Unsupervised Induction of Frame-Semantic Representations

Ashutosh Modi **Ivan Titov** **Alexandre Klementiev**
Saarland University
Saarbrücken, Germany
{amodi|titov|aklement}@mmci.uni-saarland.de

Abstract

The frame-semantic parsing task is challenging for supervised techniques, even for those few languages where relatively large amounts of labeled data are available. In this preliminary work, we consider *unsupervised* induction of frame-semantic representations. An existing state-of-the-art Bayesian model for PropBank-style unsupervised semantic role induction (Titov and Klementiev, 2012) is extended to jointly induce semantic frames and their roles. We evaluate the model performance both quantitatively and qualitatively by comparing the induced representation against FrameNet annotations.

1 Introduction

Shallow representations of meaning, and semantic role labels in particular, have a long history in linguistics (Fillmore, 1968). In this paper we focus on frame-semantic representations: a *semantic frame* is a conceptual structure describing a situation (or an entity) and its participants (or its properties). Participants and properties are associated with *semantic roles* (also called *frame elements*). For example, following the FrameNet annotation guidelines (Ruppenhofer et al., 2006), in the following sentences:

(a) [$_{COOK}$ Mary] cooks [$_{FOOD}$ the broccoli] [$_{CONTAINER}$ in a small pan].

(b) Sautee [$_{FOOD}$ the onions] [$_{MANNER}$ gently] [$_{TEMP\_SETTING}$ on low heat].

the same semantic frame $Apply\_Heat$ is evoked by verbs *cook* and *sautee*, and roles $COOK$ and $FOOD$ in the sentence (a) are filled by *Mary* and *the broccoli*, respectively. Note that roles are specific to the frame, not to the individual *lexical units* (verbs *cook* and *sautee*, in the example).[1]

Most approaches to predicting these representations, called *semantic role labeling* (SRL), have relied on large annotated datasets (Gildea and Jurafsky, 2002; Carreras and Màrquez, 2005; Surdeanu et al., 2008; Hajič et al., 2009). By far, most of this work has focused on PropBank-style representations (Palmer et al., 2005) where roles are defined for each individual verb, or even individual senses of a verb. The only exceptions are modifiers and roles $A0$ and $A1$ which correspond to proto-agent (a doer, or initiator of the action) and proto-patient (an affected entity), respectively. However, the SRL task is known to be especially hard for the FrameNet-style representations for a number of reasons, including, the lack of cross-frame correspondence for most roles, fine-grain definitions of roles and frames in FrameNet, and relatively small amounts of statistically representative data (Erk and Pado, 2006; Das et al., 2010; Palmer and Sporleder, 2010; Das and Smith, 2011). Another reason for reduced interest in predicting FrameNet representations is the lack of annotated resources for most languages, with annotated corpora available or being developed only for English (Ruppenhofer et al., 2006), German (Burchardt et al., 2006), Spanish (Subirats, 2009) and Japanese (Ohara et al., 2004).

Due to scarcity of labeled data, purely unsupervised set-ups recently started to receive considerable attention (Swier and Stevenson, 2004; Grenager and Manning, 2006; Lang and Lapata, 2010; Lang and

[1] More accurately, FrameNet distinguishes core and non-core roles with non-core roles mostly corresponding to modifiers, e.g., $MANNER$ in sentence (b). Non-core roles are expected to generalize across frames.

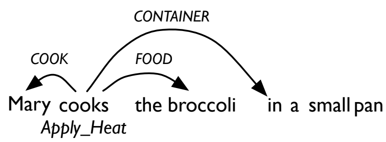

Figure 1: An example of a semantic dependency graph.

Lapata, 2011a; Lang and Lapata, 2011b; Titov and Klementiev, 2012). However, all these approaches have focused on PropBank-style representations. This may seem somewhat unnatural as FrameNet representations, though arguably more powerful, are harder to learn in the supervised setting, harder to annotate, and annotated data is available for a considerably fewer languages. This is the gap which we address in this preliminary study.

More specifically, we extend an existing state-of-the-art Bayesian model for unsupervised semantic role labeling and apply it to support FrameNet-style semantics. In other words, our method jointly induces both frames and frame-specific semantic roles. We experiment only with verbal predicates and evaluate the performance of the model with respect to some natural baselines. Though the scores for frame induction are not high, we argue that this is primarily due to very high granularity of FrameNet frames which is hard to reproduce for unsupervised systems, as the implicit supervision signal is not capable of providing these distinctions.

2 Task Definition

In this work, we use dependency representations of frame semantics. Dependency representations for SRL (Johansson and Nugues, 2008) were made popular by CoNLL-2008 and CoNLL-2009 shared tasks (Surdeanu et al., 2008; Hajič et al., 2009), but for English were limited to PropBank. Recently, English FrameNet was also released in the dependency format (Bauer et al., 2012). Instead of predicting argument spans, in dependency representation the goal is, roughly, to predict the syntactic head of the argument. The semantic dependency representation for sentence (a) is shown in Figure 1, labels on edges denote roles and labels on words denote frames. Note that in practice the structures can be more complex, as, for example, arguments can evoke their own frames or the same arguments can be shared by multiple predicates, as in right node raising constructions.

The SRL task, or more specifically frame-semantic parsing task consists, at least conceptually, of four stages: (1) identification of frame-evoking elements(FEE), (2) identification of arguments, (3) frame labeling and (4) role labeling. In this work, we focus only on the frame labeling and role labeling stages, relying on gold standard (i.e. the oracle) for FEEs and role identification. In other words, our goal is to label (or cluster) edges and nodes in the dependency graph, Figure 1. Since we focus in this study on verbal predicates only, the first stage would be trivial and the second stage could be handled with heuristics as in much of previous work on unsupervised SRL (Lang and Lapata, 2011a; Titov and Klementiev, 2012).

Additionally to considering only verbal predicates, we also assume that every verb belongs to a single frame. This assumption, though restrictive, may be reasonable in practice as (a) the distributions across frames (i.e. senses) are generally highly skewed, (b) current state-of-the-art techniques for word-sense induction hardly beat most-frequent-sense baselines in accuracy metrics (Manandhar et al., 2010). This assumption, or its minor relaxations, is relatively standard in work on unsupervised semantic parsing tasks (Poon and Domingos, 2009; Poon and Domingos, 2010; Titov and Klementiev, 2011). From the modeling prospective, there are no major obstacles to relaxing this assumption, but it would lead to a major explosion of the search space and, as a result, slow inference.

3 Model and Inference

We follow previous work on unsupervised semantic role labeling (Lang and Lapata, 2011a; Titov and Klementiev, 2012) and associate arguments with their frame specific syntactic signatures which we refer to as *argument keys*:

- Active or passive verb voice (`ACT/PASS`).
- Argument position relative to predicate (`LEFT/RIGHT`).
- Syntactic relation to its governor.
- Preposition used for argument realization.

Semantic roles are then represented as clusters of argument keys instead of individual argument occurrences. This representation aids our models in inducing high purity clusters (of argument keys) while reducing their granularity. Thus, if an argument key k is assigned to a role r ($k \in r$), all of its occurrences are labeled r.

3.1 A model for frame-semantic parsing

Our approach is similar to the models of Titov and Klementiev (2012; 2011). Please, see Section 5 for a discussion of the differences.

Our model encodes three assumptions about frames and semantic roles. First, we assume that the distribution of lexical units (verbal predicates) is sparse for each semantic frame. Second, we enforce the selectional restriction assumption: we assume that the distribution over potential argument fillers is sparse for every role, implying that 'peaky' distributions of arguments for each role r are preferred to flat distributions. Third, each role normally appears at most once per predicate occurrence. Our inference will search for a frame and role clustering which meets the above requirements to the maximal extent.

Our model associates three distributions with each frame. The first one (ϕ) models the selection of lexical units, the second (θ) governs the selection of argument fillers for each semantic role, and the third (ψ) models (and penalizes) duplicate occurrence of roles. Each frame occurrence is generated independently given these distributions. Let us describe the model by first defining how the set of model parameters and an argument key clustering are drawn, and then explaining the generation of individual frame instances. The generative story is formally presented in Figure 2.

For each frame, we begin by drawing a distribution of its lexical units from a DP prior $DP(\gamma, H^{(P)})$ with a small concentration parameter γ, and a base distribution $H^{(P)}$, pre-computed as normalized counts of all verbs in our dataset. Next, we generate a partition of argument keys B_f from CRP(α) with each subset $r \in B_f$ representing a single frame specific semantic role. The crucial part of the model is the set of selectional preference parameters $\theta_{f,r}$, the distributions of arguments x for each role r of frame f. We represent arguments by lemmas of their syntactic heads.[2] In order to encode the assumption about sparseness of the distributions $\theta_{f,r}$, we draw them from the DP prior $DP(\beta, H^{(A)})$ with a small concentration parameter β, the base probability distribution $H^{(A)}$ is just the normalized frequencies of arguments in the corpus. Finally, the geometric distribution $\psi_{f,r}$ is used to model the number of times a role r appears with a given frame occurrence. The decision whether to generate at least one role r is drawn from the uniform Bernoulli distribution. If 0 is drawn then the semantic role is not realized for the given occurrence, otherwise the number of additional roles r is drawn from the geometric distribution $Geom(\psi_{f,r})$. The Beta priors over ψ indicate the preference towards generating at most one argument for each role.

Now, when parameters and argument key clusterings are chosen, we can summarize the remainder of the generative story as follows. We begin by independently drawing occurrences for each frame. For each frame occurrence, we first draw its lexical unit. Then for each role we independently decide on the number of role occurrences. Then we generate each of the arguments (see **GenArgument** in Figure 2) by generating an argument key $k_{f,r}$ uniformly from the set of argument keys assigned to the cluster r, and finally choosing its filler $x_{f,r}$, where the filler is either a lemma or the syntactic head of the argument.

3.2 Inference

We use a simple approximate inference algorithm based on greedy search for the maximum a-posteriori clustering of lexical units and argument keys. We begin by assigning each verbal predicate to its own frame, and then iteratively choose a pair of frames and merge them. Note that each merge involves inducing a new set of roles, i.e. a re-clustering of argument keys, for the new merged frame. We use the search procedure proposed in (Titov and Klementiev, 2012), in order to cluster argument keys for each frame.

Our search procedure chooses a pair of frames to merge based on the largest incremental change to the objective due to the merge. Computing the change involves re-clustering of argument keys, so considering all pairs of initial frames containing single verbal predicates is computationally expensive. Instead, we

[2] For prepositional phrases, we take as head the head noun of the object noun phrase as it encodes crucial lexical information. However, the preposition is not ignored but rather encoded in the corresponding argument key.

Parameters:

for each frame $f = 1, 2, \ldots$:
$\phi_f \sim DP(\gamma, H^{(P)})$ [distrib of lexical units]
$B_f \sim CRP(\alpha)$ [partition of arg keys]
for each role $r \in B_f$:
$\theta_{f,r} \sim DP(\beta, H^{(A)})$ [distrib of arg fillers]
$\psi_{f,r} \sim Beta(\eta_0, \eta_1)$ [geom distr for dup roles]

Data Generation:

for each frame $f = 1, 2, \ldots$:
for each occurrence of frame f:
$p \sim \phi_f$ [draw a lexical unit]
for every role $r \in B_f$:
if $[n \sim Unif(0, 1)] = 1$: [role appears at least once]
GenArgument(f, r) [draw one arg]
while $[n \sim \psi_{f,r}] = 1$: [continue generation]
GenArgument(f, r) [draw more args]

GenArgument(f, r):
$k_{f,r} \sim Unif(1, \ldots, |r|)$ [draw arg key]
$x_{f,r} \sim \theta_{f,r}$ [draw arg filler]

Figure 2: Generative story for the frame-semantic parsing model.

prune the space of possible pairs of verbs using a simple but effective pre-processing step. Each verb is associated with a vector of normalized aggregate corpus counts of syntactic dependents of the verb (ignoring the type of dependency relation). Cosine similarity of these vectors are then used to prune the pairs of verbs so that only verbs which are distributionally similar enough are considered for a merge. Finally, the search terminates when no additional merges result in a positive change to the objective.

4 Experimental Evaluation

4.1 Data

We used the dependency representation of the FrameNet corpus (Bauer et al., 2012). The corpus is automatically annotated with syntactic dependency trees produced by the Stanford parser. The data consists of 158,048 sentences with 3,474 unique verbal predicates and 722 gold frames.

4.2 Evaluation Metrics

We cannot use supervised metrics to evaluate our models, since we do not have an alignment between gold labels and clusters induced in the unsupervised setup. Instead, we use the standard purity (PU) and collocation (CO) metrics as well as their harmonic mean (F1) to measure the quality of the resulting clusters. Purity measures the degree to which each cluster contains arguments (verbs) sharing the same gold role (gold frame) and collocation evaluates the degree to which arguments (verbs) with the same gold roles (gold frame) are assigned to a single cluster, see (Lang and Lapata, 2010). As in previous work, for role induction, the scores are first computed for individual predicates and then averaged with the weights proportional to the total number occurrences of roles for each predicate.

4.3 Model Parameters

The model parameters were tuned coarsely by visual inspection: $\alpha = 1.e$-5, $\beta = 1.e$-4, $\gamma = 1$, $\eta_0 = 100$, $\eta_1 = 1.e$-10. Only a single model was evaluated quantitatively to avoid overfitting to the evaluation set.

4.4 Qualitative Evaluation

Our model induced 128 multi-verb frames from the dataset. Out of 78,039 predicate occurrences in the data, these correspond to 18,963 verb occurrences (or, approximately, 25%). Some examples of the induced multi-verb frames are shown in Table 1. As we can observe from the table, our model clusters semantically related verbs into a single frame, even though they may not correspond to the same gold frame in FrameNet. Consider, for example, the frame *(ratify::sign::accede)*: the verbs are semantically related and hence they should go into a single frame, as they all denote a similar action.

Another result worth noting is that the model often clusters antonyms together as they are often used in similar context. For example, consider the frame *(cool::heat::warm)*, the verbs *cool*, *heat* and *warm*, all denote a change in temperature. This agrees well with annotation in FrameNet. Similarly, we cluster *sell* and *purchase* together. This contrasts with FrameNet annotation as FrameNet treats them not as antonyms but as different views on same situation and according to their guidelines, different frames are assigned to different views.

Often frames in FrameNet correspond to more fine-grained meanings of the verbs, as we can see in the example for *(plait::braid::dye)*. The three describe a similar activity involving hair but FrameNet

Induced frames	FrameNet frames corresponding to the verbs
(rush::dash::tiptoe)	rush : [Self_motion](150) [Fluidic_motion](19) dash : [Self_motion](100) tiptoe : [Self_motion](114)
(ratify::sign::accede)	ratify : [Ratification](41) sign : [Sign_agreement](81) [Hiring](18) [Text_Creation](1) accede : [Sign_Agreement](31)
(crane::lean::bustle)	crane : [Body_movement](26) lean: [Change_posture](70) [Placing](22) [Posture](12) bustle : [Self_motion](55)
(cool::heat::warm)	cool : [Cause_temperature_change](27) heat: [Cause_temperature_change](52) warm: [Cause_temperature_change](41) [Inchoative_change_of_temperature](16)
(want::fib::dare)	want : [Desiring](105) [Possession](44) fib : [Prevarication](9) dare : [Daring](21)
(encourage::intimidate::confuse)	encourage : [Stimulus_focus](49) intimidate : [Stimulus_focus](26) confuse: [Stimulus_focus](45)
(happen::transpire::teach)	happen : [Event](38) [Coincidence](21) [Eventive_affecting](1) transpire : [Event](15) teach : [Education_teaching](7)
(do::understand::hope)	do : [Intentionally_affect](6) [Intentionally_act](56) understand : [Grasp](74) [Awareness](57) [Categorization](15) hope : [Desiring](77)
(frighten::vary::reassure)	frighten : [Emotion_directed](44) vary : [Diversity](24) reassure : [Stimulus_focus](35)
(plait::braid::dye)	plait : [Hair_configuration](11) [Grooming](12) braid : [Hair_configuration](7) [Clothing_parts](6) [Rope_manipulation](4) dye : [Processing_materials](18)
(sell::purchase)	sell : [Commerce_sell](107) purchase : [Commerce_buy](93)
(glisten::sparkle::gleam)	glisten : [Location_of_light](52) [Light_movement](1) sparkle : [Location_of_light](23) [Light_movement](3) gleam : [Location_of_light](77) [Light_movement](4)
(forestall::shush)	forestall : [Thwarting](12) shush : [Silencing](6)

Table 1: Examples of the induced multi-verb frames. The left column shows the induced verb clusters and the right column lists the gold frames corresponding to each verb and the number in the parentheses are their occurrence counts.

gives them a finer distinction. Arguably, implicit supervision signal present in the unlabeled data is not sufficient to provide such fine-grained distinctions.

The model does not distinguish verb senses, i.e. it always assigns a single frame to each verb, so there is an upper bound on our clustering performance.

4.5 Quantitative Evaluation

Now we turn to quantitative evaluation of both frame and role induction.

Frame Labeling. In this section, we evaluate how well the induced frames correspond to the gold standard annotation. Because of the lack of relevant previous work, we use only a trivial baseline which places each verb in a separate cluster (*NoClustering*). The results are summarized in Table 3.

As we can see from the results, our model achieves a small, but probably significant, improvement in the F1-score. Though the scores are fairly low, note that, as discussed in Section 4.4, the model is severely penalized even for inducing semantically plausible frames such as the frame *(plait::braid::dye)*.

Role Labeling. In this section, we evaluate how well the induced roles correspond to the gold standard annotation. We use two baselines: one is the syntactic baseline *SyntF*, which simply clusters arguments according to the dependency rela-

	PU	CO	F1
Our approach	78.9	71.0	**74.8**
NoFrameInduction	79.2	70.7	**74.7**
SyntF	69.9	73.3	**71.6**

Table 2: Role labeling performance.

tion to their head, as described in (Lang and Lapata, 2010), and the other one is a version of our model which does not attempt to cluster verbs and only induces roles (*NoFrameInduction*). Note that the *NoFrameInduction* baseline is equivalent to the *factored* model of Titov and Klementiev (2012). The results are summarized in Table 2.

First, observe that both our full model and its simplified version *NoFrameInduction* significantly outperform the syntactic baseline. It is important to note that the syntactic baseline is not trivial to beat in the unsupervised setting (Lang and Lapata, 2010). Though there is a minor improvement from inducing frames, it is small and may not be significant.[3]

Another observation is that the absolute scores of all the systems, including the baselines, are significantly below the results reported in Titov and Klementiev (Titov and Klementiev, 2012) on the CoNLL-08 version of PropBank in a comparable setting (auto parses, gold argument identification): 73.9 % and 77.9 % F1 for *SyntF* and *NoFrameInduction*, respectively. We believe that the main reason for this discrepancy is the difference in the syntactic representations. The CoNLL-08 dependencies include function tags (e.g., *TMP*, *LOC*), and, therefore, modifiers do not need to be predicted, whereas the Stanford syntactic dependencies do not provide this information and the model needs to induce it.

It is clear from these results, and also from the previous observation that only 25% of verb occurrences belong to multi-verb clusters, that the model does not induce sufficiently rich clustering of verbs. Arguably, this is largely due to the relatively small size of FrameNet, as it may not provide enough evidence for clustering. Given that our method is quite efficient, a single experiment was taking around 8 hours on a single CPU, and the procedure is highly parallelizable, the next step would be to use a much larger and statistically representative corpus to induce the representations.

[3]There is no well-established methodology for testing statistical significance when comparing two clustering methods.

	PU	CO	F1
Our approach	77.9	31.4	**44.7**
NoClustering	80.8	29.0	**42.7**

Table 3: Frame labeling performance.

Additional visual inspection suggest that the data is quite noisy primarily due to mistakes in parsing. The large proportion of mistakes can probably be explained by the domain shift: the parser is trained on the WSJ newswire data and tested on more general BNC texts.

5 Related Work

The space constraints do not permit us to provide a comprehensive overview of related work. Aside from the original model of Titov and Klementiev (2012), the most related previous method is the Bayesian method of Titov and Klementiev (2011). In that work, along with predicate-argument structure, they also induce clusterings of dependency tree fragments (not necessarily verbs). However, their approach uses a different model for argument generation, a different inference procedure, and it has only been applied and evaluated on biomedical data. The same shallow semantic parsing task has also been considered in the work of Poon and Domingos (2009; 2010), but using a MLN model and, again, only on the biomedical domain. Another closely related vein of research is on semi-supervised frame-semantic parsing (Fürstenau and Lapata, 2009; Das and Smith, 2011).

6 Conclusions

This work is the first to consider the task of unsupervised frame-semantic parsing. Though the quantitative results are mixed, we showed that meaningful semantic frames are induced. In the future work, we intend to consider much larger corpora and to focus on a more general set-up by relaxing the assumption that frames are evoked only by verbal predicates.

Acknowledgements

The authors acknowledge the support of the MMCI Cluster of Excellence, and thank Caroline Sporleder, Alexis Palmer and the anonymous reviewers for their suggestions.

References

Daniel Bauer, Hagen Fürstenau, and Owen Rambow. 2012. The dependency-parsed framenet corpus. In *International conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkey.

A. Burchardt, K. Erk, A. Frank, A. Kowalski, S. Pado, and M. Pinkal. 2006. The SALSA corpus: a german corpus resource for lexical semantics. In *LREC*.

Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling. In *CoNLL*.

D. Das and N.A. Smith. 2011. Semi-supervised frame-semantic parsing for unknown predicates. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1435–1444. Association for Computational Linguistics.

D. Das, N. Schneider, D. Chen, and N.A. Smith. 2010. Probabilistic frame-semantic parsing. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 948–956. Association for Computational Linguistics.

K. Erk and S. Pado. 2006. Shalmaneser–a toolchain for shallow semantic parsing. In *Proceedings of LREC*, volume 6. Citeseer.

Charles J. Fillmore. 1968. The case for case. In Bach E. and Harms R.T., editors, *Universals in Linguistic Theory*, pages 1–88. Holt, Rinehart, and Winston, New York.

Hagen Fürstenau and Mirella Lapata. 2009. Graph alignment for semi-supervised semantic role labeling. In *EMNLP*.

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

Trond Grenager and Christoph Manning. 2006. Unsupervised discovery of a statistical verb lexicon. In *EMNLP*.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL-2009), June 4-5*.

Richard Johansson and Pierre Nugues. 2008. Dependency-based semantic role labeling of PropBank. In *EMNLP*.

Joel Lang and Mirella Lapata. 2010. Unsupervised induction of semantic roles. In *ACL*.

Joel Lang and Mirella Lapata. 2011a. Unsupervised semantic role induction via split-merge clustering. In *ACL*.

Joel Lang and Mirella Lapata. 2011b. Unsupervised semantic role induction with graph partitioning. In *EMNLP*.

Suresh Manandhar, Ioannis P. Klapaftis, Dmitriy Dligach, and Sameer S. Pradhan. 2010. Semeval-2010 task 14: Word sense induction and disambiguation. In *Proceedings of the 5th International Workshop on Semantic Evaluation*.

K.H. Ohara, S. Fujii, T. Ohori, R. Suzuki, H. Saito, and S. Ishizaki. 2004. The japanese framenet project: An introduction. In *Proceedings of LREC-04 Satellite Workshop Building Lexical Resources from Semantically Annotated Corpora(LREC 2004)*, pages 9–11.

Alexis Palmer and Caroline Sporleder. 2010. Evaluating FrameNet-style semantic parsing: the role of coverage gaps in FrameNet. In *COLING*.

M. Palmer, D. Gildea, and P. Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Hoifung Poon and Pedro Domingos. 2009. Unsupervised semantic parsing. In *EMNLP*.

H. Poon and P. Domingos. 2010. Unsupervised ontology induction from text. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 296–305. Association for Computational Linguistics.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. 2006. Framenet ii: Extended theory and practice. available at `http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126`.

C. Subirats. 2009. Spanish framenet: A frame-semantic analysis of the spanish lexicon. *Berlin/New York: Mouton de Gruyter*, pages 135–162.

Mihai Surdeanu, Adam Meyers Richard Johansson, Lluís Màrquez, and Joakim Nivre. 2008. The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In *CoNLL 2008: Shared Task*.

Richard Swier and Suzanne Stevenson. 2004. Unsupervised semantic role labelling. In *EMNLP*.

Ivan Titov and Alexandre Klementiev. 2011. A Bayesian model for unsupervised semantic parsing. In *ACL*.

Ivan Titov and Alexandre Klementiev. 2012. A bayesian approach to semantic role induction. In *Proc. EACL*, Avignon, France.