y. Ding; J. Eisner; D. Gildea; T. Koo; K. Parton; G. Penn; D. Radev; and O. Rambow (2002). Natural language generation in the context of machine translation. Technical report. Center for Language and Speech Processing,

- Johns Hopkins University, Baltimore. Summer Workshop Final Report.
- Ienco, D.; S. Villata and C. Bosco (2008). Automatic extraction of subcategorization frames for Italian. In Proceedings of the International Conference on Language Resources and Evaluation, LREC'08.
- Kawahara, D and Kurohashi, S. (2010): Acquiring Reliable Predicate-argument Structures from Raw Corpora for Case Frame Compilation, In Proceedings of the International Conference on Language Resources and Evaluation, LREC'10.
- Korhonen, A. (2002). Subcategorization acquisition. Ph.D. thesis, University of Cambridge Computer Laboratory.
- Korhonen, A. (2010). Automatic Lexical Classification - Bridging Research and Practice. In *Philosophical Transactions of the Royal Society*. 368: 3621-3632.
- Korhonen A. and Y. Krymolowski (2002). On the Robustness of Entropy-Based Similarity Measures in Evaluation of Subcategorization Acquisition Systems. In Proceedings of the Sixth CoNLL.
- Lapata, M. F. Keller, and S. Schulte im Walde (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, 30(4):419-435.
- Lenci R, McGillivray B, Montemagni S, Pirrelli V. (2008) Unsupervised acquisition of verb subcategorization frames from shallow-parsed corpora. In Proceedings of the International Conference on Language Resources and Evaluation, LREC'08.
- Marimon, M., Fisas, B., Bel, N., Arias, B., Vázquez, S., Vivaldi, J., Torner, S., Villegas, M. and Lorente, M. (2012). The IULA Treebank. In Proceedings of the International Conference on Language Resources and Evaluation, LREC'12.
- McCarthy, D. and J. Carroll (2003). Disambiguating nouns, verbs and adjectives using automatically acquired selectional preferences. *Computational Linguistics* 29:4.
- Messiant, C. (2008). A subcategorization acquisition system for french verbs. In Proceedings of the ACL2008 (Student Research Workshop).
- Nivre, J and Hall, J. (2005). MaltParser: A language-independent system for data-driven dependency parsing. In Proceedings of the 4th Workshop on Treebanks and Linguistic Theories (TLT).
- Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., Marinov, S. and Marsi, E. (2007). Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13:95–135.
- O'Donovan R, Burke M, Cahill A, van Genabith J, Way A. (2004) Large-scale induction and evaluation of lexical resources from the penn-ii treebank. In Proceedings of ACL 2004.
- Padró, LL. and Stanilovsky, E. (2012); FreeLing 3.0: Towards Wider Multilinguality. In Proceedings of the International Conference on Language Resources and Evaluation, LREC'12.
- Padró, M., Ballesteros, M., Martínez, H. and Bohnet, B. (2013). Finding dependency parsing limits over a large Spanish corpus. In Proceedings of IJCNLP 2013.
- Preiss, J., Briscoe, E. J. and A. Korhonen. (2007). A System for Large-scale Acquisition of Verbal, Nominal and Adjectival Subcategorization Frames from Corpora. In Proceedings of ACL 2007.
- Surdeanu, M.; Harabagiu, S.; Williams, J. and Aarseth, S. (2003). Using predicate-argument structures for information extraction. In Proceedings of ACL 2003.
- Schulte im Walde, S. and C. Brew (2002). Inducing German semantic verb classes from purely syntactic subcategorisation information. In Proceedings of ACL 2002.
- Sun, L.; A. Korhonen and Y. Krymolowski. (2008a), Automatic Classification of English Verbs Using Rich Syntactic Features. In Proceedings of the Third International Joint Conference on Natural Language Processing (IJCNLP 2008).
- Sun, L.; A. Korhonen and Y. Krymolowski. (2008b). Verb Class Discovery from Rich Syntactic Data. Ninth International Conference on Computational Linguistics and Intelligent Text Processing (CICLING 2008).