First approach toward Semantic Role Labeling for Basque

Haritz Salaberri, Olatz Arregi, Beñat Zapirain
IXA Group, University of the Basque Country (UPV-EHU)
Manuel Lardizabal pasealekua, 1 - 20018 Donostia-San Sebastiáñ
{haritz.salaverri, olatz.arregi, benat.zapirain}@ehu.es

Abstract
In this paper, we present the first Semantic Role Labeling system developed for Basque. The system is implemented using machine learning techniques and trained with the Reference Corpus for the Processing of Basque (EPEC). In our experiments the classifier that offers the best results is based on Support Vector Machines. Our system achieves 84.30 F1 score in identifying the PropBank semantic role for a given constituent and 82.90 F1 score in identifying the VerbNet role. Our study establishes a baseline for Basque SRL. Although there are no directly comparable systems for English we can state that the results we have achieved are quite good. In addition, we have performed a Leave-One-Out feature selection procedure in order to establish which features are the worthiest regarding argument classification. This will help smooth the way for future stages of Basque SRL and will help draw some of the guidelines of our research.

Keywords: Semantic Role Labeling, PropBank/VerbNet, Basque

1. Introduction
The main task of semantic role labeling (SRL), sometimes also called shallow semantic parsing, is to detect the semantic relations held between the predicate of a sentence and its associated participants and properties as well as their classification into specific roles. Predicates can be of two types: nominal or verbal. Our work focuses on verbal predicates. Annotating text with semantic roles can help determine who did what to whom, where, when, and how within the events described in the text. As is stated in (Márquez et al., 2008) the predicate of a clause (a verb in our case) establishes what took place, and other sentence constituents express the participants in the event (such as who and where), as well as further event properties (such as when and how).
The developed system labels the predicate arguments with PropBank (Kingsbury and Palmer, 2003) and VerbNet (Schuler, 2005) role sets and it is able to label corpora on a large scale. The annotation of semantic roles is important for the development of advanced tools and applications such as machine translation (Boas, 2002), question answering (Shen and Lapata, 2007) and text summarization (Melli et al., 2005); therefore, it can be concluded that the developed system fulfills the need for automatically annotating semantic roles within large Basque corpora.
Regarding the type of syntactic information used for learning, we distinguish two types of semantic role labeling: the dependency-based and the constituent-based. As is stated in (Surdeanu et al., 2008) dependency syntax (on which dependency-based SRL relies) represents grammatical structures by means of labeled binary head-dependent relations rather than phrases; therefore, head-dependent pairs are identified and labeled when annotating syntax with dependencies. In constituent-based syntax (on which constituent-based SRL relies), on the other hand, grammatical structures are represented by means of phrases (Carreras and Márquez, 2005). In the EPEC corpus used in our study the syntactic dependencies are annotated following the dependency-based formalism used in the Prague Dependency Treebank corpus (Hajic, 1998).

2. EPEC Corpus
EPEC (Aduriz et al., 2006) is a 127,000 word (+10,000 sentence) sample collection of written Standard Basque. It is a strategic resource for the processing of Basque and it has already been used for the development and improvement of several tools (Aldabe et al., 2013). Half of this collection was obtained from Euskaldunon Egunkaria², the only daily newspaper written entirely in Basque.
Syntactic in EPEC is annotated following the dependency-based formalism used in the Prague Dependency Treebank, which was also used in the German NEGRA corpus (Skut et al., 1997). This formalism was chosen over the constituent-based formalism used in the English Penn Treebank corpus (Marcus et al., 1993) due to the good adaptability it offers regarding the free word order displayed by Basque syntax. In addition, annotating syntax by using dependency relations implies a model strongly based on hierarchy where linear order plays a secondary role and gives the possibility to use functional information.

2.1. Semantic roles in EPEC and differences regarding PropBank
The number of predicate arguments that have been identified in the corpus is 54,500. From these arguments, 35,500 have been manually annotated with semantic roles. These 35,500 arguments have been used as a training set for the SRL system developed in this study.
We have analyzed how often PropBank argument instances are mapped to VerbNet roles in both PropBank and EPEC. This way we will be able to better understand the results obtained by the system we have developed and to somehow compare our results to the ones for English SRL systems that use PropBank (the Wall Street Journal corpus) as a training corpus.

1http://www.euskaracorpusa.net
2Now called Berria: http://www.berria.info/
As (Loper et al., 2007) states, when PropBank was created an explicit effort was made to use A0 for arguments that fulfill Dowty’s criteria for "prototypical agent" and A1 for arguments that fulfill the criteria for "prototypical patient". As a result, these two argument labels are significantly more consistent across verbs than A2, A3 and A4 (as shown in table 1). (Loper et al., 2007) also states that despite this effort there still are some inter-verb inconsistencies for even A0 and A1. These inter-verb inconsistencies are clearly visible in table 1: in PropBank 2% of A0 argument instances are mapped into the VerbNet role theme and 47% of theme roles are mapped into A1. In EPEC, on the other hand (this is where our training corpus differs from PropBank), 18% of Theme roles are mapped into A0 and 52% are mapped into A1, thus we can clearly state that the inter-verb inconsistency between A0 and A1 is much bigger for Basque EPEC than for English PropBank. The reason why the verb inconsistency between A0 and A1 is so big in EPEC lies in the fact that as opposed to PropBank, when the corpus was created, no effort was made to maintain A0 as a "prototypical agent" and A1 as a "prototypical patient", instead, these arguments where randomly assigned across the verbs in EPEC. This difference in verb inconsistency will clearly be an important factor for correctly interpreting the results obtained by our EPEC based SRL system and to be able to know where we stand exactly regarding state-of-the-art systems.

3. Data Format

The data format used in our experiments is intended to follow the column-based format from the Conll08 4 shared task (closed-track) (Surdeanu et al., 2008), as we understand this can be thought of as a standard for SRL related tasks. There are some minor differences though: We use additional linguistic information such as name entity and declension case information that was not provided in the original shared task. This is the reason why, as can be seen by the example sentence Argentinara joan zen taldea egongo da Pau Orthezen kontra. (The team that went to Argentina will play against Pau Orthez) shown in figure 1, additional columns were considered. The explanation to use declension case as a feature lies on the fact that as opposed to English, Basque is a morphologically rich (agglutinative) language and declension case offers very meaningful information. Nevertheless, minor differences apart, general rules followed by data in Conll08 prevail in the data used in our experiments. The general rules are the following:

- The files contain sentences separated by a blank line.
- A sentence consists of one or more tokens and the information for each token is represented on a separate line.
- A token consists of at least 11 fields. The fields are separated by one or more whitespace characters (spaces or tabs). Whitespace characters are not allowed within fields.

For explanatory reasons the columns in figure 1 have been labeled from C1 to C15. The information hold by each column is:

- C1: Token counter, starting at 1 for each new sentence.
- C2: Unsplit word form or punctuation symbol.
- C3: Predicted lemma of C2.
- C4: PoS tag from the Treebank.
- C5: PoS subcategory tag.
- C6: Declension case tag.
- C7: Name entity tag.
- C8: Number entity tag.
- C9: Syntactic head of the current token, which is either a value of C1 or 0.
- C10: Syntactic dependency relation to C9. The relations considered are: ncsbj, ncobj, nczobj, nc-mod, ncpred (non-clausal subject, object, indirect object, ...), ccomp_obj, ccomp_subj, cmod (clausal finite object, subject, modifier), xcomp_obj, xcomp_subj, xcomp_zobj, xmod, xpred (clausal non-finite object, subject, indirect object, ...).
- C11: Rolesets of the semantic predicates in the sentence.
- C12-C15: Columns with argument labels for each semantic predicate following textual order. PropBank and VerbNet argument labels for each predicate.

As (Loper et al., 2007) states, when PropBank was created an explicit effort was made to use A0 for arguments that fulfill Dowty’s criteria for "prototypical agent" and A1 for arguments that fulfill the criteria for "prototypical patient". As a result, these two argument labels are significantly more consistent across verbs than A2, A3 and A4 (as shown in table 1). (Loper et al., 2007) also states that despite this effort there still are some inter-verb inconsistencies for even A0 and A1. These inter-verb inconsistencies are clearly visible in table 1: in PropBank 2% of A0 argument instances are mapped into the VerbNet role theme and 47% of theme roles are mapped into A1. In EPEC, on the other hand (this is where our training corpus differs from PropBank), 18% of Theme roles are mapped into A0 and 52% are mapped into A1, thus we can clearly state that the inter-verb inconsistency between A0 and A1 is much bigger for Basque EPEC than for English PropBank. The reason why the verb inconsistency between A0 and A1 is so big in EPEC lies in the fact that as opposed to PropBank, when the corpus was created, no effort was made to maintain A0 as a "prototypical agent" and A1 as a "prototypical patient", instead, these arguments where randomly assigned across the verbs in EPEC. This difference in verb inconsistency will clearly be an important factor for correctly interpreting the results obtained by our EPEC based SRL system and to be able to know where we stand exactly regarding state-of-the-art systems.

### Table 1: Percentages indicating how often PropBank argument instances are mapped to VerbNet roles.

|        | A0 % | A1 % | A2 % | A3 % | A4 % |
|--------|------|------|------|------|------|
| PropBank | Agent 85 | Theme 47 | Recipient 22 | Asset 33 | Location 89 |
|        | Theme 2 | Topic 23 | Extent 15 | Theme2 14 | Beneficiary 5 |
|        | Experiencer 7 | Patient 11 | Predicate 14 | Recipient 13 | - |
| EPEC | Agent 77 | Theme 52 | Attribute 41 | Location 41 | Destination 41 |
|        | Theme 18 | Topic 22 | Destination 14 | Destination 20 | Location 30 |
|        | Topic 2 | Product 10 | Location 13 | Beneficiary 18 | Attribute 16 |

4http://barcelona.research.yahoo.net/dokuwiki/
4. Development scheme

Typically, the role labeling task consists of identifying the constituents of each target predicate (argument identification) and labeling them with semantic roles (argument classification). Nevertheless, in order to identify and then classify these arguments, SRL systems first have to identify the target predicate (predicate identification) and then assign a certain sense number to it (predicate classification) (Che et al., 2008). In these early stages of Basque role labeling we focus just on the argument identification and classification task by making use of the manually identified and classified verbs in the EPEC corpus.

4.1. Argument identification

In our semantic role labeler, the argument identification process is performed automatically by following a systematic processing strategy over the set of dependency relations annotated in the corpus. This strategy is described in (Surdeanu et al., 2008) and basically consists of identifying as the head of a semantic argument the token inside the argument boundaries whose head is a token outside the argument boundaries. Figure 2 shows the syntactic dependencies for the example sentence from figure 1.

![Figure 2: Syntactic dependencies marked in the example sentence.](image)

4.2. Argument classification

The semantic layer in terms of semantic roles from EPEC follows both the PropBank and the VerbNet models, meaning that the arguments of each predicate have been manually annotated with both the PropBank and VerbNet tags, as can be seen in figure 1. This gave us the opportunity to build two classifiers: one classifier to label the arguments with PropBank tags and another classifier to label the arguments with VerbNet tags.

In order to estimate the performance of our system (for both classifiers), we carried out 10-fold cross-validation over the training set. The performance of the system has been tested using several learning algorithms such as Support Vector Machines (SVM), decision trees and random decision trees. These algorithms are well known to have a good performance regarding NLP tasks.

4.2.1. Features

We have considered several typical features in order to train both the PropBank and VerbNet classifiers.

- **Predicate lemma**: Lemma for the proposition predicate.
- **Argument lemma**: Lemma for the argument head.
- **Argument PoS**: Part-of-Speech category for the argument head.
- **Argument PoS subcategory**: Part-of-Speech subcategory for the argument head.
- **Declension case**: Declension case for the argument.
- **Syntactic function**: Syntactic function for the argument.
• **Argument position**: Position of the argument according to the predicate.
• **Distance in words**: Distance in number of words between the argument and the predicate.
• **Distance in arguments**: Distance in number of arguments between the argument and the predicate.
• **Frame**: Predicate-argument structure for the proposition.
• **Syntactic frame**: Argument position inside the frame (Xue and Palmer, 2004).
• **Name entity**: Entity of the argument (if any). It can be Organization, Place or Person.
• **Number entity**: Number entity of the argument (if the argument is a numerical value). It can be Date, Price etc.

These features have been widely used in machine learning-based role labeling since the foundational work (Gildea and Jurafsky, 2002). Some examples for English SRL include (Palmer et al., 2010) and (Carreras and Màrquez, 2005). The listed features have also been used in other languages such as Chinese (Xue and Palmer, 2005) and Swedish (Johansson et al., 2012).

5. Results

In order to evaluate the performance of the argument classification process we used standard precision, recall and f₁ measures. The overall results achieved when classifying the arguments with PropBank and VerbNet roles are shown in tables 2 and 3 respectively.

| PropBank | SVM | P   | R   | F₁  |
|----------|-----|-----|-----|-----|
|          |     | 84.30 | 84.60 | 84.30 |
|          | DT  | 84.00 | 83.20 | 83.90 |
|          | RDT | 77.40 | 78.30 | 77.70 |

Table 2: PropBank SRL performance. (SVM: Support Vector Machines, DT: Decision Trees, RDT: Random Decision Trees)

| VerbNet | SVM | P   | R   | F₁  |
|---------|-----|-----|-----|-----|
|          |     | 83.10 | 83.10 | 82.90 |
|          | DT  | 81.70 | 81.80 | 81.50 |
|          | RDT | 72.20 | 72.90 | 72.10 |

Table 3: VerbNet SRL performance. (SVM: Support Vector Machines, DT: Decision Trees, RDT: Random Decision Trees)

As can be noticed in the result tables, the best performance when labeling arguments with PropBank role tags (84.30 F₁ score) is achieved by using a Support Vector Machines classifier. SVM also performs the best when labeling arguments with VerbNet role tags (82.90 F₁ score). The learning algorithm that gets, by far, the worst results for both PropBank and VerbNet is the Random Decision Trees algorithm.

In fact, when RDT is used for the VerbNet classifier results drop more than 10 absolute points. Table 4 shows the f₁ score achieved by both our classifiers for each role tag in their respective rolesets. These results have been achieved by the best-performing SVM algorithm.

|          | PropBank | VerbNet |
|----------|----------|---------|
| Arg0     | 95.00    | 89.70   |
| Arg1     | 93.70    | 96.20   |
| Arg2     | 81.60    | 91.60   |
| Arg3     | 57.90    | 79.20   |
| Arg4     | 15.40    | 10.50   |

Table 4: f₁ score for PropBank and VerbNet role labels with SVM classifiers.

The results table shows that PropBank core arguments (Arg0 to Arg4) are labeled with a f₁ score that progressively decreases from 95.00 to 15.40. Results for PropBank adjuncts, on the other hand, vary markedly depending on the type. Negation (NEG) and cause (CAU) adjuncts for instance are labeled with a 99.20 and a 80.50 f₁ score while the score for adverb (ADV) and dislocation (DIS) adjuncts is 50.80 and 41.60, respectively.

The results for VerbNet show that roles that are not adjuncts are labeled with a f₁ score that goes from 96.20 to 66.90, whereas all but three (Experiencer, Predicate and Source) have a score above 80. Regarding VerbNet adjuncts, the f₁ scores look a lot like the scores achieved for PropBank adjuncts.

6. Analysis

In order to analyze and to better understand the results obtained by our semantic role labeler, we have compared these results to the ones reported for CoNLL 2005 datasets by (Zapirain et al., 2008). Table 6 shows the results for both Basque and English. Our results have been obtained using
a SVM classifier; the results in (Zapirain et al., 2008), on the other hand, were obtained using a Maximum Entropy Markov Model.

| Without the feature | PropBank       | VerbNet       |
|---------------------|----------------|---------------|
|                      | P   | R   | F1  | P   | R   | F1  |
| Pred. lemma         | 78.30 | 77.50 | 77.10 | 67.40 | 68.20 | 66.10 |
| Argument lemma      | 79.90 | 80.40 | 79.90 | 78.70 | 79.00 | 78.50 |
| Argument PoS        | 84.20 | 84.50 | 84.20 | 83.00 | 83.00 | 82.80 |
| Argument subPoS     | 84.00 | 84.20 | 83.90 | 82.60 | 82.50 | 82.30 |
| Declension case     | 75.20 | 76.10 | 75.30 | 73.60 | 73.90 | 73.40 |
| Syntactic function  | 82.00 | 82.20 | 81.90 | 80.90 | 80.90 | 80.60 |
| Argument position   | 84.30 | 84.60 | 84.30 | 83.10 | 83.10 | 82.90 |
| Distance in words   | 84.30 | 84.60 | 84.30 | 83.10 | 83.10 | 82.90 |
| Distance in arguments| 84.30 | 84.50 | 84.30 | 83.10 | 83.10 | 82.90 |
| Frame               | 84.40 | 84.70 | 84.40 | 83.30 | 83.30 | 83.10 |
| Syntactic frame     | 84.50 | 84.60 | 84.30 | 83.40 | 83.40 | 83.10 |
| Name entity         | 84.30 | 84.60 | 84.30 | 83.20 | 83.20 | 83.00 |
| Number entity       | 84.40 | 84.60 | 84.30 | 83.10 | 83.10 | 82.90 |
| **ALL**             | 84.30 | 84.60 | 84.30 | 83.10 | 83.10 | 82.90 |

Table 5: Leave-One-Out results for PropBank and VerbNet

Table 6: English SRL Vs. Basque SRL

It can be noted at first glance that the results of our system are significantly higher than the results in (Zapirain et al., 2008). The reason for these high values is that, as we have previously stated in section 4, in our system the dependency parsing, predicate identification and classification subtasks needed in order to label arguments with semantic roles have been performed manually and not automatically as in (Zapirain et al., 2008). In addition, the semantic role labeling performed by our system uses dependency-based syntax and not constituent-based syntax as in (Zapirain et al., 2008). Performing argument identification is a much more complex task in constituent-based syntax. This is another very important factor to be taken into account when comparing both systems.

When comparing the results achieved for core arguments we have noticed that the results for Arg1 and Arg2 improve about 15 absolute points, while results for Arg0 and Arg4 improve to a lesser degree (5 and 6.5 points). One of the reasons (apart from the previously mentioned) why Arg2 improves 16 points is that, as shown in table 1, in EPEC, Arg2 argument instances are in 41% of the cases mapped into VerbNet Attribute instances while in PropBank Arg2 argument instances are much more sparsely mapped into VerbNet roles, most frequent mappings being Recipient (22%), Extent (15%) and Predicate (14%).

When comparing VerbNet roles, on the other hand, we have noticed that our results improve in a range of 5 to 15 points over the results reported for English. There are some exceptions though: Product for example improves 30 points, and Stimulus 24, Experiencer, on the other hand, does not improve but worsens in 20 absolute points.

Table 1 shows that the Experiencer role is mapped to Arg0 in PropBank, this being the third most frequent VerbNet role mapped to Arg0. In addition, there are no significant inter-verb inconsistencies for Experiencer in the English corpus. In EPEC, on the contrary, the number of Experiencer instances mapped to PropBank roles is significantly smaller taking into account that it does not appear in table 1. We presume there is a greater inter-verb inconsistency among the Experiencer role instances in Basque than in English and that is the reason why the results are 20 points worse.

Finally, when we compare the results for adjuncts in both languages, we notice that there are some minor differences. In general, we see that some adjuncts are labeled better in
Basque and some others are labeled better in English. The reasons for these differences lie on the nature of each language. For example, modal verbs are much easier to detect in English than in Basque.

6.1. Feature selection

Feature selection is a fundamental problem in many different areas where machine learning is used. As is stated in (Novakovic, 2009) feature selection shrinks the dimensionality of feature space and removes redundant, irrelevant, or noisy data. This, on the other hand, reduces the number of resources used (especially in terms of time) by the learning algorithm, improving the data quality and therefore the performance of the classifier.

In order to be able to perform a feature selection process in future stages of Basque SRL, we have determined the impact of each individual feature in the argument classification task. For this purpose, we have followed a Leave-One-Out (LOO) procedure over the training data for both PropBank and VerbNet train sets. This procedure evaluates the worth of each feature that has been initially considered by iteratively removing the information relative to that feature and by then training the classifier with the rest of features. Results corresponding to SVM based classifiers are shown in table 5.

As can be seen in table 5, there are some features that worsen the result for both PropBank and VerbNet. For the VerbNet classifier the features with a negative impact are the Frame (from 82.90 to 83.10), the Syntactic Frame (from 82.90 to 83.10) and the Name Entity (from 82.90 to 83.00). For the PropBank classifier, on the other hand, the only feature that produces a negative impact is the Frame feature (from 84.30 to 84.40). It may also be noted in table 5 that the worth of each feature, the importance regarding classification, varies from PropBank to VerbNet. The four worthiest features in PropBank, listed from the worthiest to the least worthy, are: The Declension case (from 84.30 to 75.30), the Predicate lemma (from 84.30 to 77.10), the Argument lemma (from 84.30 to 79.90) and the Syntactic function (from 84.30 to 81.90). The four worthiest features in VerbNet, on the other hand, are: the Predicate lemma (from 82.90 to 66.10), the Declension case (from 82.90 to 73.40), the Argument lemma (from 82.90 to 78.50) and the Syntactic function (from 82.90 to 80.60).

We have performed an experiment where we have removed, from the initial set of features, the ones that do not have a positive impact. Then we have trained both the PropBank and VerbNet classifiers with the features left. These features are: the Predicate lemma, the Argument lemma, the Argument PoS category, the Argument PoS subcategory, the Declension case and the Syntactic function. The results for SVM based classifiers are shown in table 7.

|          | P     | R     | F1    |
|----------|-------|-------|-------|
| PropBank | 84.20 | 84.30 | 84.00 |
| VerbNet  | 82.90 | 82.80 | 82.60 |

Table 7: Results for the feature selection experiment

As can be seen in table 7, although the removed features individually taken did not improve the results, the F₁ scores achieved without these features decrease in 0.3 points. Consequently, we can state that the combination of the removed features produces an improvement. Nevertheless, we can also state that the most valuable information, the worthiest information, regarding role labeling is gathered by the features that have not been removed from this experiment.

7. Conclusion

In this paper we have presented the first results on semantic role labeling for Basque using the Reference Corpus for the Processing of Basque (EPEC) and several machine learning methods such as Support Vector Machines and decision trees. We have achieved 84.30 F₁ score when labeling predicate arguments according to PropBank and a 82.90 F₁ score when labeling predicate arguments according to VerbNet. Our system establishes with these scores the baseline for Basque SRL.

Regarding the comparison we have performed, we are aware that both systems are hardly comparable to each other due to the great differences that lie between them. Nevertheless, we conclude that the results of our system are quite good for arguments that go from A0 to A5. Results for adjuncts on the other hand appear to follow some language-nature guided behavior; despite that, the average result for adjuncts is quite good as well. In addition, we have analyzed the impact of each individual feature regarding the argument classification subtask and came with the conclusion that removing some features can help reduce the processing time drastically with a F₁ score reduction of just 0.3 points.

8. Acknowledgements

This research has been supported by the University of the Basque Country (IXA Group, Research Group of type A (2010-2015))(IT34410)).

9. References

Itziar Aduriz, María Jesús Aranzabe, Jose Mari Arriola, Aitziber Atutxa, Diaz A de Ibarraza, Nerea Ezeiza, Koldo Gojenola, Maite Oronoz, Aitor Soroa, and Ruben Urizar. 2006. Methodology and steps towards the construction of epec, a corpus of written basque tagged at morphological and syntactic levels for automatic processing. *Language and Computers*, 56(1):1–15.

Itziar Aldabe, Itziar Gonzalez-Dios, Iñigo Lopez-Gazpio, Ion Madrazo, and Montse Maritxalar. 2013. Two approaches to generate questions in basque. *Procesamiento del Lenguaje Natural*, 51:101–108.

Hans Christian Boas. 2002. Bilingual framenet dictionaries for machine translation. In *LREC*.

Xavier Carreras and Lluís Màrquez. 2005. Introduction to the conll-2005 shared task: Semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 152–164. Association for Computational Linguistics.

Wanxiang Che, Zhenghua Li, Yuxuan Hu, Yongqiang Li, Bing Qin, Ting Liu, and Sheng Li. 2008. A cascaded syntactic and semantic dependency parsing system. In...
Proceedings of the Twelfth Conference on Computational Natural Language Learning, pages 238–242. Association for Computational Linguistics.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. Computational linguistics, 28(3):245–288.

Jan Hajic. 1998. Building a syntactically annotated corpus: The prague dependency treebank. Issues of valency and meaning, pages 106–132.

Richard Johansson, Karin Friberg Heppin, and Dimitrios Kokkinakis. 2012. Semantic role labeling with the swedish framenet. In LREC, pages 3697–3700.

Paul Kingsbury and Martha Palmer. 2003. Propbank: the next level of treebank. In Proceedings of Treebanks and lexical Theories, volume 3.

Edward Loper, Szu-Ting Yi, and Martha Palmer. 2007. Combining lexical resources: mapping between propbank and verbnet. In Proceedings of the 7th International Workshop on Computational Linguistics, Tilburg, the Netherlands.

Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. Computational linguistics, 19(2):313–330.

Lluís Márquez, Xavier Carreras, Kenneth C Litkowski, and Suzanne Stevenson. 2008. Semantic role labeling: an introduction to the special issue. Computational linguistics, 34(2):145–159.

Gabor Melli, Yang Wang, Yudong Liu, Mehdi M Kashani, Zhongmin Shi, Baohua Gu, Anoop Sarkar, and Fred Popowich. 2005. Description of squash, the sfu question answering summary handler for the duc-2005 summarization task. safety, 1:14345754.

Jasmina Novakovic. 2009. Using information gain attribute evaluation to classify sonar targets. In 17th Telecommunications forum TELFOR.

Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. Semantic role labeling. Synthesis Lectures on Human Language Technologies, 3(1):1–103.

Karin Kipper Schuler. 2005. Verbnet: A broad-coverage, comprehensive verb lexicon.

Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In EMNLP-CoNLL, pages 12–21.

Wojciech Skut, Brigitte Krenn, Thorsten Brants, and Hans Uszkoreit. 1997. An annotation scheme for free word order languages. In Proceedings of the fifth conference on Applied natural language processing, pages 88–95. Association for Computational Linguistics.

Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Márquez, and Joakim Nivre. 2008. The conll-2008 shared task on joint parsing of syntactic and semantic dependencies. In Proceedings of the Twelfth Conference on Computational Natural Language Learning, pages 159–177. Association for Computational Linguistics.

Nianwen Xue and Martha Palmer. 2004. Calibrating features for semantic role labeling. In EMNLP, pages 88–94.

Nianwen Xue and Martha Palmer. 2005. Automatic semantic role labeling for chinese verbs. In IJCAI, volume 5, pages 1160–1165. Citeseer.

Benat Zapirain, Eneko Agirre, and Lluís Márquez. 2008. Robustness and generalization of role sets: Propbank vs. verbnet. In ACL, pages 550–558.